IEEE

10th ANNIVERSARY

AICT2016
International
Conference
12-14 October 2016, Baku

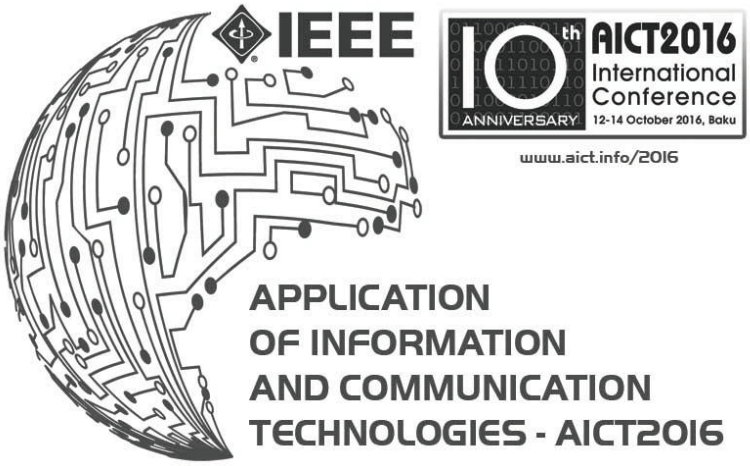
www.aict.info/2016

APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGIES - AICT2016

CONFERENCE PROCEEDINGS

12-14 October 2016, Baku, Azerbaijan

www.aict.info/2016



CONFERENCE PROCEEDINGS

2016 10th International Conference on Application of Information and Communication Technologies (AICT)

IEEE Catalog Number: CFP1656H-PRT
ISBN: 978-1-5090-1840-6

Additional copies of this publication are available from
Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
+1 845 758 0400
+1 845 758 2633 (FAX)
email: curran@proceedings.com

© Designed by Designing and Publishing Department of Qafqaz University, 2016.

IEEE Catalog Number: CFP1656H-PRT
ISBN: 978-1-5090-1840-6

CONFERENCE COMMITTEES:

CONFERENCE HONORABLE CHAIRMEN

Mr. Mikail Jabbarov, *Minister of Education of Azerbaijan*
Mr. Ramin Guluzade, *Minister of Communications and High Technologies of Azerbaijan*

CONFERENCE CHAIRMEN

Professor Havar Mamedov, *Rector of Qafqaz University, Azerbaijan*
Professor Abel Maharramov, *Baku State University Rector, Azerbaijan*
Assoc.Professor Elmar Gasimov, *Baku Higher Oil School Rector, Azerbaijan*
Professor Nargiz Pashayeva, *Lomonosov Moscow State University Baku branch Rector, Azerbaijan*
Professor Xalıq Yahudov, *Azerbaijan Technical University Rector, Azerbaijan*
Professor Mohn Harun Abdullah, *University Malaysia Sabah Vice-Chancellor, Malaysia*
Professor Rasim Aliguliyev, *Institute of Information Technologies of ANAS Director, Azerbaijan*
Professor Telman Aliyev, *Institute of Control Systems of ANAS Director, Azerbaijan*
Professor Manzoor H. Soomro, *ECO Science Foundation (ECOSF) President, Pakistan*
Professor Alexander Khoroshilov, *UNESCO Institute for Information Technologies in Education (IITE)*

CONFERENCE VICE CHAIRS

Professor Niftali Gocayev, *Qafqaz University, Azerbaijan*
Professor Aydin Kazimzade, *Baku State University, Azerbaijan*
Professor Azad Tagizade, *Lomonosov Moscow State University Baku branch, Azerbaijan*
Assoc.Professor Ag. Asri Ag. Ibrahim, *University Malaysia Sabah, Malaysia*
Professor Oktay Nusratov, *Institute of Control Systems of ANAS, Azerbaijan*
Dr. Rashid Alakbarov, *Institute of Information Technologies of ANAS, Azerbaijan*

GENERAL CHAIRS

Assoc.Professor Abzetdin Adamov, *Qafqaz University, Azerbaijan*

TECHNICAL PROGRAM CO-CHAIRS

Professor *H.Levent Akin*, *Bogazici University, Turkey*
Assoc.Professor *Vincent Guyot*, *ESIEA/LIP6, France*
Professor *Asoke Talukder*, *IIIT, Bangalore, India*
Professor *Serdar Korukoğlu*, *Ege University, Turkey*
Professor *In Jeong Lee*, *ITIRC Vice-president, Hoseo University, Korea*
***Dr. Tae Woon Kang*,** *Professor and Consultant, National Research Foundation, Korea*
Assoc.Professor *Kshetrimayum Rakhesh Singh*, *IIT-Guwahati, Guwahati, Assam, India*
***Dr. Nazim Agoulmine*,** *University of Evry Val d'Essonne, France*
***Professor Vladimir Kureychik*,** *Southern Federal University, Russia*
***Dr. Gennady Smorodin*,** *EMC Academic Alliance Russia & CIS, Russia*

PANEL CHAIR

Assoc.Professor Abzetdin Adamov, *Qafqaz University, Azerbaijan*

KEYNOTES, TUTORIALS AND PUBLICATION CHAIR

Assoc.Professor Abzetdin Adamov, *Qafqaz University, Azerbaijan*

LOCAL ORGANIZING COMMITTEE

Professor Khalil Ismailov, *Dean of Engineering Faculty, Qafqaz University*
Assoc.Professor Abzetdin Adamov, *Director of Applied Research Center CeDAWI, Qafqaz University*
Assoc.Professor Babek Abbasov, *Head of Information Technology and Systems Department, Qafqaz University*
Professor Hamzaga Orujov, *Head of Department of Mathematics, Qafqaz University*
Assoc.Professor Vugar Musayev, *Institute of Information Technologies of ANAS*
Professor Fakhreddin Isayev, *Head of Department of Mathematics and Informatics, Qafqaz University*
Dr. Anar Rustemov, *Information Technology and Systems Department, Qafqaz University*
Asgar Mammadli, *CeDAWI Research Center, Qafqaz University*
Sevil Imanova, *Industrial Engineering Department, Qafqaz University*
Etibar Seyidzade, *Computer Engineering Department, Qafqaz University*
Khayyam Masiyev, *Computer Engineering Department, Qafqaz University*
Anar Khocayev, *Computer Engineering Department, Qafqaz University*
Rauf Suleymanli, *Computer Engineering Department, Qafqaz University*

ADVISORY AND STEERING COMMITTEE

Niftali Gocayev, *Vice Rector for Research and Science, Qafqaz University, Azerbaijan*
Victor Kureichik, *Taganrog Institute of Technology, Southern Federal University, Russia*
Mehmet Ulema, *Manhattan College, USA*
Adnan Yazici, *Middle East Technical University, Turkey*
Raouf Boutaba, *University of Waterloo, Canada*
Guy Omidyar, *IFIPTC6 WG6.8 Past Chair, USA*
Insepov Zinetulla, *Nazarbayev University Research and Innovation System, Kazakhstan*

TECHNICAL PROGRAM COMMITTEE MEMBERS

Cem Ersoy, *Bogazici University, IEEE ComSoc Chapter Chair, Turkey*
Asoke K Talukder, *IIIT Bangalore, India*
Etibar Seyidzade, *Qafqaz University, Azerbaijan*
Agasi Melikov, *National Aviation Academy, Azerbaijan*
Nadir Alishov, *Ukrain National Akademy of Science, Ukrain*
Petr Sosnin, *Ulyanovsk State Technical University, Russia*
Abzetdin Adamov, *Qafqaz University, Azerbaijan*
Galina Kedrova, *Moscow State Lomonosov University, Russia*
Khalil Ismailov, *Qafqaz University, Azerbaijan*
Ali Sahintas, *Qafqaz University, Azerbaijan*
Mehmet Ulema, *Manhanttan College, USA*
Masoud Mashhouri, *Carleton University, Canada*
Kamil Saraç, *University of Texas at Dallas, USA*
Ramiz Alguliyev, *Institute of Information Technologies of ANAS, Azerbaijan*
Galina Mehdiyeva, *Baku State University, Azerbaijan*
Anar Rustemov, *Qafqaz University, Azerbaijan*
Yashar Hajiyev, *Qafqaz University, Azerbaijan*
Babek Abbasov, *Qafqaz University, Azerbaijan*
Reijo Savola, *VTT Technical Research Centre of Finland, Finland*
Alexander Afanasyev, *Ulyanovsk State Technical University, Russia*
Ali Pourmohammad, *Amirkabir University of Technology, Iran*
Eyyup Gulbandilar, *Eskisehir Osmangazi University, Turkey*
Lyazzat Atymbayeva, *Kazakh National Research Technical University named after K.I. Satpayev, Kazakhstan*
Mohamed Musa, *Sudan University of Science and Technology, Sudan*
Muammer Akcay, *Kütahya Dumlupınar University, Turkey*
Murat Koyuncu, *Atilim University, Turkey*
Samir Rustamov, *ADA University, Institute of Control Systems of Azerbaijan National Academy of Sciences, Azerbaijan*
Victor Abrukov, *Chuvash State University, Russia*
Efendi Nasibov, *Dokuz Eylül University, Turkey*
Kashif Nisar, *University Malaysia Sabah, Malaysia*
Mohd Helmy Abd Wahab, *Universiti Tun Hussein Onn Malaysia, Malaysia*
Tofig Hasanov, *Kyoto Institute of Technology, Japan*
Hasan Ogul, *Bashkent University, Turkey*

TABLE OF CONTENTS

SESSION 1 – BIG DATA MANAGEMENT AND APPLICATION

SESSION 2 – DATA MINING AND DATA ENGINEERING

SESSION 3. CYBER SECURITY ISSUES

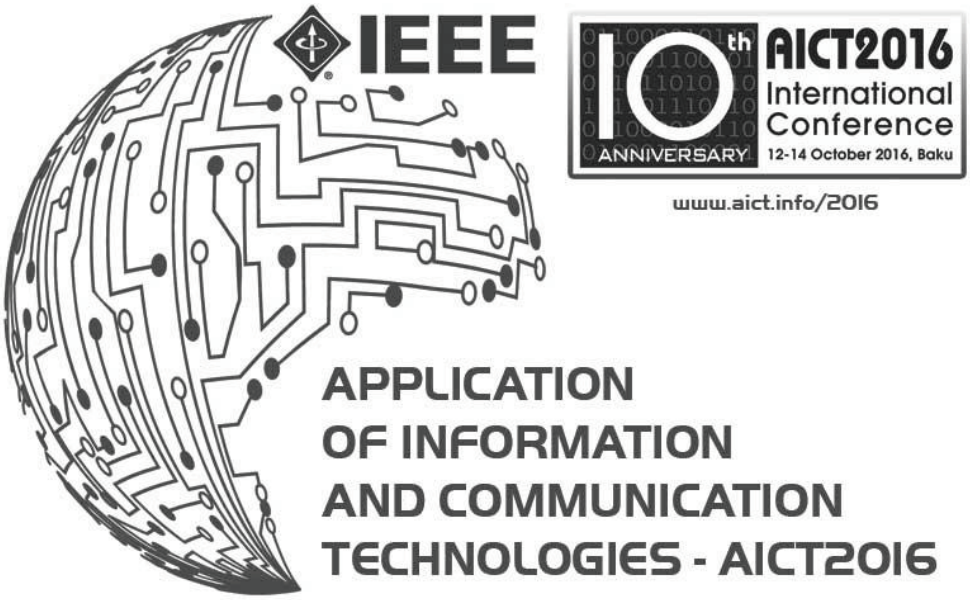
SESSION 4. LATEST TRENDS IN ICT APPLICATION

SESSION 5. COMMUNICATION, NETWORK AND HARDWARE

SESSION 6. ICT IN GOVERNANCE, BUSINESS ADMINISTRATION, FINANCE AND ECONOMY

SESSION 7. ICT IN EDUCATION, HUMANITIES, SOCIAL SCIENCES AND RESEARCH

SESSION 8. ICT IN MEDICINE AND HEALTH CARE



CONFERENCE PROCEEDINGS

SESSION 1.

BIG DATA MANAGEMENT AND APPLICATION

12-14 October 2016, Baku, Azerbaijan

The case study approach to learning Text Mining

Yerzhan Baiburin,
Zheniskul Zhantassova
S. Amanzholov East Kazakhstan State University
Oskemen, Kazakhstan
ebaiburin@gmail.com

Aliya Nugumanova,
Aigul Syzdykpayeva
D. Serikbayev East Kazakhstan State Technical University
Oskemen, Kazakhstan
yalisha@yandex.kz

Igor Bessmertny
Saint Petersburg National Research University of ITMO
Saint Petersburg, Russia
igor_bessmertny@hotmail.com

***Abstract*— In this paper we discuss the case study approach to learning Text Mining techniques. We propose a novel framework that supports case-based learning and implements the SMART-goal setting methodology.**

***Index Terms*— case study, Text Mining, learning approach.**

I. Introduction

In this paper, we propose a novel framework that provide students to learn Text Mining in a smart practical way. The key idea of a proposed framework is based on the SMART-goal setting technique. The technique is used for creating goals, which are specific (S), measurable (M), attainable (A), realistic (R) and timed (T). Nowadays this technique is actively used in pedagogy to improve students' motivation and engagement.

When the technique is applied to learning Text Mining, learning smart goals can be defined on the base of a simple and an efficient template (see table 1). For example, a learning goal can be set as follows: «Learn to solve the task of keyword extraction with the precision more than 60% using the Chi-squared method for the geology domain texts by the 16th June».

TABLE I. The Smart goal template

S	Learn to solve the <SPECIFIC TASK>
M	With the <SUCCESS CRITERIA>
A	Using the <SET OF AVAILABLE TOOLS>
R	For the <REAL DOMAIN>
T	By the <DATE>

We believe that the most efficient way to achieve learning goals formulated in such "smart" manner is case-based learning. Case-based learning also referred to as case study learning is a pedagogical method, which used "to generate an in-depth, multi-faceted understanding of a complex issue in its real-life context" [1]. For this reason it is very capable to cover any real tasks and solve any real problems however complicated they may appear at first sight.

In our opinion, case study learning is completely relevant to realizing smart goals of the learning process. Firstly, case studies are specific, realistic and time-bounded by default. Secondly, when using case study learning, a teacher aims to prepare case materials in order to make learning outcomes attainable and measurable.

Therefore, the aim of the paper is to discuss the design of a framework that provides smart-goals oriented case study learning. The proposed framework should be an effective tool for setting case study goals, for organizing case study materials to achieve defined goals, for providing all necessary means and methods to fulfill a case study research.

The rest part of the paper is organized as follows. The related work is given in Section 2. In Section 3, we formulate the ideology of the proposed framework. The paper is concluded in Section 4.

II. Related work

Text Mining and broader Information Technology are demanding "with respect to professional skills in problem-solving, teamwork, project management, professional ethics and values" [2]. This motivates pedagogues to focus on using of problem-based and case-based learning tools. Nowadays the market of learning systems is saturated with a great number of tools that help pedagogues and students to hone students' professional skills in project and time management, as well as in collaboration work.

However, professional skills are not limited only to management and job performance; the main goal of any course is to increase the level of problem-oriented knowledge of students. In that regard, it is especially important to create new programs and interfaces for case-based learning, which is a very efficient way to augmentation of problem-oriented knowledge in Information Technology and its branches.

As it pointed out in [3], "case studies on their own do not help students to learn – by themselves they are just collections of information". To make cases more valuable and useful it is necessary to link them with "authentic activities that reflect the experience of real word practitioners" [4].

This idea is widely used in many research works. For example, authors of [5] propose a case library, which is a collection of usability engineering case studies for teaching human-computer interaction. In addition, they develop a browser that presents case structures in a direct manner. Authors of [3] create a tool called Visualizing a Development Record. Its

function is the trail of information that others can follow to trace the creation of an interface.

Work [6] is the most relevant to our study. It presents "a generalized interactive case-based learning system with multimedia support, which is easy to generate and play cases". The system is designed to teach programming languages. It is based on XML, describing the information of cases, including the case's title, problem, analysis, playing sections' information, and path information of all the files (program files, text, audio, video, image files) in the database. The most important module of the system is case generating module. With the help of this module, teachers can generate a case courseware by uploading a program code and audio, video, image file for explanation of code lines.

At last, we should to mention some works, which are devoted to issues of realization of case-based learning systems. For example, authors of [2] propose four requirements for interactive case study libraries design. The requirements can be summarized into for concepts: authenticity, social interaction, resource accumulation/updates and communities of practice. Work [7] examines case-based learning in online education from three aspects: instructional design, facilitation and technology support. It formulates several recommendations and limitations for better designing case-based learning activities.

III. Proposed Framework

The proposed framework consists of five modules by the number of components of the smart goal (see Figure 1). The first module (S) allows to define a specific case study (enter a title, enter case description, upload video and audio materials, etc.). The second module (M) allows selecting criteria that will measure if the case is solved successfully. The third module (A) allows selecting a set of predefined methods (services) that will be used to solve the case. The forth module (R) allows select a domain for the case. And, the last module (T) allows setting a duration of the case.

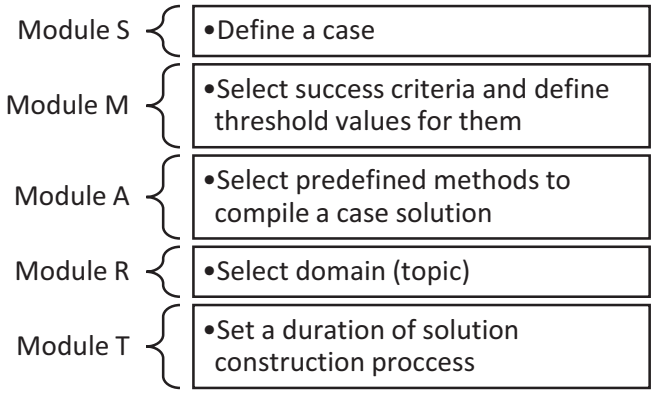


Figure 1 Architecture of the proposed framework

Selecting criteria means that one or more indicators can be selected from a dropdown list. There are three main indictors in Text Mining area. These are precision, recall and F-measure. However, a teacher can be add new indicators to the list and associate them with predefined methods and cases. The list of indicators associated with ontology construction cases is shown in Figure 2.

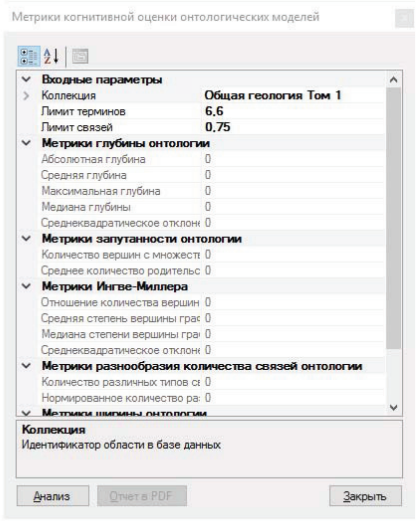


Figure 2 Selection of indicators

Predefined methods are realized as units (services) that can be compiled together in order to solve the case. For example, the full set of predefined methods can consists of lemmatization, tokenization, stop-words excluding, terms-by-documents matrix construction and other similar services. It is a student's privilege to select or reject any method when constructing the solution of the case. Connection of units into the common solution shown in Figure 3.

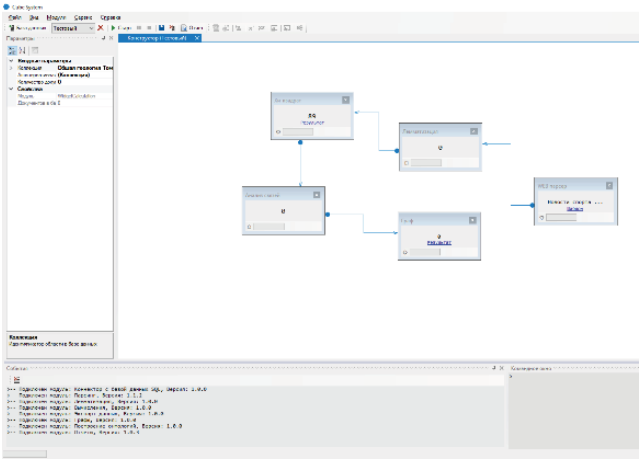


Figure 3 Connection of units into the solution

Domains are realized as predefined collections of texts related to different topics. Domain selection means that case experiments will be conducted on texts of a specific topic. Domain selection module is shown on Figure 4.

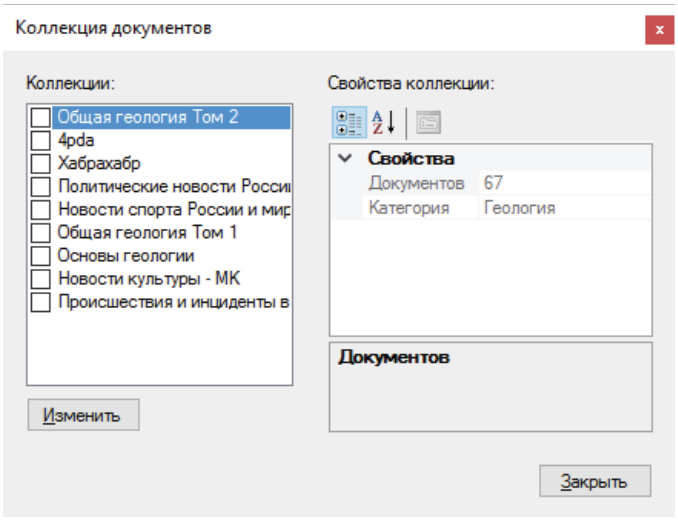


Figure 4 Domain selection module

A pilot version of the proposed framework was implemented in the form of web-application. This application was proposed to master students of Informatics to implement their experiments during course project development. The results of two specific case studies are shown in figure 5 and 6.

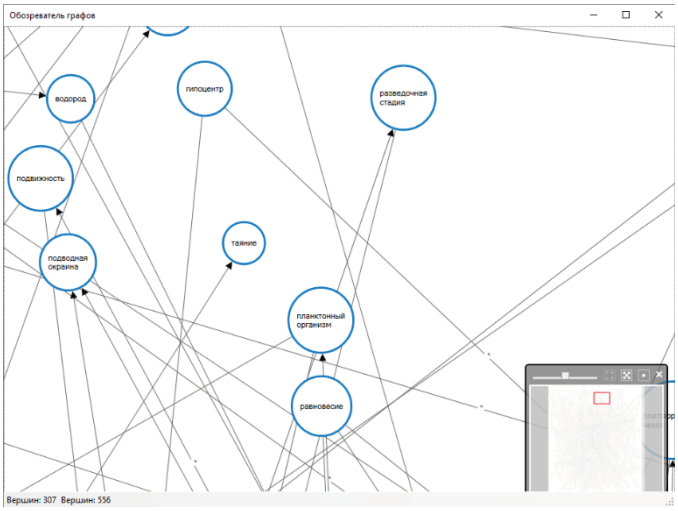


Figure 5 The result of a case study devoted to ontology construction

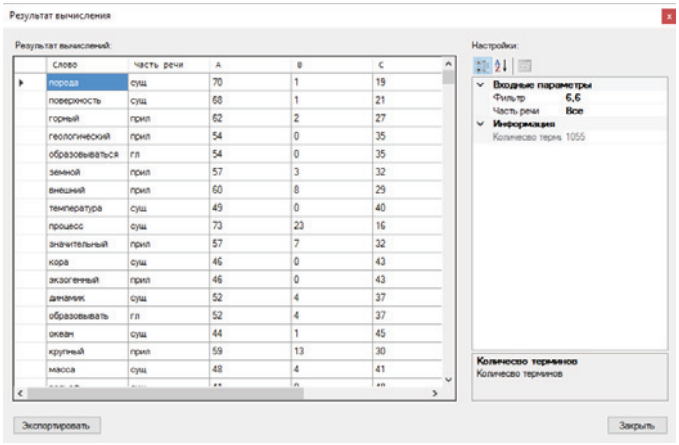


Figure 6 The result of a case study devoted to keyword extraction

IV. Conclusion

In this paper, we explore the problem of computer-supported case learning for Text Mining area. We realize the computer framework that is based on smart goal setting approach. The proposed framework consists of five modules by the number of components of a smart goal. The most important module is the third one, which provides case solution building by choosing methods from the predefined set of services. Through the implementation of services, we can make the proposed framework flexible and a really smart. Therefore, we provide students with a tool, which assists them to create solutions of any complexity.

Our future work will be directed towards integration of the proposed framework with the curriculum development system.

References

[1] Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. BMC medical research methodology, 11(1), 1.

[2] Jiang H., Ganoe C., Carroll J. M. Four requirements for digital case study libraries //Education and Information Technologies. – 2010. – Т. 15. – №. 3. – P. 219-236.

[3] Berry B. et al. Making a Case for HCI: Exploring Benefits of Visualization for Case Studies //Proc. of World Conf. on Educ. Multimedia, Hypermedia & Telecom. EDMEDIA. – 2006.

[4] Bennett, S. Harper, B. & Hedberg, J. (2002). Designing real life cases to support authentic design activities, Australian Journal of Educational Technology,18(2), 1-12

[5] Carroll J. M., Rosson M. B. A case library for teaching usability engineering: Design rationale, development, and classroom experience //Journal on Educational Resources in Computing (JERIC). – 2005. – Т. 5. – №. 1. – С. 3.

[6] Yintao L. et al. An implementation of a case-based learning system //International Conference on Web-Based Learning. – Springer Berlin Heidelberg, 2006. – С. 165-175.

[7] Lee S. et al. A Review of Case-based Learning Practices in an Online MBA Program: A Program-level Case Study //Educational Technology & Society. – 2009. – Т. 12. – №. 3. – С. 178-190

Large Scale System Management Based on Markov Decision Process and Big Data Concept

Sagit Valeev, Natalya Kondratyeva
Dept. of Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
vss2000@mail.ru, knv24@mail.ru

***Abstract* — Processing Big Data in large scale system is a new direction of development of information technologies, including methods and algorithms of streaming data in real time mode. In the paper, application of Markov Decision Process System for large scale system management based on Big Data concept is proposed.**

***Index Terms* — large scale system, efficiency, Markov decision process, lifecycle, modeling, Big Data.**

I. Introduction

Nowadays, under condition of economy complexity and globalization, most of service and production industries, for example, aviation industry, are considered as large scale systems.

However, the main findings of the analysis undertaken in this area of research should be pointed as lack of engineering management methods in regarding aviation industry. It disables considering aviation as a complex system. So, currently, it's impossible to manage and develop aviation industry integrally.

Modern aviation industry is a complex large scale system that includes designing, testing, production, operation, maintains services etc. Thus, it seems reasonable to provide information logistical support throughout the life cycle of single processes in this area. On the other hand, it allows building stochastic models to manage aviation industry in real-time mode under uncertainties and risks conditions. It is possible to provide using Big Data concept [1 – 4].

The concept of Big Data combines techniques and technologies that extract knowledge from huge data streams. Active interest from researchers and practitioners in the field of aviation to Big Data technology due to the implementation of high-speed telecommunication data channels for dynamic objects, the implementation of MapReduce, Hadoop technologies as well as Cloud Computing [5 – 8].

An important place in the framework of standardization of design solutions based on the idea of open systems takes implementation of CALS technologies. It allows forming the basis for the creation and usage of the unified information environment during the design, production, testing, and operation. At the same time, the application of Big Data on the stages of design, testing and operation of control systems and aircraft power plant is only at the initial stage today.

In the paper, Markov decision process system building for large scale system management based on Big Data concept is considered.

II. Markov Decision Process System Building for Large Scale System Management

As the decision making approach in the case of large scale system, Markov decision processes (MDP) can be applied [9]. MDP is a mathematical model applied for decision making in situations where outcomes are partly random and partly under the control of a decision maker.

Definition 1. MDP means a stochastic system characterized by a 5-tuple $\langle \mathbf{S}, \mathbf{A}, A, p, g \rangle$, where the components are as follows: **S** is a finite set of discrete states and **A** is a finite set of control actions. Mapping **A**: $\mathbf{S} \to p(\mathbf{A})$ is the availability function that renders a set of actions available to each state where p denotes the power set. The transition function is given by p: $\mathbf{S} \times \mathbf{A} \to \Delta(\mathbf{S})$, where $\Delta(\mathbf{S})$ is the set of all probability distributions over **S**. Let $p(y \mid x, a)$ denote the probability of arrival at state y after executing action $a \in A(x)$ in state x. The immediate-cost function is defined by g: $\mathbf{S} \times \mathbf{A} \to \mathbf{R}$, where $g(x, a)$ is the cost of taking action a in state x.

Definition 2. A control policy determines the action to take in each state. A deterministic policy, $\pi : \mathbf{S} \to \mathbf{A}$, is simply a function from states to control actions. A randomized policy, $\pi : \mathbf{S} \to \Delta(\mathbf{A})$, is a function from states to probability distributions over actions. We denote the probability of executing action a in state x by $\pi(x)(a)$ or, for short, by $\pi(x, a)$.

In terms of logistics approach, models of large scale system services can be represented as a Markov chain (Fig. 1), where $p_{1,2}$ – the probability of transition from state S_1 to state S_2 relevant to processes of production to operation, $p_{2,1}$ – the probability of transition from state S_2 to state S_1.

Let us consider a more detailed representation of the scheme. Traditionally, production and operation stages of life cycle of large scale system products have week information logistics interconnection I_k, I_m (Fig. 1).

The responsibility for production to operation stages transition as rule is problem of large scale system (transitions $p_{1,2}$ and $p_{2,1}$).

It is possible to decompose each transition to some more detailed stages. For instance testing of new production units can be considered as separate processes too.

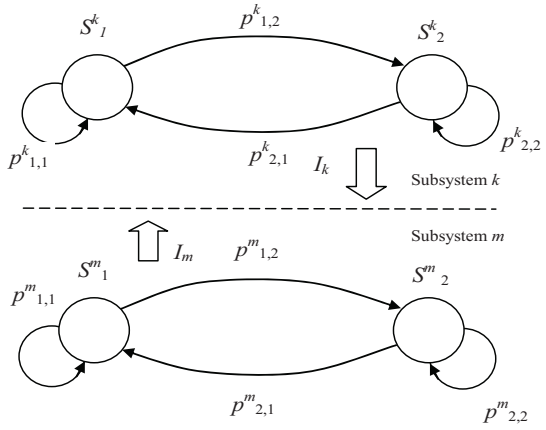


Fig. 1. Transition graph corresponding to the processes of life cycle of large scale system services in general

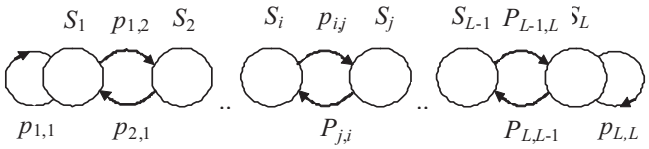


Fig. 2. Detailed transition graph corresponding to the processes of life cycle of large scale system products

Thus, it is appropriate to make the presented in Fig. 1 graph more detailed (Fig. 2).

It is possible to describe the graph in Fig. 2 as the following. Let $S = \{s_1, \ldots, s_L\}$ $(j = 1 \ldots L)$ be the state space which is the same for every step of different processes presenting life cycle of large scale system objects. One-dimensional array of probability distribution on the set of indices j for the n-th step (transition) is denoted as $\pi_n = (\pi_n^{(1)}, \ldots, \pi_n^{(L)})$. Herewith $\pi_n^{(j)}$ is probability of the event when an large scale system object is in a state j on n-th step.

Let's determine a state probability π_n distribution on every single step. Let us define T as an event corresponding to the situation when large scale system object is in a state j on $(n + 1)$-th step. In accordance with agreed notation, the event probability will be $P(T) = \pi_{n+1}^{(j)}$. Let us define Si as an event when aviation object is in a state i on n-th step, therefore $P(S_i) = \pi_n^{(i)}$. Thus p_{ij} is probability the aviation object is state j on $(n + 1)$-th step if he was in a state i on n-th step, i.e. $p_{ij} = p(T \mid S_i)$. We can find the event probability by the total probability equation:

$$\pi_{n+1}^{(j)} = \sum_{i=1}^{L} \pi_n^{(i)} p_{ij}, \quad j = 1, \ldots, L. \tag{1}$$

Equation (1) allows sequential (step by step) determining the change in the probability distribution of aviation object's states while aviation industry service providing. The initial distribution of probabilities should be a priori given.

The probabilities are directly dependent on the risk of failure $R_i(I)$ at each step of the product life cycle.

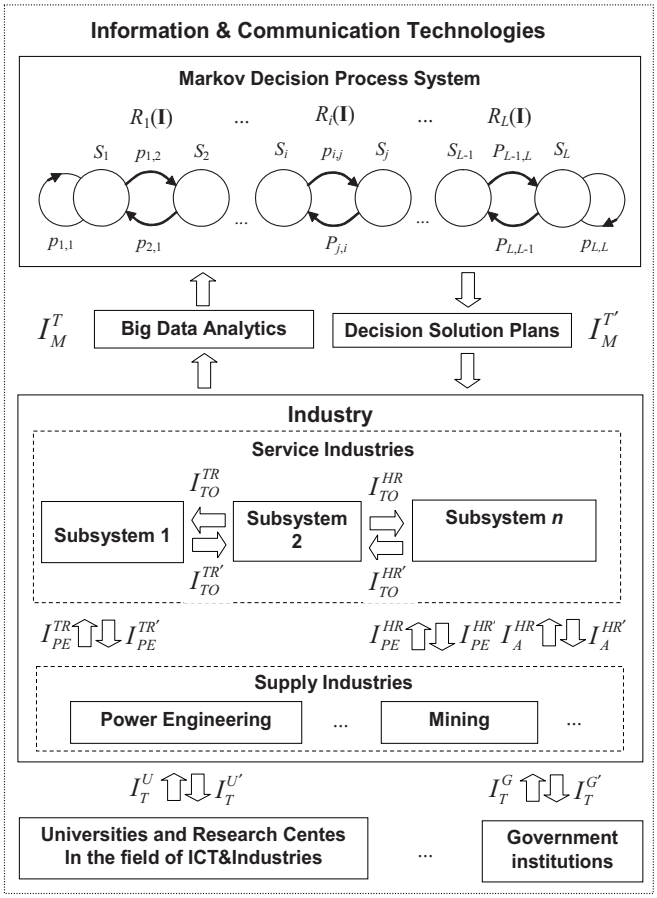


Fig. 3. General scheme of information flows in large scale system for Markov Decision model building

The components of information vector I are presented on the example of general scheme of information flows in industry as a large scale complex system (Fig. 3).

For example, in aviation industry, service industries and supply industries may be allocated. Service industries interact with a product directly on the base of different organizational technical system. Supply industries are related with product indirectly and include power engineering and so on. All these subsystems are connected with material and information flows. In this case, we need to minimize the risk of service failure Ri(I):

$$\sum_{i-1}^{L} R_i(I) \rightarrow \min \tag{2}$$

Interrupting the information flows or information lack or invalidation lead to increasing the risks $R_i(I)$.

Therefore, on authors' opinion, it is need to provide Markov decision process system for aviation industry services as well as system information support through all the aviation industry life cycle elements.

III. Big Data Processing for Markov Decision Process System Building

As it can be seen, such the approach requires huge information content processed at national and world wide level.

Herewith, the following urgent tasks may be allocated for the flows processing (Fig. 4):

- Data mining as the tool of information flows analysis;
- Building a large amount data warehouse;
- Forecasting and optimizing processes in aviation industry;
- High technology tools and equipments to provide information support of aviation industry services.

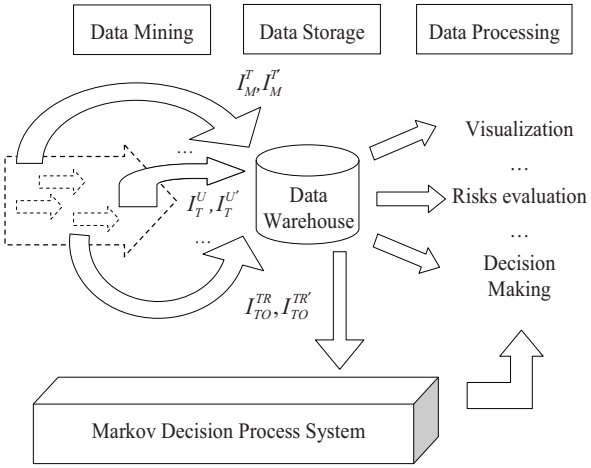


Fig. 4. Information processing of data flows

Multi-agent approach is suggested to build the Markov Decision Process System model of industrial service life cycle. For now, it is applied effectively for business process optimization as well as complex social and technical systems management and control [10]. Let's present the example of strategy for agent in MDP of airport management.

Airports are large scale systems that operate in unique and evolving physical, financial and regulatory environments. For estimation airport efficiency, statistics of total movement, air transport movements, terminal passengers, and freight are applied. The UK Civil Aviation Authority collects statistics from more than 60 UK Airports for the period of 1981-2015 years [11]. So, we can say that some kind of Big Data technologies was applied (Fig.4).

Let's find basic strategy for agent in MDP on the basis of these data. Data of total movement, air transport movements, terminal passengers, freight for 1981-2015 years are presented in Fig. 5 – 8.

For obtaining the main action to reduce costs of airline company the theoretic-information approach based on transfer entropy notion was applied [12, 13].

The transfer entropy extends the concept of mutual information to provide a direction sensitive measure of information flow between two time series. Formally, the transfer entropy from time series Y to X is given by

$$T_{Y\to X} = \sum p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log \frac{p(x_{n+1} \mid x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} \mid x_n^{(k)})}, \qquad (3)$$

where x_{n+1} is the value of X at time $n + 1$, and $x_n^{(k)}$ $(y_n^{(l)})$ is the $k(l)$ lagged values of $X(Y)$ at time n.

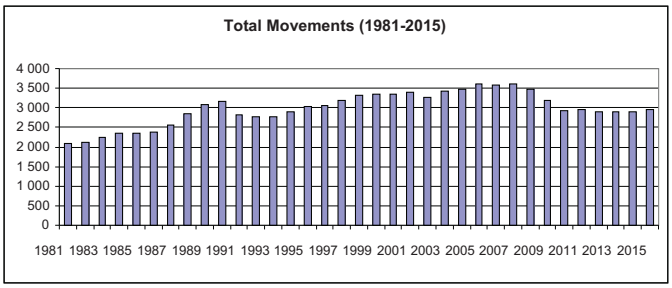


Fig. 5. Total Movements Time Series

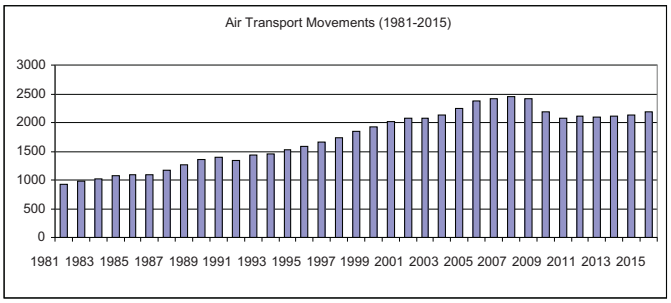


Fig. 6. Air Transport Movements Time Series

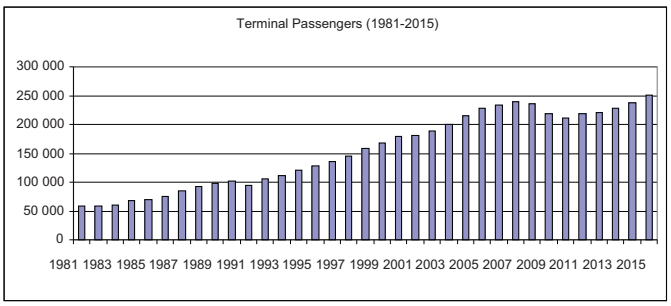


Fig. 7. Terminal Passengers Time Series

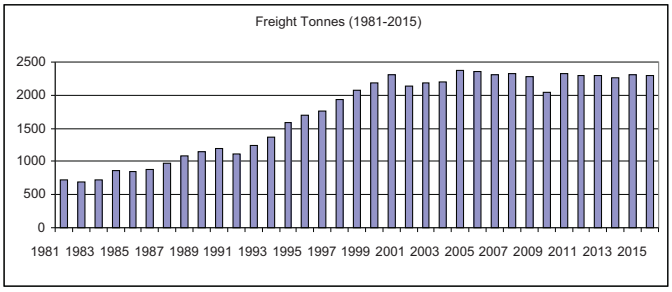


Fig. 8. Freight Time Series

The definition of transfer entropy assumes X as Markov process. The transfer entropy measures the additional amount of information Y contains about X over the information contained in the Markov embedding.

The results of calculation of transfer entropy for statistics data sets are presented in Table 1.

TABLE I. TRANSFER ENTROPY FROM *Y* TO *X*

Y → X	Total Movements	Air Transport Movements	Terminal Passengers	Freight Tonnes
Total Movements	0	0.2	0.3	0
Air Transport Movements	0.2	0	0	0.2
Terminal Passengers	0.3	0	0	0.7
Freight Tonnes	0	0	0	0

The results of analysis of transfer entropy calculations determine the agent action in MDP is control amount of freight or related to fuel consumption. That is why at low cost airlines companies, aircrafts are operating with a minimum set of optional equipment, further reducing costs of acquisition and maintenance, as well as keeping the weight of the aircraft lower and thus saving fuel.

IV. CONCLUSION

Scope of Big Data technologies is very wide. In particular, it covers a complex of organizational and technical systems, such as aircraft systems, aircraft building enterprises, etc. Data on the functioning of these systems is characterized by high volume, heterogeneous structure and a high frequency of updating.

The tendency of large scale systems development can be world wide integration of its elements on the basis of common information space. Under these circumstances, ICT tasks look even more important. These tasks are data mining, building data warehouse, high technology tools and equipments creating, and forecasting and optimizing processes in large scale systems on the base of Big Data analytics. The tasks may be solved with help of interconnected information and communication systems based on innovative technologies.

Big Data approach is suggested to build the MDP model of large scale system. The MDP model implementation will improve the information structure of the large scale system, which, in turn, will reduce costs and increase revenues in the area.

REFERENCES

[1] S. S. Valeev, "Big Data technologies in aviation," in Proc. of the 2nd International Conference "Information Technologies for Intelligent Decision Making Support", 15-17 May, 2014, Ufa: USATU, 2014, pp. 150-152.

[2] S. S. Valeev, N. V. Kondratyeva, "Reference model for large scale industry management based on Big Data concept," in Proc. of the 4th International Conference "Information technologies for Intelligent Decision Making Support," Vol. 1, Russia, Ufa, May 17-19, 2016. Ufa: USATU, 2016, pp. 15-18.

[3] A. Guichvarov, N. Kondratieva, "Technical and economic assessment of aircraft engines fatigue testing on base of simulation modeling," in Proc. of AIAA/ASME/SAE/ASEE Joint Propulsion Conf. & Exhibit. 37th, Salt Lake City, Utah, 8-11 July, 2001, AIAA-2001-3817.

[4] A. S. Guichvarov, N. V. Kondratyeva, "Method of GTE fatigue tests optimization on the basis of lifecycle modeling," in Aviation Industry, 2003, No. 3, pp. 21-25.

[5] A. Adamov, "The Truth about Cloud Computing as new Paradigm in IT," in Proc. of the 3rd IEEE International Conference on Application of Information and Communication Technologies (AICT2009), Azerbaijan-Baku, 14-16 October, 2009.

[6] G. Smorodin, O. Kolesnichenko, "Big Data as the big game changer: Big Data-driven world needs Big Data driven ideology," in Proceedings of The 9th IEEE International Conference on Application of Information and Communication Technologies (AICT2015), Russia, Rostov-on-Don, 14-16 October, 2015, pp. 40-43.

[7] Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Eds. EMC Education Services. Indianapolis: Wiley & Sons, Inc., 2015, pp. 55-86.

[8] Information Storage and Management: Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments, 2nd ed., Eds. EMC Education Services. Indianapolis: Wiley & Sons, Inc., 2015, pp. 67-93.

[9] Cs.Csaji and L.Monostori, "Value Function Based Reinforcement Learning in Changing Markovian Environments," in Journal of Learning Research, No.9, (2008), pp. 1679-1709.

[10] Y.Shoham and K.Leyton-Brown, Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press, 2009.

[11] Civil Aviation Authority. https://www.caa.co.uk/. Access 10 June, 2016.

[12] T. Schreiber, "Measuring information transfer," in Phys. Rev. Lett. 85, 461 (2000).

[13] S. Arya, T. Malamatos, and D. M. Mount, "A Simple Entropy-Based Algorithm for Planar Point Location," in ACM Trans. Algorithms, No. 3 (2007).

Approaches for Interactive Browsing of Large Image Datasets

(Invited Paper)

Gerald Schaefer
Department of Computer Science
Loughborough University Loughborough, U.K.

***Abstract*—With the number of digital images growing at a rapid rate, effective approaches for organising large image repositories are highly sought after. In this paper, we summarise a number of such approaches that allow for intuitive browsing of image databases. Based on visual features extracted from the images, complete image libraries can be visualised in such a way that similar images are located close to each other in the visualisation. This, coupled with appropriate browsing operations, allows for interactive exploration of even large scale image collections as demonstrated by the Hue Sphere Image Browser and Honeycomb Image Browser approaches.**

***Index Terms*—image databases; image libraries; image database navigation; image browsing**

I. Introduction

Since the number of digital images is growing at a rapid rate, methods that are able to deal with and manage these expanding libraries, ideally in an automated fashion, are in high demand. As images are often unannotated [1], content-based image retrieval (CBIR) techniques, which are based on expressing visual similarity between images based on extracted image features [2], [3], [4], [5], have shown great promise.

In this paper, we show how the same features can also be exploited to develop intuitive and effective user interfaces that allow for visualisation and interactive browsing of large image repositories, with an emphasis on some of the systems that we have developed for this purpose in recent years.

Such image database navigation systems [6], [7] typically provide a visualisation of an entire image collection [8] coupled with browsing operations for further exploration of the database [9], and should lead to faster identification and retrieval of images of interest compared to conventional applications [10]. Visualisation methods can in general be divided into mapping-based approaches that employ dimensionality reduction techniques to derive a projection from the high-dimensional features space to the low-dimensional visualisation space [11], [12], [13], [14], clustering-based techniques that are based on grouping images into clusters [15], [16], [17], [18], and graph-based approaches which organise an image collection on a graph or network structure [19], [20], [21], [22].

II. Dimensionality Reduction-based Image Database Visualisation and Browsing

CBIR features extracted from images are typically high-dimensional. On the other hand, image browsers need to organise an image collection in a low-dimensional visualisation space, typically the two-dimensional space defined by a computer screen. The underlying idea of most image database navigation systems is to generate a visualisation in such a way that images that are visually similar, as established based on the extracted features, should be located close to each other. This makes dimensionality reduction techniques a good candidate for this task. Fig. 1 shows an example where multi-dimensional scaling (MDS) [25] was employed on colour histograms [26] extracted from about 1,500 images. As is

Fig. 1. Visualisation of a dataset of about 1,500 images (the UCID [23] image dataset) based on colour histograms and MDS.

Fig. 2. Part of a PCA visualisation of the MIR Flickr 25000 dataset [24].

Fig. 3. MDS grid view of a dataset of 5,000+ images (the MPEG-7 dataset [30]).

apparent, the image collection is organised by colour and thus provides an intuitive layout for the user to explore the dataset.

Principal component analysis (PCA) can also be employed for generating a suitable projection and offers a computationally less expensive but also less flexible alternative. Fig. 2 shows part of a PCA visualisation (again, based on colour features) developed for remote access of image databases [27]. We note that in contrast to Fig. 1 the overlap between images has been eliminated. The reason for this is that image overlap has been shown to lead to less efficient browsing [28] which can be addressed by organising images on a lattice without overlap coupled with image spreading techniques.

As can be seen in Fig. 2, if a large number of images are to placed on screen this can only be achieved by scaling the thumbnail size down accordingly. An alternative way to provide access to large scale image collection is a hierarchical arrangement. Fig. 3 shows an example of an image collection laid out on a hierarchical MDS grid [29].

In this approach, the initial view of Fig. 3 shows the images arranged by colour (using MDS on colour histogram features) on a regular lattice where each shown image corresponds to a cluster of pictures (of similar colour content). The user

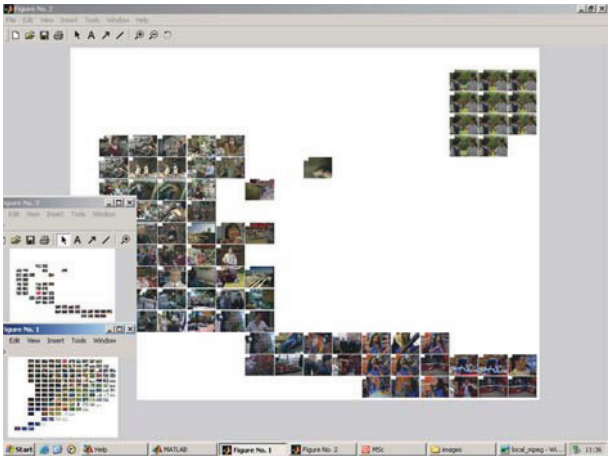


Fig. 4. MDS grid after zooming into a cluster of images.

Fig. 5. Initial view of a Hue Sphere Image Browser session.

has then the possibility to delve further into the database by opening an image cluster to access the images contained there. Based on the size of the dataset, there might be several levels of the browsing hierarchy as shown in Fig. 4 which gives a screenshot of the application after the user has browsed to the third layer of the visualisation hierarchy.

III. Hue Sphere and Honeycomb Image Browsers

While approaches based on dimensionality reduction or clustering have been shown to lead to useful image database navigation systems, one common disadvantage is their computational complexity for calculating the browsing data structure, since both dimensionality reduction and clustering algorithms are computationally demanding, especially for larger datasets.

To address this, we have developed the Hue Sphere Image Browser [31], [32], [33], [34] as an efficient image browsing system without the need of complex data processing algorithms. Here, we also arrange images based on colour, but do not do so based on colour histograms or other complex colour features but on the median colour in HLS colour space of which we retain the hue and lightness components.

Fig. 6. Hue Sphere Image Browser after opening a cluster of images.

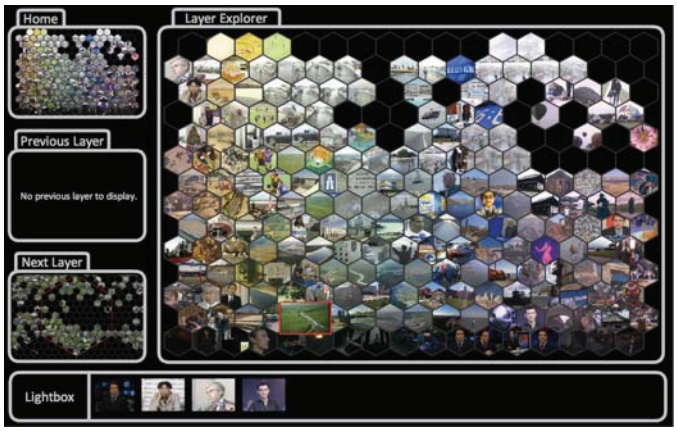


Fig. 7. Initial view of a Honeycomb Image Browser session.

Fig. 9. Mobile Hue Sphere Image Browser.

Consequently, each image is described by two colour co-ordinates which are then mapped to longitude and latitude values to yield a spherical visualisation space upon which the image thumbnails are projected. This, combined with the grid layout and hierarchical organisation employed in the MDS grid, leads to an intuitive and efficient approach to visualise image databases as can be seen from Fig. 5. In contrast to the MDS grid and other approaches though, calculation of the entire browsing structure is very fast, as the mapping between images and their grid cells requires only a few computations.

Browsing can be performed by rotating the image sphere which will bring up images of different colours, tilting the sphere to focus on darker or brighter images, and optical zoom to view larger thumbnail images. Further exploration of the image repository is enabled by opening an image cluster as illustrated in Fig. 6.

The Honeycomb Image Browser [35] shares many of the features of the Hue Sphere Browser. Here also, the median image colour is employed and images are placed on a regular lattice. This lattice however is one comprising a space filling arrangement of hexagons which has the advantage that all neighbouring cells are equidistant and that neighbouring rows of cells are visually displaced as can be seen in Fig. 7.

In the Honeycomb Browser, again each image can be a representative image for a cluster and image browsing can be done by panning and zooming as well as exploring image clusters as shown in Fig. 8.

IV. Mobile Image Browsing

Increasingly, images are not stored and accessed on computers or websites, but on mobile devices such as smart phones due to the ease of capturing images using their integrated cameras and the ubiquity of mobile devices. Clearly, the reduced screen size represents a challenge, while browsing is performed using a variety of touch gestures (e.g. a two-finger pinch for optical zooming) [36].

Fig. 9 shows the Hue Sphere Image Browser ported to a tablet [37], while Fig. 10 shows a mobile version of the Honeycomb Image Browser [38], [39].

V. Conclusions

In this paper, we have given an overview and examples of approaches developed for intuitive visualisation and exploration of image repositories. Based on extracted visual features, images can be organised in such a way that similar images are located close to each other in the generated visualisation. This principle, combined with browsing operations

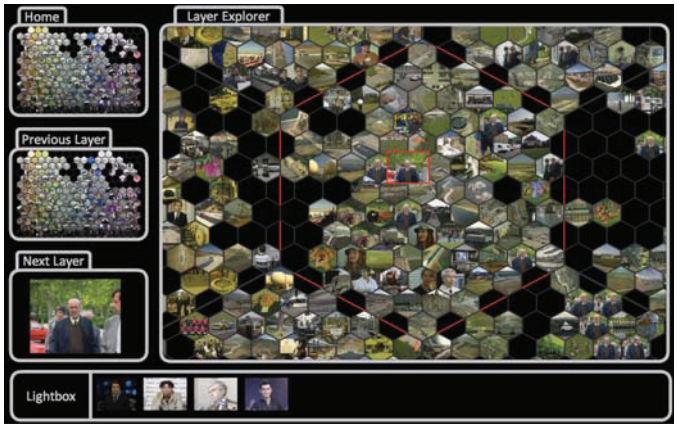


Fig. 8. Honeycomb Image Browser after opening a cluster of images.

Fig. 10. Mobile Honeycomb Image Browser.

to allow further exploration of the image database, forms the basis of effective image database navigation systems.

While most image browsing systems require computationally expensive dimensionality reduction or clustering algorithms, our Hue Sphere Image Browser and Honeycomb Image Browser systems allow the underlying browsing structure to be calculated efficiently while providing intuitive user interfaces for interactive exploration of large image collections, both on desktop computers and on mobile devices.

REFERENCES

[1] K. Rodden, "Evaluating similarity-based visualisations as interfaces for image browsing," Ph.D. dissertation, University of Cambridge Computer Laboratory, 2001.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1249–1380, 2000.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, 2008.

[4] G. Schaefer, "Content-based image retrieval: Some basics," in *2nd International Conference on Man-Machine Interactions*, ser. Advances in Intelligent and Soft Computing, vol. 103, 2011, pp. 21–29.

[5] ——, "Content-based image retrieval: Advanced topics," in *2nd International Conference on Man-Machine Interactions*, ser. Advances in Intelligent and Soft Computing, vol. 103, 2011, pp. 31–37.

[6] D. Heesch, "A survey of browsing models for content based image retrieval," *Multimedia Tools and Applications*, vol. 40, no. 2, pp. 261–284, 2008.

[7] W. Plant and G. Schaefer, "Visualisation and browsing of image databases," in *Multimedia Analysis, Processing and Communications*, ser. Studies in Computational Intelligence. Springer, 2011, vol. 346, pp. 3–57.

[8] ——, "Visualising image databases," in *IEEE International Workshop on Multimedia Signal Processing*, 2009, pp. 1–6.

[9] ——, "Navigation and browsing of image databases," in *International Conference on Soft Computing and Pattern Recognition*, 2009, pp. 750–755.

[10] ——, "Evaluation and benchmarking of image database navigation tools," in *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, vol. 1, 2009, pp. 248–254.

[11] Y. Rubner, L. Guibas, and C. Tomasi, "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval," in *Image Understanding Workshop*, 1997, pp. 661–668.

[12] I. Keller, T. Meiers, T. Ellerbrock, and T. Sikora, "Image browsing with PCA-assisted user-interaction," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2001, pp. 102–108.

[13] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. S. Huang, "Visualization and user-modeling for browsing personal photo libraries," *Int. Journal of Computer Vision*, vol. 56, no. 1-2, pp. 109–130, 2004.

[14] G. P. Nguyen and M. Worring, "Interactive access to large image collections using similarity-based visualization," *Journal of Visual Languages and Computing*, vol. 19, no. 2, pp. 203–224, 2008.

[15] S. Krischnamachari and M. Abdel-Mottaleb, "Image browsing using hierarchical clustering," in *IEEE Symposium on Computers and Communications*, 1999, pp. 301–307.

[16] J. Y. Chen, C. A. Bouman, and J. C. Dalton, "Hierarchical browsing and search of large image databases," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 442–455, March 2000.

[17] I. Bartolini, P. Ciaccia, and M. Patella, "Adaptively browsing image databases with PIBE," *Multimedia Tools and Applications*, vol. 31, no. 3, pp. 269–286, 2006.

[18] A. Gomi, R. Miyazaki, T. Itoh, and J. Li, "CAT: A hierarchical image browser using a rectangle packing technique," in *12th Int. Conference on Information Visualization*, 2008, pp. 82–87.

[19] C. Chen, G. Gagaudakis, and P. Rosin, "Similarity-based image browsing," in *Int. Conference on Intelligent Information Processing*, 2000, pp. 206–213.

[20] D. Heesch and S. Rüger, "NNk networks for content-based image retrieval," in *European Conference on Information Retrieval*, 2004, pp. 253–266.

[21] M. Dontcheva, M. Agrawala, and M. Cohen, "Metadata visualization for image browsing," in *18th Annual ACM Symposium on User Interface Software and Technology*, 2005.

[22] M. Worring, O. de Rooij, and T. van Rijn, "Browsing visual collections using graphs," in *Int. Workshop on Workshop on Multimedia Information Retrieval*, 2007, pp. 307–312.

[23] G. Schaefer and M. Stich, "UCID - An Uncompressed Colour Image Database," in *Storage and Retrieval Methods and Applications for Multimedia*, ser. Proceedings of SPIE, vol. 5307, 2004, pp. 472–480.

[24] M. J. Huiskes and M. S. Lew, "The MIR Flickr retrieval evaluation," in *ACM Int. Conference on Multimedia Information Retrieval*, 2008.

[25] J. B. Kruskal and M. Wish, *Multidimensional scaling*. Sage Publications, 1978.

[26] M. J. Swain and D. H. Ballard, "Color indexing," *Int. Journal of Computer Vision*, vol. 7, no. 11, pp. 11–32, 1991.

[27] W. Plant and G. Schaefer, "Exploring mapping-based visualisations of large remote image databases," in *6th International Symposium on Visual Information Communication and Interaction*, 2013, pp. 111–112.

[28] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "Evaluating a visualisation of image similarity as a tool for image browsing," in *IEEE Symposium on Information Visualization*, 1999, pp. 36–43.

[29] G. Schaefer and S. Ruszala, "Image database navigation on a hierarchical MDS grid," in *International Conference German Pattern Recognition Society*, ser. Springer Lecture Notes on Computer Science, vol. 4174, 2006, pp. 304–313.

[30] Moving Picture Experts Group, "Description of core experiments for MPEG-7 color/texture descriptors," Tech. Rep. ISO/IEC JTC1/SC29/WG11/ N2929, 1999.

[31] G. Schaefer and S. Ruszala, "Image database navigation: A globeal approach," in *International Symposium on Visual Computing*, ser. Springer Lecture Notes on Computer Science, vol. 3804, 2005, pp. 279–286.

[32] ——, "Hierarchical image database navigation on a hue sphere," in *International Symposium on Visual Computing*, ser. Springer Lecture Notes on Computer Science, vol. 4292, 2006, pp. 814–823.

[33] G. Schaefer, "A next generation browsing environment for large image repositories," *Multimedia Tools and Applications*, vol. 47, no. 1, pp. 105–120, 2010.

[34] G. Schaefer and M. Stuttard, "An on-line tool for browsing large image repositories," in *International Conference on Information Retrieval and Knowledge Management*, 2010, pp. 102–106.

[35] W. Plant and G. Schaefer, "Image retrieval on the honeycomb image browser," in *17th IEEE International Conference on Image Processing*, 2010, pp. 3161–3164.

[36] G. Schaefer, "Mobile image browsing," in *16th IEEE Mediterranean Electrotechnical Conference*, 2012, pp. 141–143.

[37] D. Ahlstroem, K. Schoeffmann, M. Hudelist, and G. Schaefer, "A user study on image browsing on small screens," in *20th ACM International Conference on Multimedia*, 2012.

[38] G. Schaefer, M. Tallyn, D. Felton, D. Edmundson, and W. Plant, "Intuitive mobile image browsing on a hexagonal lattice," in *Visual Communications and Image Processing*, 2012.

[39] ——, "Interactive exploration of image collections on mobile devices," in *8th International Conference on Active Media Technology*, 2012.

Investigation of Algorithms and Models of Optimization of Cloud Applications and Services in Virtual Infrastructure of Data Center

Irina Bolodurina
Department of Applied Mathematics
Orenburg State University
Orenburg, Russia
prmat@mail.osu.ru

Denis Parfenov
Faculty of Distance Learning Technologies
Orenburg State University
Orenburg, Russia
fdot_it@mail.osu.ru

***Abstract*— Data centers are widely used for the placement of highly loaded applications and solutions used to process large amounts of data (Big Data). The study developed a model of cloud application and services. The novelty of this model is the use of methods of intellectual analysis and prediction of dynamic characteristics in the study of multicomponent systems. In order to provide flexibility built models used agent-oriented programming. Plurality objects of data center modeled as agents. Each node in this case is an platform of agents, which controls other agents. Agents that act as data resources for cloud applications and cloud services. A distinctive feature of the proposed model is a set of individual parameters needed to perform tasks. It formalizing reflected in the quality of service compliance requirements as part of resource characteristics.**

***Index Terms*—Cloud computing; computing resources; data center; software-defined infrastructure; software-defined networks; software-defined storage; cloud applications; cloud service.**

I. Introduction

In recent years, cloud computing has become a popular approach, used to provide an access to the services and applications for operation of business processes [1]. There are three main models to the provision of cloud computing services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The usage of these approaches in deploying cloud computing platforms has many advantages, such as reliability and quality of service [2]. At the same time, there are a number of limitations, caused both by consumers and by providers of cloud services. For the consumer, cloud resources are endless in terms of scalability. However, if we consider the economic aspect of their consumption, then their ability to scale significantly narrows. From the side of cloud service provider the set of services and computing power are limited [3-4]. In order to maximize the economic aspect of cloud services by increasing the number of users, providers have to apply policies for the flexible usage of allocated resources, while minimizing operation costs. Thereby in today's virtual and physical data centers the problem of resources and cloud applications management is an important issue, because it has a direct impact on the operation costs.

In the past few years the large IT corporations (such as Amazon, Google, Salesforce, IBM, Microsoft, Oracle) develop a renewed approach to the management of the resources and objects in data centers, used for the cloud applications. The main trend in this sphere is the optimization of the data centers resource consumption. In recent our researches we have developed the approaches to the storage optimization of cloud applications' data and to improvement in efficiency of the access to cloud resources [9]. However, they do not solve the assignment problem of cloud application instances in cloud environment. In addition, the review of researches in this field has shown that the problem of optimizing resource selection for the specific types of cloud applications is insufficiently investigated [5-7].

The main objective of our investigation is to construct a models to describe the basic principles of interaction between resources and performance targets for the optimal distribute load and effective launch of cloud applications in the data center. Under the task of organizing the interaction between resources and cloud applications we signify the following list of actions:

1) formalizing of the requirements to execute request to find free networking and computing resources in data center;

2) making plan of a reconfiguration and segmentation of data center resources;

3) reconfiguration and execution of the request for placement and the launch of cloud applications, including the connection to the services and data sources to be placed on the optimal storage devices.

In this part of our investigation, we developed model of multi-agent resource management in data center, built on a software-defined infrastructure that supports self-organizing key objects.

The most effective analysis and forecasting of dynamic characteristics in the investigation of complex systems provide simulation methods [8]. To provide flexibility built models expedient it is to use agent-oriented method. Plurality objects

The research work was funded by Russian Foundation for Basic Research, according to the research projects No. 16-37-60086 mol_a_dk and 16-07-01004.

of data center modeled as agents. Each node in this case is an platform of agents, which controls other agents. Agents that act as data resources for cloud applications and cloud services. A distinctive feature of the proposed model is a set of individual parameters needed to perform tasks. It formalizing reflected in the quality of service compliance requirements as part of resource characteristics.

Consider the basic elements of the multi-agent model. There are many resources of the data center:

$$R = (R_1, R_2, \ldots, R_n) \quad (1)$$

each resource $R_j, j = \overline{1 \ldots N}$ characterized by:

- cost S_j which is expressed in units carrying capacity and is appointed by the resource based on the current state of the data center infrastructure;
- specify the maximum execution time $T\max_j$;
- access priority / starting order P_j;
- a subset of $H_j^R = (H_{j,1}^R, H_{j,2}^R, \ldots, H_{j,k}^R)$ which is a set of parameters that describe the resource, in this case $H_{j,i}^R$ - characterized by the available volume for each of the parameters;
- subset of $E_j^R = (E_{j,1}^R, E_{j,2}^R, \ldots, E_{j,k}^R)$, which represents the efficiency of resource usage for each of the parameters;
- data vector of allowable execution time of the task $PV(R_j)$;
- consider the set of nodes of the data center $WN = (WN_1, WN_2, \ldots, WN_l)$.

Each node $WN_l, l = \overline{1 \ldots L}$ is characterized by:

- vector properties $PV(WN_l)$ - characterize the key parameters of the node data center infrastructure;
- a subset of the resources, so that every resource forms a graph $R_j = (WN(t), Connections(t), Servis(t))$, in which the relationships are branches of computing nodes or network entities, and the arcs are connections to the required services.

In addition, in the model of the data center and software-defined infrastructure stands out a variety of tasks, formalized as the requirements for cloud applications work $O = (O_1, O_2, \ldots, O_m)$, where $O_m, m = \overline{1 \ldots M}$. Each task is characterized by:

- a set of connected services $Y = \{y_{m,j}\}$;
- subset $H_m^O = (H_{m,1}^O, H_{m,2}^O, \ldots, H_{m,k}^O)$ - set of parameters required to start the service, as the main of which are the characteristics that affect the quality of service (maximum execution time, response time, etc.).

In the composition of the multi-agent model of resource control each agent manager is a specific resource and is characterized by its parameters H_m^O, priority run P, the maximum execution time $T\max_j$, etc. The objective for each agent is to find resources that can perform required tasks with minimal effort. To do this, the agent need to calculate the local optimal schedule, taking into account the cost of resources, as a result of simulating the interactions between tasks and resources.

We propose the approach to solve the above problems. It is based on the management of cloud application assignment to appropriate resources for its work.

For understanding the operation principles of a cloud application, we need to define its place in the infrastructure of the virtual data center. Date center is a dynamic object, changing in time *t*, its state can be formalized as a directed graph of the following form:

$$VirtDC = (Node(t), Connect(t), CloudAppl(t)), \quad (2)$$

where vertices $Node(t) = \{Node_1, \ldots, Node_n\}$ are active elements of the virtual DC cloud infrastructure (computing nodes, storage systems and others); directed edges $Connect(t) = \{Connect_1, \ldots, Connect_v\}$ are active user connections to the virtualized applications; $CloudAppl(t) = \{CloudAppl_1, \ldots, CloudAppl_c\}$ are active instances of the cloud applications, launched on the virtual resources.

The major feature of cloud applications is the approach, in which users have access to them and to their services, and they do not know anything about their actual location. Most often, users know only the address of aggregation node and application name. The cloud system automatically selects the optimal virtual machine for the request, which will have been processed on it.

Before we talk about the resource allocation for the cloud applications, it is necessary to determine their structure, the basic parameters and the key characteristics of their operation, affecting the efficiency of their usage. For this purpose, we have developed a generalized model of cloud application.

The generalized model of cloud application is a multilayer structure formalized in a form of graphs, describing the connections of individual elements. The model can be represented in the form of three basic slices, detailing the connections of the specific objects of virtual cloud infrastructure: applications, related services and allocated resources.

The cloud application is a weighted directed acyclic graph of data dependencies:

$$CloudAppl = (G, V), \quad (3)$$

Its vertices G are tasks, which get information from the sources or process it in accordance with the users' requests;

directed edges V are dependencies of the tasks on the data sources between corresponding vertices.

Each vertex $g \in G$ is characterized by the following tuple:

$$g = (\mathrm{Re}\, s, NAppl, Utime, SchemeTask), \quad (4)$$

where *Res* are the resource requirements; *NAppl* is the number of application instances; *Utime* is the estimation of the users request execution time; *SchemeTask* are communication schemes of data transmission between sources and computing nodes.

Each directed edge $v \in V$ connects the application with the required data source. It is characterized by the following tuple:

$$v = (u, v, Tdata, Mdata, Fdata, Vdata, Qdata), \quad (5)$$

where *u* and *v* are linked vertices; *Tdata* is the type of transmitted data; *Mdata* – the access method to the information source (REST, JSON and others); *Fdata* – the physical type of accessed object (file in the storage system, local file, distributed database, data services and so on); *Vdata* – the traffic volume estimated on the accessed data (in Mb), the requirements for the quality of services.

The originality of the model is in the fact that for each application the consolidated assessment of its work with data sources is calculated. It allows to predict the performance of the whole cloud system.

As mentioned earlier, cloud service in one of the key slices in the generalized model of cloud application. Cloud service serves as an autonomous data source for the application, for which it acts as a consolidated data handler. Generally, cloud service is highly specialized and designed to perform a limited set of functions. The advantage of connecting cloud application to the service is in an isolated data processing, in contrast to the direct access to the raw data, when cloud application does not use a service. The usage of services reduces the execution times of user requests. Cloud service is formalized as a directed graph of data dependencies. The difference lies in the fact that from the user point of view cloud service is a closed system.

Cloud service can be formalized as a tuple:

$$CloudServ = (AgrIP, NameServ, Format), \quad (6)$$

where *AgrIP* is the address of aggregation computing node; *NameServ* – the service name; *Format* – the required format of output data.

The aggregator of a service selects optimal virtual machine, on which it is executed. In addition, all its applications are distributed between predefined virtual machines or physical servers. Their new instances are scaled dynamically, depending on the number of incoming requests from cloud applications, users or other services.

To describe the assignment of cloud applications and services in the virtual data center infrastructure, we have also implemented the model of cloud resource. Cloud resource represents an object of virtual data center, which describes the behavior and the characteristics of the individual infrastructure elements, depending on its current state and parameters. The objects of virtual data center are disk arrays, including detached storage devices, virtual machines, software-defined storages, databases of various kinds (SQL/NoSQL) and others. In addition, each cloud service or application imposes requirements on the number of computing cores, the RAM and disk sizes, the presence of special libraries on physical or virtual nodes, used to launch their executing environments.

Each cloud resource can be formalized as follows:

$$Cloud\,\mathrm{Re}\, s = (T\,\mathrm{Re}\, s, Param, State, Core, Rmem, Hmem, Lib), \quad (7)$$

where *TRes* is the type of resource; *Param* – the set of parameters; *State* – the state of resource; *Core* – the number of computing cores; *Rmem* – the size of RAM; *Hmem* – the size of disk; *Lib* – the libraries requirements.

The distinctive feature of the proposed model is the resource universality, which allows to explore it from the user's points of view (as a closed system) and from the point of view of virtual data center software-defined infrastructure (as an open system).

The novelty of the model is the simultaneous description of the data placement, the associated applications and the state of virtual environment, considering network topology.

We developed the model of the software-defined storage, which details the resource model of the virtual data center. It is represented in the form of a directed multigraph, its vertices are the virtual data center elements, which are responsible for applications' data placement (e.g. virtual disk arrays, DBMS and so on):

$$Stg_{ki} = \left(MaxV_{ki}, P_{ki}^{stg}, Vol_{ki}(t), \overline{R}_{ki}(t), \overline{W}_{ki}(t), s_{ki}^{stg}(t)\right), \quad (8)$$

where $MaxV_{ki} \in N$ is the maximum storage capacity in Mb; $P_{ki}^{stg} = \{p_{kij}^{stg}\}_j$ – the set of network ports; $Vol_{ki}(t) \in N \cup \{0\}$ – the available storage capacity in Mb; $\overline{R}_{ki}(t)$ and $\overline{W}_{ki}(t)$ are read and write speeds; $s_{ki}^{stg}(t) \in \{"online","offline"\}$ – its state.

II. Practical part

We prepare scheduler for software-defined storage (qStg). The qStg acts as a scheduler and does not delay packets; it is also useful for lowering latency when traffic does not need to be slowed. Each class acts as a priority queue, where class 1 has the highest priority and class 3 has the lowest priority. We use a filter based on the type flows of applications and services to determine the class to which a packet will be queued. A priority is assigned to the containers when they are created. Since each date storages has an unique IP address, a filter rule for that IP address is added to the qStg. All packets are checked for rules and then enqueued to a class based on the flow filter rule of the qStg. For example, all the packets from video

applications with high priority are enqueued to class 1. Similarly, the packets with other level with default and low priorities are added to class 2 and 3, respectively. The packets are enqueued to a class as they arrive, but they get dequeued in the ascending order of the class. The qStg scheduler checks for packets in the queue of class 1; if no packets are available to dequeue, the queue of class 2 is checked, and then similarly the queue of class 3 is checked. The dequeuing of packets from the queue of different classes enforces the scheduling policy and priorities to applications queries.

The competitive advantage of the developed control algorithm software-defined storage compared to existing analogues is heuristic analysis of new data types in the process of downloading files in the cloud platform. Thus through storage virtualization is performed transparently to the client mirroring data on multiple storage devices that provides an increase in speed of data allocation in terms of maintaining the integrity and redundancy. Formation of the self-organizing software-defined storage on the basis of virtual machines and containers can not only reduce the risks associated with the loss of inaccessibility of data, but also provides intelligent analysis of demand data which is formed on the basis of the card placement of virtual machines and containers. The algorithm is based on data placement in software-defined storage facilities on a model that allows describing the structure and connections of virtual devices, machines and containers of data. The model is based on multi-agent approach in the organization of storage. The agents collect system status. This information is analyzed with the use of machine learning algorithms (Data Mining). The output of the analysis obtains a map of location devices within the cloud platform tied to physical devices, as well as a map of the demand generated the data. By analyzing the two cards and the heuristic algorithm predicting cloud management system decides on reconfiguration or moving virtual storage devices as well as the rotation and redistribution of data between different nodes of the system. This map of the location of dynamic objects is generated not only by the extent of occurrence of the events download or reading data but a set interval of time.

The algorithm of placing cloud application in the software-defined storage is used for increasing the productivity of the system components of the cloud. Due to the efficient allocation of data streams between running instances of virtual machines and containers provided with not only the quality of service, but also a compact arrangement of the devices. To resolve this issue as one of the elements in the basis of the developed algorithm, the study placing cloud application in software-defined storage facilities used aggressive version of the algorithm Backfill used to optimize the performance of tasks in the Grid. The task of increasing the efficiency of use of available storage devices based on the data, had been received from the agents and management nodes cloud system through the dynamic resource management in conditions of limited consumption of the computing power, is relevant for the cloud. This is primarily due to the economic performance of cloud platforms.

In addition to using elements Backfill algorithm produced a series of improvements relating to direct data access mechanism to improve the efficiency of cloud optimization system. When dealing with services located in the cloud system does not exclude the situation in which the user's request for service may be employed multiple data stores with different access characteristics. It is necessary to prepare the access time to optimize the reading when using such data cloud system. To do this, the algorithm builds a series of internal rules, thereby adjusting to the flow of user requests every instance storage. As a result, query execution plans with the same intensity at different times may be distributed differently. Rebuilding of the rules takes place in accordance with the demand for resources to efficiently manage the distribution and dynamic load balancing.

III. Experimental part

For the experiment we have used the cloud system implemented at Orenburg State University. It includes 4 OpenFlow switches (2 HP 3500yl, 2 Netgear GSM7200), 8 computing nodes (32 GB RAM, 4 cores), 1 server (32 GB RAM, 8 cores) with OpenFlow controller and 1 server (32 GB RAM, 4 cores) for monitoring function. As a selected fat tree topology with three levels. Objects of network connected by speed 1000 Mbit / s.

We will study workload characteristics, requirement analysis, the theory of QoS in software-defined storages and I/O scheduling strategies. We obtain such goals by execution mechanisms and dynamic robust I/O scheduling algorithms for multi-type resources allocation. In the current progress, scheduling for software-defined storages has been proposed for the SSD/HDD hybrid storage. The preliminary evaluation in some benchmarks shows that software-defined storages can gain better performance compared with other strategies.

To evaluate the effectiveness of the algorithm data placement in software-defined storage facilities built in the context of models of software-defined infrastructure, we have conducted research work in the cloud system built on the basis Openstack with different parameters. Thus as reference data for comparison in the experiment standard algorithms are used in cloud systems as well as traditional storage systems. For the experimental study a prototype cloud environment, which includes the main components as well as software modules developed algorithms for modifying the processing of user requests data in a software-defined storages.

In the OpenStack cloud system, the module is implemented that applies an algorithm developed by placing the data in the software-defined storage facilities for the management of computing resources and cloud system efficient allocation of virtual machines to physical hosts, as well as related data. In an experiment designed to analyze the data stream of requests, similar to real traffic cloud infrastructure based on the data log records access to certain types of resources, classified by data type and structure of the query. Retrospective reproducible requests amounted to 3 years, while for load experiment used the averaged data. The data are distributed to a pool of virtual machines on the following criteria: the type of customers to

make an appeal to the data, the type of service demanded by the connection. The number of simultaneous requests to the system was 100,000, which corresponds to the maximum number of potential users of the system.

All created queries are reproduced consistently at three pilot sites. This restriction is due to the need to compare the results with the physical storage systems are not capable of reconfiguration. The main difference is the use of experimental sites SSDs. In addition to the platforms to analyze the effectiveness formed 3 groups of experiments aimed at intensive operation for reading (experiment 1), write (experiment 2) and concurrent read and write data (experiment 3).

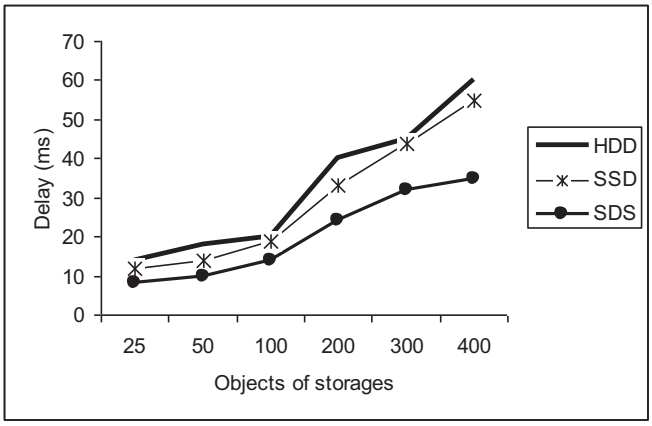


Fig. 1. Result of experiment 1 (read)

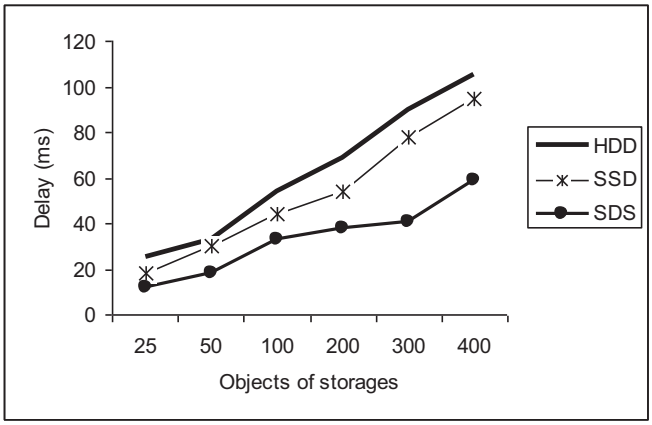


Fig. 2. Result of experiment 2 (write)

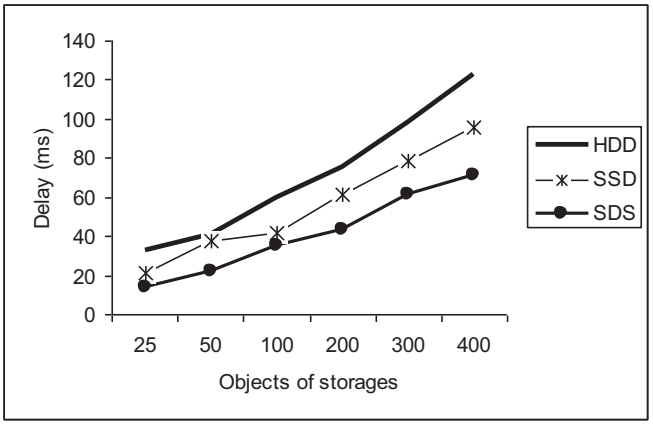


Fig. 3. Result of experiment 3 (read and write)

The experiment lasted for one hour corresponding to the longest period of time of peak load, recorded in real traffic. After analyzing the data of experimental studies we proved that the software-defined storage more efficient, regardless of the type of physical devices. The findings support the use of the algorithm and software-defined storage to provide efficient services to the cloud. The results of the experiments can be concluded to reduce by 20-25% the number of failures in service when placing the data in the software and message-driven data warehouses. In addition, in a pilot study assessed the amount of storage used for experimental platforms 2 and 3, as for the physical storage resources do not support scaling in real time. Due to the optimal allocation of resources for each compute node is guaranteed to work together to ensure all running instances of the application that meets the requirements of potential users. At the same time thanks to the work of the algorithm data placement in program-controlled vaults of opportunity of the release of 20 to 30% of the resources allocated computing nodes. Thus, the proposed algorithm can be used for any computing system architectures including inhomogeneous physical node configuration and VMs.

IV. Conclusion

Thus, assessing the overall result of work the algorithm of of placing cloud application in software-defined storage can gain performance from 20 to 25% compared with the physical storages and virtual machines. This may be necessary when storage systems work with high intensities requests. In addition, reducing of the number of allocated virtual resources allows for more efficient scale cloud systems, and provision of a safety margin with a sharp increase in the intensity of use of the selected applications.

References

[1] Bein D., Bein W., Venigella S. “Cloud Storage and Online Bin Packing” Proc. of the 5th Intern. Symp. on Intelligent Distributed Computing, 2011, Delft: IDC, P. 63-68.

[2] Nagendram S., Lakshmi J.V., Rao D.V., et al “Efficient Resource Scheduling in Data Centers using MRIS” Indian J. of Computer Science and Engineering, 2011, V. 2. Issue 5, P. 764-769.

[3] Arzuaga E., Kaeli D.R. “Quantifying load imbalance on virtualized enterprise servers” Proc. of the first joint WOSP/SIPEW international conference on Performance engineering, 2010, San Josa, CA: ACM, P.235-242.

[4] Mishra M., Sahoo A. “On theory of VM placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach” Cloud Computing (CLOUD), IEEE International Conference, 2011, Washington: IEEE Press, P.275-282.

[5] Korupolu M., Singh A., Bamba B. “Coupled placement in modern Data Centers” IEEE Intern. Symp. on Parallel & Distributed Processing. N. Y.: IPDPS, 2009. P.1-12.

[6] Singh A., Korupolu M., Mohapatra D. “Server-storage virtualization: integration and load balancing in Data Centers” Proc. of the 2008 ACM/IEEE Conf. on Supercomputing. Austin: IEEE Press, 2008. P.1- 12.

[7] Plakunov A., Kostenko V. “Data center resource mapping algorithm based on the ant colony optimization” Proc. of Science and Technology Conference (Modern Networking Technologies) (MoNeTeC), Moscow: IEEE Press, 2014. P.1- 6.

[8] Darabseh, A., Al-Ayyoub, M., Jararweh, Y., Benkhelifa, E., Vouk, M., Rindos, A. “SDStorage: A Software Defined Storage Experimental Framework” Proc. of Cloud Engineering (IC2E), Tempe: IEEE Press, 2015. p.341- 346.

[9] Parfenov D., Bolodurina I. “Approaches to the effective use of limited computing resources in multimedia applications in the educational institutions” WCSE 2015-IPCE, 2015.

Big Data Integration Architectural Concepts for Oil and Gas Industry

Rasim M. Alguliyev[1], Ramiz M. Aliguliyev[2], Makrufa S. Hajirahimova[3]
[1,.2,.3]Institute of Information Technology of ANAS
Baku, Azerbaijan
[1]rasim@science.az, [2]r.aliguliyev@gmail.com, [3]makrufa@science.az

***Abstract*— In recent years economic an social activity fields is based on data. Oil and gas industry leaders understand the value of big data and are interested in digital oil industry becoming a reality. Here is big data is analysed as a key component in based decision making in oil and gas industry during exploration, drilling and production. In oil and gas industry architectural model is offered for integration and convergence of big data, business analyticss and transaction data.**

***Key words*—oil and gas industry, big data, business analyticss, business intelligence, predictive analyticss, data mining, decision-making.**

I. INTRODUCTION

Better learning desire of the earth's subsurface encourage the oil and gas companies to gather more quickly, different types of data, more information for informed decision-making.

In oil and gas industry exploration, during the drilling, extraction and processing huge amount of data is generated, collected. Thus, oil an gas company always collect data from underground and surface objects (oil wells, etc.) by thousands of sensors (for example, IT traffic of Chevron passess 1.5 terabyte per a day), in the real-time mode the carry out control assets and environmental monitoring [1-3], 2D, 3D, 4D visualization, advanced geophysical modeling, simulation for supporting operations during seismic studies in the exploration phase.

Naturally, there is valuable information that it is clearly not known in these data and they are serious knowledge in based decision making. But data collection is so great and complex and that's why it is impossible to process them with traditional methods. For improve of their business strategies oil and gas companies are looking for ways of maximum using from big data. In this respect, big data and big data analytics is more important for the oil and gas industry [3-6]. Because data analysis is an invaluable tool in planning and decision making during the exploration and development of oil and hydrocarbon.

On the other hand, collected information becomes different formats according to source, as well as type (structured and non-structured). The different data formats like DLIS, LIS, SEGx,Videos, Docs, PDFs, XLS, CSV, graphics, OGP Px, XML, RDBMS etc are present. The intelligent wells or intelligent producing fileds like real time drilling data continuously feed our systems along with senor or telemetric data at many places. This in turn creates a serious problem in the use of them together. Because currently it is impossible to obtain necessary information when needed for business-analytics from big data based on available managing platform of exploration and production data. That's why there are need new approaches for maximum benefit, create value from data that collected various formats from different sources. The main problem is the integration of data which gathered through different platforms. In this case, architectural model is proposed that ensured the integration of big data, businnes analytics and intellectual analysis technologies in oil and gas industry in order to solve the problem.

II. BIG-DATA SOURCES, CHARACTERISTICS, OPPORTUNITIES AND PROBLEMS IN OIL-GAS FIELD

A. Sources

Oil and gas branch is an extremely competitive environment at all times. However, oil and gas companies face a number of challenges: limited natural energy resources, demand, frequent changes prices of the oil and gas, environmental risks, competition from emerging new energy sources, etc. According to experts the successful resolution of these and other problems industry dependents on the effective use of data collected from a large volume [7,8].

Generally, there are many sources of data in the oil and gas industry [9], and they can be grouped into two classes:

- Company data and operational data;
- Data of monitoring and management of technological processes.

Over the years, collected and well-structured data from oil and gas companies operational systems, such as CRM, ERP, warehouse management systems, data marts and etc. These data are formatted and cleaned the purpose of creation reports and analyzes.

Collected data in real-time from oil wells, during geophysical modeling and geological, seismic studies and drilling and production thousands sensors and etc. in the exploration phase refer to the second group of data sources. This group includes also the historical data collected during geological and geophysical studies, delivery, pricing and etc.

Demographic data, environmental and climate data and external data feeds (for example, social media) as a source of data.

Of course, in oil and gas industry is very important all data collected from all sources for creation of new business strategies and correct decisions.

B. Characteristics

Collected data is characterized by "5V" and sometimes "6V" during the exploration and production of oil and gas deposits [5, 8]:

Volume, exploration, drilling, production, seismic and other data;

Variety, structured (standard models of data as (PPDM, SEG-Y, WITSML, PRODML, RESML); unstructured (picture, audio, videos and etc.); partially structured (*processed data during analysis and interpretation, daily reports on drilling and etc.*)

Velocity, data stream incoming from drilling rigs (EDR, LWD, MWD and so on.) and sensors (Flow, Pressure, ROP, etc.) during real-time.

Veracity– Data quality can vary greatly and this variability may affect the results of the analysis.

- improving the quality of data;
- implementation of data integration models;
- seismic, drilling and production data harmonization;
- initial processing for identifying anomalies in data.

It should be noted that the quality of the data of particular importance for data integration. Does not qualitative data should be cleaned before the conversion and integration.

Value: Data also have capacity to create value.

- acceleration of initial oil extraction;
- increasing production;
- cost reduction;
- minimize the risks.

C. Opportunities and challenges

Overall, application of big data analytics in all stages (exploration, production, processing, retail sales) of oil-gas extraction is possible [10]. Main goal consists of following [11,12]:

- to find more oil and gas deposits with great velocity and accuracy;
- minimum damage to the environment by providing safe, efficient and improve oil extraction;
- to refine and distribute the natural energy resources with the most effective and most affordable value;
- to manage relationships with customers, find new profitable markets and opportunities;
- mitigation of the risks especially in safety, ecology and so on. Field.

Oil-gas companies obtain the following opportunities by analysis of data [10,13,14]:

- Maintance of competition ability during planning, exploration, developent and production phases of oil-gas deposits;
- Increasing an efficiency of production in connection with technical service and forecasting;
- Reducing of initial oil extraction, operating costs and increasing of productivity at all stages;
- Providing automated accessibility of necessary information in necessary time.

In other words, companies can reduce production costs, improve decision making,increase the efficiency of operations and business processes by analysis of big data, can obtained new ideas, work new business models by the convergence of different types of data.

Big data provides an opportunity to optimize the stage of geological exploration and drilling, as well as mining and manufacturing, insure from risks. During the investigation of new resources the using of big data and advanced analytics find the seismic traces and combined them with previous forgotten seismic results. It gives very important results.

By reference to historical data associated with drilling and production review of the assumptions and judgments during the new researches could be great help to the researches in the areas that limited with the environmental protection rules. For example, information such as climate change and ice streams may allow analysts to study the effects of storms with processes in drilling [14-16].

In [17] are shown obtaining ways of greatest benefit by avoiding from typical mistakes when improper use of big data and analytics.

The essential directions of analytical works contain in oil-gas exploration: determining of seismic traces; acceleration of exploration; the creation of new scientific models; improving of the search quality; assessment of field and evaluation of perspectives and etc.

In real-time during drilling operations can help to make right decision based on determining of potential problems with the help of predictive analytics, statistical models, heuristic judgments for forecasts [18].

Previously know the structure and anomaly of soil means to save millions in terms of labor and provision in drilling. The role of large amounts of clarifying data is great in reduce the risks, decision-making during drilling the wells if we consider high costs in drilling and in some cases resulted failure in oil industry

Issues that can improve drilling and completion operations of big data and analytics are following:

- *Construction and evaluation of drilling models:* it will help update the models based on big data from sensors that placed on the drilling rig and optimize drilling parameters;

- *Increasing drilling accuracy and safety*: early detection of anomalies which affect drilling and prevention of undesirable events such as blow, blasting;

- *Optimization of drilling process:* early detection of negative factors that affect to drilling operations;

- *Optimization of financial assets*- determine the optimum value by using computing technologies;

- *Decision-making in real-time:* drilling information and its assessment should be carried out in real-time and this information should be used in decision-making, creation of new scientific models and etc.

Oil-gas industry faced to 3 main problems in exploration and production (processing) stage:

- *Data management.* It combines the collection, processing and storageof large volume of structured and unstructured big data that allows analysis and efficient search of data by using of the analytical and statistical methods.

- *Quantification of data*- it combines application of statistical and analytical methods for determination of issuance of forecasts and more accurately identification ideas that have prediction importance.

- *Risk assessment*- it is prognosis analysis which are performed by mathematical models of collected known risky data for how to work with unknown risks.

It should be noted that the of giants leading information technology, such as IBM, Oracle, Microsoft, Teradata, etc. exists's software products for big data and analysis in the oil and gas industry [2,10,19-21].

III. BIG DATA INTEGRATION ARCHITECTURAL CONCEPTS FOR OIL AND GAS INDUSTRY

For the implementation concept of big data in the oil and gas industry requires the integration of heterogeneous data. Because data can be more valuable if it combined or related with other data.

Generally, data integration is combining of data which placed in various sources and presentation to the user in unification form. That is, data integration rather, a set of technologies to create the most appropriate representation of data for to grasp the connection between separate components of of this set. It is possible to provide data integration in physical (bring data to a single format from different sources), logical (is intended acessibility possibility to data in different sources according to a single global scheme) and semantic (semantic characteristics of predmet fields are intended) levels. The key issues of integration are following: architecture of integration system; the development of user interface, the development of integration model that supported data from different sources, the integration of metadata and etc.

It should be noted that the ETL (Extract, Transform, and Load), ELT (Extract, Load and Transform), EAI (Enterprise application integration, CDC (Change Data Capture), database replication and etc. are traditional methods of data integration.. Some of them are oriented to a batch mode, and some in the real-time [9,22].

Big data integration (BDI) solution is markedly different from the traditional tasks designed to support ETL, such as virtualization and data storage. Despite all this, the BDI is different from the existing methods of integration mainly features as a large volume, high speed, diversity and etc [23]. The distinction, in the context of the oil and gas field it can be summarized as follows.

The first difference is determined by a very large number of data sources in the industry of oil and gas. For example, data from hundreds and thousands of sensors during exploration, drilling, production, transportation and processing of oil and gas. This also includes demographic data, digitized historical data, enterprise data from information systems and other.

The second difference is characterized by different types (structured, unstructured, semi-structured) and data formats (XLS, Docs, video, graphics, UCP Px, XML). Since heterogeneous data sources (seismic, sensors during exploration, drilling, production, transportation and processing of petroleum and gas), and when collecting, processing and management of data using various automated software platforms.

The next difference is formulated as follows. The main data sources are very dynamic and continuously evolving. Suffice it to note that oil and gas companies in real time with thousands of sensors continuously collect data from underground and surface facilities. Since a huge amount of data is continuously collected newly available and can be analyzed at any time.

Another difference in the oil and gas industry is the quality of the data. This distinction reflects the quality and reliability of the data source. The low quality of the data, especially unstructured (graphics, video, test, etc.), is one of the most serious problems. Several ensure data reliability problems are specific to the oil and gas industry. The reliability of the data may depend on the quality of the oil and gas appliances and sensors. All these factors influence the results of the analysis of large data. And it may end in the acceptance of wrong decisions.

The solution of this problem, ie the integration big data can be divided into several steps. Although these tasks ourselves are complicated, the basic principles integration remain the same. The first stage of any integration begins with the establishment of the data schema (schemamapping), which

can be a single-level or multi-level with division global and local data. The next procedure (recordrlinkage) involves the allocation entries close on logic, but derived from different sources. Unlike the classical integration in BDI in addition to the logical associations must take into attention dynamism of the sources and the necessary procedures for complex event processing and filtering. Since various noises create additional complexity.

From previous 2 stages *data fusion* is a new procedure and did not need in traditional integration systems. Its main purpose is eliminates conflicts between data and select the appropriate one.

In IDC's study it is noted that the application of big data technologies is in experimental stage in oil and gas industry. Now, big data analytics solutions for oil-gas industry are created by the companies as Microsoft, IBM (InfoSphere BigInsights), Oracle, Hitachi (Hitachi Data Systems), Amazon and etc. In particular, it should be emphasized project of data integration (Oracle Data Integrator - ODI) for real-time implemented by Oracle company. ODI provides a data integration between Oracle Databases and IBM DB2, Microsoft SQL Server, MySQL and other bases.

Peak Indicators offers a wide range of data integration services for relational, unstructured and cloud data sources.

iGATE Corporation plans to establish a system named Oil Well Log Analytics. This system services to integrate oil exploration and production data in a single connected platform.

The first attempts to implement BDI taken Talend products companies (Open Studio for Big Data), Syncsort (DMExpress), Pentaho (Data Integration) and several others.

Despite that IT markets offer a wide range of tools for the integration of structured and unstructured big data sources in oil and gas industry there are open problems.

For solving the problem of the proposed architectural model is presented in figure 1.

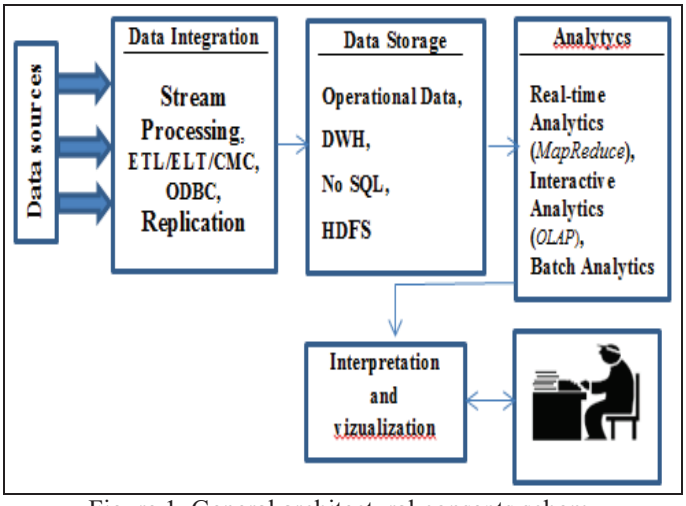


Figure 1. General architectural consepts schem

As seen from the scheme, the concept include of blocks integration, storage, analytics and visualization .In the concept of provided for different levels (*Data Accessibility, Data Platform, Consolidated Data*) of integration.

In [24] outlines the main modern architectural platforms, such as OLTP (transaction processing), OLAP (Online Analytical Processing), Data Warehouse (DW), Massively Parallel Processing (MPP) databases, distributed parallel file systems, Hadoop, NoSQL, Hadoop Distributed File System (HDFS), MapReduce, key-value stores. In the proposed concept is meant the use of these technology platforms.

IV. Conclusion

There is no doubt that the is currently oil and natural gas is the main source of energy and amounts 30% of the world energy.

Oil-gas industry sector rapidly adapt to the information era has entered to digital turnaround. During this period oil and gas companies that gatehered information with terabytes and sometimes petabytes per a day need new analytical tools to get ideas in data collection and interpret them accurately for predict potentials, keep costs to a minimum and so on.

İn analytics of data the main issue is the integration of data that incoming from different sources and in different formats (structured and unstructured). The application of proposed integration architecture is possible in the fields oil and gas industry such as exploration, drilling, processing and so on.

References

[1] J.Feblowitz, “The Big Deal About Big Data in Upstream Oil and Gas, Paper & presentation”, IDC Energy Insights, October 2012

[2] Unified Data Solution for Oil & Gas, http://www.hrgresearch.com/

[3] M.Brulé., 2013, Tapping the power of Big Data for the oil and gas industry, IBM Software White Paper for Petroleum Industry, May 2013.

[4] Big Data In Big Oil: How Shell Uses Analytics To Drive Business Success http://www.forbes.com/sites/teradata/2015/

[5] H.Murtaza, “Beyond the hype: Big data concepts, methods, and analytics”, International Journal of Information Management, 2015, vol. 35, pp. 137–144.

[6] S.Amir, B. Atle, “Knowledge Based Oil and Gas Industry”, Knowledge Based Oil and Gas, 2011.

[7] http://oslo.spe.org/bigdata

[8] M.S. Hajirahimova, Opportunities and challenges Big Data in oil and gas industry, National Supercomputer Forum (NSKF), 2015, Russia, Pereslavl

[9] Improving Oil & Gas Performance with Big Data. http://www.oracle.com/us/technologies/big-data/big-data-oil-gas-2515144.pdf

[10] A. Hems, A. Soofi., E. Perez., “Drilling for New Business Value: How innovative oil and gas companies are using Big Data to out maneuver the competition”, A Microsoft White Paper, May 2013

[11] X. Wu, X. Zhu, G.Q. Wu, W. Ding, “Data mining with bigdata”, IEEE Transactionson Knowledge and Data Engineering, 2014, vol.26, issue 1, pp.97–107.

[12] A. Baaziz., L.Quoniam, “The information for the operational risk management in uncertain environments: Case of Early Kick

Detection while drilling of the oil or gas wells", International Journal of Innovation and Applied Studies (IJIAS), 2013, vol. 4, no. 1.

[13] M. Seshadri, Big Data Science Challenging The Oil Industry, CTO Global Services, EMC Corporation, March 2013.

[14] Exploring Oil Data: A Reporter's Handbook, http://openoil.net/exploring-oil-data/

[15] A. Steed, D.M. Ricciuto, G. Shipman etc., "Big data visualanalytics for exploratory earth system simulation analysis Chad", Computers &Geosciences, 2013,vol.61, pp.71–82.

[16] R. Bertocco, V,Padmanabhan ,. Big Data analytics in oil and gas: Converting the promise into value,. http://www.bain.com/

[17] O.E.Gundersen, L.J. Boe, Successful implementation of Big Data and Advanced analytics in O&G, (Predictive Real-Time Analytics Drives Decision Support for Oil Well Drilling.

[18] O. E. Evensen, Big Data in Oil & Gas for Improved Decision Making and Operational Efficiency, IBM.

[19] R. K.Perrons, H. Adam, "Cloud computing in the upstream oil & gas industry: A proposed way forward", Energy Policy, 2013, vol. 56, pp.732–737.

[20] J. Hollingsworth , Big Data For Oil & Gas, Oracle Oil & Gas Industry Business Unit, March 2013

[21] Panorama des solutions de BIG DATA: Hadoop et son écosystème, Journal du Net, http://www.journaldunet.com/, 2013.

[22] Krishman K. Data Warehousing in the Age of Big Data,2013, p.371.

[23] X. L .Dong, D. Sirvastava, "Big Data integration", Data Engineering (ICDE), 2013 IEEE 29th International Conference on, 1245-1248

[24] E. Begoli. A Short Survey on the State of the Art in Architectures and Platforms for Large Scale Data Analysis and Knowledge Discovery from Data, 2012, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.8676

Parallelizing Shortest Path Algorithm for Time Dependent Graphs with Flow Speed Model

Mehmet Akif Ersoy
Dept. of Computer Engineering
Bogazici University
Istanbul, Turkey
Email: mehmet.ersoy@boun.edu.tr

Can Özturan
Dept. of Computer Engineering
Bogazici University
Istanbul, Turkey
Email: ozturaca@boun.edu.tr

***Abstract*—Various sequential algorithms for the shortest path problem on time dependent graphs are appearing in the literature. However, these algorithms mostly suffer from long running times and huge memory requirements. These problems are making them unsuitable for navigation applications which need to run on real time data with fast response times. For the shortest path problem with time dependent flow speed model, we propose parallel algorithms based on Modified Dijkstra algorithm in order to speed-up the running time of the sequential algorithm without requiring much more memory. We develop three different parallel implementations by using Cuda and OpenMP: These are (i) a Cuda based version, (ii) an OpenMP based version and (iii) a hybrid Cuda and OpenMP based version. We get up to 10-fold speedup in the OpenMP version, and 17-fold speed up in the other two versions.**

***Index Terms*—parallel, time dependent, shortest path**

I. Introduction

In these days, state of the art navigation programs mostly benefit from (i) real-time data gathered through some devices installed on cars and trucks by companies that have an agreement with them, (ii) from camera data placed on mostly used route paths, (iii) from satellite data [1]. Navigation systems have to take into consideration the other cars in their routes and traffic densities in areas where they travel. Thus, efficient solutions to the shortest path problem on time dependent graphs are required for navigation applications.

A *time dependent graph (TDG)* is a graph in whose structure and parameters may change over time. Some of the previous works treat TDGs as graphs where dynamic node/edge additions/deletions are allowed. Some other papers consider TDGs as graphs where the edge-delay, which is the total time spent to get from start of the edge to end of the edge in a given time, changes over time [2]–[7]. In this work, we take the model proposed by Sung et al. [8], where the speed of a traveling vehicle changes in specified time intervals. This approach is more realistic when considering cars traveling on urban roads.

A commonly satisfied property in time dependent graphs is called the *FIFO (First-In-First-Out) property* which is also named as the *non-passing property* in the literature. This property states that if the same route from node i to node j is traveled then the one that leaves earlier from node i, arrives earlier at node j than the one that leaves later. This property is a realistic assumption that we also see in daily traffic. The FIFO property enables us to develop efficient and polynomial algorithms for time dependent shortest path problem (TDSPP) [9]. TDSPP is an NP-hard problem on TDG without the FIFO [10] property. In this work, we consider TDG in which the FIFO property holds.

II. Previous Work

One of the most studied problem on time dependent graphs is the shortest path problem. In this section, we review the previous work that appears in the literature. The problem that is mostly studied on TDGs involve graphs in which lengths of edges are time dependent.

A. Shortest Path Algorithms on TDG with Varying Edge Lengths

One of the proposed methods for solving TDSPP is to use time expanded graphs. This method converts a TDG to a static graph. It creates a copy of each element of TDG for each time instance. When the number of time intervals increase, this method needs to create a huge graph and the algorithm run time can become non-polynomial [5]. Time expansion method is often used where the FIFO property does not hold.

Some research has been done in order to speed up the TDSPP by exploiting preprocessing of the graph before running the algorithm. Goldberg et al. [11], propose landmark-based ALT algorithm on static graphs. Ohshima et al. [5], present a modified version of this algorithm which runs on dynamic graphs with FIFO property. With the help of some preprocessing, they achieve up to 4 times speed up with respect to the Modified Dijkstra algorithm which was first proposed by Dreyfus [4] in order to solve shortest path problem on TDGs.

Ding et al. [2] and Kanoulas et al. [6] study shortest path problem on time dependent graphs for a given starting interval from the source node. Kanoulas et al. proposes an algorithm which is an extension of the A* algorithm and Ding et al. proposes a Dijkstra based algorithm for FIFO and non-FIFO graphs. They both examine the least total travel time from a source node v_s to a destination node v_d with the departure time from v_s given as a time interval input. Chabini et al. [7], [12] proposes Decreasing Order of Time (DOT) algorithm for all-to-one fastest paths problem for all departure times and prove that it has the optimal worst-case running time

complexity for this problem. They assume the edge lengths are time dependent variables but will be static after a finite number of time intervals. The running complexity of DOT algorithm is $\mathcal{O}(SPP + |V|M + |E|M)$ [7], where SPP is the static shortest path problem, $|V|$ is the number of vertices and $|E|$ is the number of edges in the graph.

B. Parallel Algorithms

A lot of work appear on the parallelization of algorithms for shortest path problem on graphs which are not time dependent [13]. There are also some works that involve parallel programming of TDSPP. These works consider the TDG model in which the edge-lengths are time dependent [14]–[16]. Chabini and Ganugapati [14] propose parallel algorithms based on the DOT algorithm given by [7], [12]. They implement distributed and shared memory parallel implementations that achieve approximately 4 fold speedup with 6 processors on input graphs with 1000 nodes, 3000 edges and 100 time intervals [14]. Tremblay et al. [15] also propose parallelization works for the DOT algorithm and the time dependent least time path algorithm given by [17]. They achieve up to 13 fold speedup on 15 processors [15]. Also, Ziliaskopoulos et al. [16] examine parallel designs for the algorithm given in [17]. They improve the algorithm run time by achieving two fold speedup on shared memory implementation with 4 CPUs, and achieve 7 fold speedup with message passing algorithm using parallel virtual machine (PVM). In theses parallel algorithms, they consider the cases where the time varying travel times for edges will be constant after a known time interval.

C. Flow-Speed Model

Flow-Speed Model is introduced by Sung et al. in [8]. The main idea that distinguishes this model from the other models used on TDGs is that the flow speed on each edge depends on the time interval instead of time dependent edge lengths that is considered in most of the TDG problems in previous works.

Let $G = (V, E)$ be a graph, where V is set of vertices and E is set of edges and l_{ij} be the non-negative length of an edge from vertex i to vertex j. Consider dividing the time horizon into K time intervals in which an interval represented by $[f_k, f_{k+1})$, where $k = 0, 1, ..., K-1$ and $0 \leq f_0 < f_1 < f_2 < ... < f_{k-2} < f_{k-1}$. Note that implying $t \in [f_x, f_{x+1})$ means that $f_x \leq t < f_{x+1}$. Let $v_{k_{(i,j)}}$ be the non-negative flow speed on the edge (i, j) in the time interval $[f_k, f_{k+1})$. For each vertex $j \in V$, define a value $t_j = min_{i \neq j}\{T(t_i, (i, j))\}$, where $T(t_i, (i, j))$ is the travel time from vertex i to j starting at time t_i. Then, the purpose of this model becomes finding values of t_j for all vertices $j \in V$, where the starting time from the starting node given as t_s.

D. Modified Dijkstra Algorithm

The algorithm that we work on is similar to Dijkstra based algorithm which was first proposed by Dreyfus et al. in [4] for edge-length varying TDGs. Sung et al. presents Modified Dijkstra (MD) algorithm (Figure 1) for solving problems using the flow-speed model on time dependent graphs. They take the complexity of Dijkstra's algorithm as $\mathcal{O}(|V|^2 + |E|)$, where $|V|$ is the number of vertices and $|E|$ is the number of edges in graph. MD algorithm's run time complexity is $\mathcal{O}(|V|^2 + |E|K)$ [8], where K is the maximum number of time intervals scanned in ArrivalTime function that is presented in Figure 2.

There are two differences between MD algorithm and Dijkstra's algorithm. In MD algorithm, there could be an update for each neighbour of that node without taking into consideration whether that neighbour is visited yet or not. On the other hand, Dijkstra's algorithm only updates unvisited neighbours. The second difference between these two algorithms is calculating the time passed to get from the node which has the smallest arrival time amongst the unvisited nodes to a neighbour of that node. This difference is because of the fact that MD algorithm deals with TDGs whereas Dijkstra's algorithm solves the shortest path problem for static graphs. *ArrivalTime* function in the MD algorithm calculates the arrival time to a neighbour by considering the current time.

Algorithm Modified Dijkstra

1: $S \leftarrow \emptyset; U \leftarrow V$
2: $t_i = \infty$ for each vertex $i \in V$
3: $t_s = 0$ and $pred(s) = 0$
4: **while** $|S| < n$ **do**
5: let $i \in U$ be a vertex for which $t_i = \min\{t_j : j \in U\}$
6: $S \leftarrow S \cup \{i\}$
7: $U \leftarrow U - \{i\}$
8: **for** each$(i, j) \in A(i)$ **do**
9: **if** $t_j > ArrivalTime(t_i, (i, j))$ **then**
10: $t_j = ArrivalTime(t_i, (i, j))$ and $pred(j) \leftarrow i$
11: **end if**
12: **end for**
13: **end while**

Fig. 1: Modified Dijkstra Algorithm [8].

III. Parallel Algorithms

The major challenge of the shortest path problem on TDG is that the solution requires a lot of time and memory. To address this challenge, we resort to parallelization of the algorithm for multi-core and GPU systems. Parallel programming is carried out using OpenMP and Cuda. In this work, MD algorithm is taken as a sequential solution to the shortest path problem on TDG and parallelized.

A. Time Consuming Parts of Sequential Modified Dijkstra Algorithm

We use a CPU profiler program called `Very Sleepy` [18] to analyze the sequential implementation of the MD algorithm. In the MD algorithm, we observe that there are mainly two parts of the algorithm which consumes most of the run time of the program and which can be parallelized without causing too

ArrivalTime$(t_i, (i,j))$ **Function**

1: $Arclength \leftarrow l_{ij}$
2: let $k \in \{0, 1, 2, ..., K\}$ be an index for which $f_{k_{(i,j)}} \leq t_i < f_{k+1_{(i,j)}}$
3: $Arclength \leftarrow Arclength - v_{k_{(i,j)}} \times (f_{k+1_{(i,j)}} - t_i)$
4: **while** $Arclength > 0$ **do**
5: $k \leftarrow k + 1$
6: $Arclength \leftarrow Arclength - v_{k_{(i,j)}} \times (f_{k+1_{(i,j)}} - f_{k_{(i,j)}})$
7: **end while**
8: $ArrivalTime(t_i, (i,j)) \leftarrow f_{k+1_{(i,j)}} + Arclength/v_{k_{(i,j)}}$
9: **return**

Fig. 2: Function for Calculating Arrival Time [8]

much overhead. The analysis results show that the most time consuming part of this algorithm is finding an unvisited node with the smallest arrival time. This is shown on the 5th line of the MD algorithm pseudo code given in Figure 1. Another time consuming part is where we update minimum arrival times of every neighbour of the node with the smallest arrival time amongst the unvisited nodes and calculate the arrival times of these neighbours while updating. This part is the for loop on lines 8-12 in Figure 1.

B. OpenMP Based Implementation

In our OpenMP implementation, we parallelize (i) the finding of the unvisited node with minimum distance and (ii) the updating of minimum arrival times of the neighbours of that node. When we update the minimum arrival time for a node, we compare the current arrival time of the updated node with the new calculated arrival time. New arrival time is the sum of current time and the time passed to get from current minimum distanced node to neighbour node. Therefore, updating of the arrival times of each neighbour are independent from each other. Hence, we can safely parallelize this part without considering any race condition between threads.

Finding of the minimum distanced unvisited node is done by using the parallel reduction operator in OpenMP. After the reduction, we apply a critical region where we only compare the arrival times gathered by each processor. This critical region does not introduce a serious overhead.

C. Cuda/Thrust Based Implementation on a GPU

In our Cuda implementation also, we try to minimize the time where we find the element with minimum distance from among all the unvisited elements (i.e. line 5 in Modified Dijkstra's Algorithm in Figure 1). This is where the most of the running time is consumed, and hence, it is appropriate to run it in parallel on a GPU. We used $thrust :: min\_element$ function from the Cuda/Thrust library to find the element with minimum distance in a $thrust :: device\_vector < Node >$. We need to define a binary predicate that is used for comparison by Thrust. Here, we consider not only the minimum distance criterion for the element but also whether the element has been visited or not.

D. OpenMP and Cuda Hybrid Implementation

We also provide a hybrid implementation of shortest path algorithm based on both Cuda and OpenMP. In this version, we store the graph on host as well on GPU device because one is used from OpenMP and the other one is used from Cuda functions. We use Cuda for finding the unvisited element with minimum distance. We use $thrust :: min\_element$ function which is defined in thrust/extrema class in a similar fashion as we use it in our Cuda implementation. On the other hand, this hybrid implementation differs from the Cuda version in that, it uses the returned result to get the position of the element in our host $nodes$ vector. Then, it uses OpenMP to update the arrival times of the neighbours of the current node. Finally, it updates the GPU device $nodesD$ vector by only updating the changed members of the host $nodes$ vector. Note that, the update of the Cuda vector $nodesD$ is not in the OpenMP parallel code region. Hence there are no race conditions among the threads while the Cuda vector is updated.

Hybrid algorithm runs faster than the Cuda version because it uses OpenMP parallelization in the process of updating neighbors of the minimum distanced unvisited node. This is a small improvement when we consider very large graphs, since most of the running time is spent on the finding minimum distanced element. As will be shown in the next experiments and results section, this algorithm runs faster than the OpenMP implementation too. However, the hybrid version has a drawback ; it needs to store the graph data both in $std :: vector$ on the host as well as in $thrust :: device\_vector$ on the GPU. Therefore, it requires twice as much memory than the two versions.

IV. Experiments and Results

Our goal is to improve the running time performance of the MD algorithm proposed by by Sung et al. in [8]. Performance is the main issue of TDSPP algorithms, so we try to reduce algorithm run time using parallel processing. We compare our parallel implementations with our implementation of the sequential MD algorithm. We give metrics that we use to evaluate performance our parallel algorithms. We provide information on the tests and the hardware environment on which we carry out our tests. After that, we show the results obtained from our tests with respect to the performance metrics.

A. Performance Metrics and the Test Environment

In this section, we describe performance metrics, and test environment and input parameters we use in our tests. We illustrate the performance of our parallel implementations using the following metrics:

- Running time of the three implementations,
- Speedup of parallel implementations over the sequential implementation,
- Efficiency of our OpenMP implementation,
- Scalability of our OpenMP implementation.

Our tests are carried out on a system with 12-Core, 2.80GHz Intel(R) Xeon(R) CPU X5660 with 8 GB Memory and an NVIDIA GF110 [GeForce GTX 580] GPGPU accelerator. In the result figures below, unless it is stated otherwise, OpenMP and hybrid implementations are tested using all of the 12 processors of the computer. We do not turn on optimization parameters while compiling and running our implementations.

B. Input Graph Generation

For our test cases, we use Graph500 reference code to generate random graphs. This code generates a graph with vertices and edges between them. We provide two inputs to the graph generation algorithm:

1) The number of vertices to be generated in base 2 logarithm : We create test graphs by giving this input in the range 10 to 22 (i.e. number of vertices in the range 2^{10} to 2^{22}).
2) Maximum number of edges to be generated: We provide this input as 4,8,16,32,64 times the number of vertices to be generated.

After creating graphs with the Graph500 program, we use the DFS algorithm to find reachable nodes from the first node. After that, we remove all the unreachable nodes from graph before running any algorithm. 46 graphs $(G1, .., G46)$ were generated in this way. Graphs $G1, \ldots, G16$ were very small graphs that required less than 1 second running time. Therefore, we do consider these in rest of the discussion. The information about the generated graphs $(G17, .., G46)$ after removing unreachable nodes can be found in Table I.

We set random integer weights (edge lengths) for each edge in the range 1 to 10 units using the C $rand()$ function. To be able to test time dependent graphs, we generate random integer velocities in the range 1 to 4 (units per time interval) on edges so that the velocity of a car going through an edge changes in every time interval. This interval can be set as well, but in our tests we set the time interval to 1.

C. Results

In this section, we give the results of our tests for the metrics of run time, speedup, efficiency and scalability. In Table I, $|V_R|$ means the number of reachable nodes and $|E_R|$ means the remaining number of edges after unreachable nodes are removed.

1) Run Time Observations for Parallel Algorithms: Firstly, we test our parallel algorithms under different sizes of graphs to observe the change in their running times. We achieve this by running our algorithms with increasing numbers of nodes and edges in the input graph. In order to observe the performance due to the changes in the number of edges, we used graphs which have the *same ratio* of the number of edges to the number of nodes in (in this case, the ratio is approximately 8). The graphs used in Figure 3 and Figure 4 are: G20, G24, G35, G37, G41, G42 and G45.

In Figure 3, run times of the Cuda, OpenMP and hybrid algorithms are plotted against the different number of edges in the input graph. It is inferred that OpenMP algorithm runs faster than the other implementations involving Cuda in small and medium sized inputs where the number of edges changes from 200K to 4M. However, when the number of edges go up to 5M, the Cuda implementations run faster than the OpenMP version. As can be seen in the rest of the results, run times of the Cuda and hybrid algorithms are very close to each other. In Figures 3 and 4, plots of these algorithms overlap.

TABLE I: Test Graphs

Graph	$\|V_R\|$	$\|E_R\|$	$\|E_R\|$ / $\|V_R\|$
G17	16,036	161,832	10.0
G18	19,654	327,037	16.6
G19	22,740	628,033	27.6
G20	26,798	206,590	7.7
G21	35,069	450,770	12.8
G22	42,087	895,957	21.2
G23	48,149	1,722,340	35.7
G24	51,030	418,251	8.1
G25	72,737	1,124,437	15.4
G26	86,495	2,236,552	25.8
G27	98,567	4,292,432	43.5
G28	118,329	1,307,774	11.0
G29	146,411	2,636,679	18.0
G30	173,399	5,188,036	29.9
G31	197,090	9,992,620	50.7
G32	28,023	67,795	2.4
G33	50,467	186,147	3.6
G34	75,304	424,621	5.6
G35	102,093	896,498	8.7
G36	172,325	933,288	5.4
G37	245,143	2,061,600	8.4
G38	323,929	4,314,274	13.3
G39	409,860	9,007,004	21.9
G40	319,769	1,567,005	4.9
G41	486,299	3,762,438	7.7
G42	583,850	5,785,179	9.9
G43	866,253	17,585,481	20.3
G44	834,701	5,385,893	6.4
G45	1,164,208	11,606,886	9.9
G46	1,527,605	23,990,183	15.7

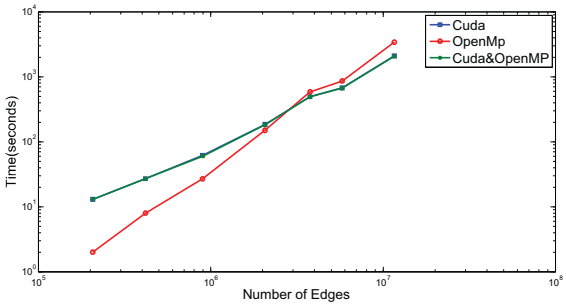


Fig. 3: Parallel Algorithms' Run Times with Respect to the Number of Edges in Graph.

In Figure 4, it can be observed that the run times of algorithms increase when the number of nodes grows. It can be seen that OpenMP runs faster than Cuda and hybrid algorithm in small and medium sized graphs. However, after the number of vertices reaches 500K and more then the OpenMP implementation takes more time than the two Cuda

based algorithms. In this figure, the ratio between the number of edges and the number of nodes is preserved as in Figure 3.

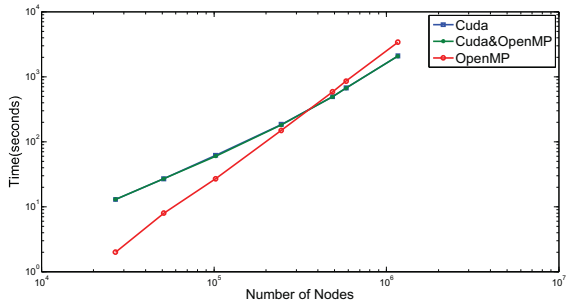


Fig. 4: Parallel Algorithms' Run Times with Respect to the Number of Nodes in Graph.

2) Comparison of Algorithms' Run Times: We test the sequential, OpenMP, Cuda and hybrid algorithms with respect to different size of graphs. Input graphs used in this section's figures are categorized as follows:

- Small and Medium Sized Graphs: G19, G21, G24, G25, G35, G29,G37, G40.
- Large Graphs: G38, G39, G41, G42, G44, G45, G46.

In Figure 5, we test our four algorithms on small and medium sized graphs. In the very small graphs the sequential algorithm runs faster than the Cuda and hybrid algorithm. However, when we increase the graph size a little, we can see that the sequential algorithm run time is increasing rapidly with the graph size, but the parallel algorithms' run times do not increase as much as that of the sequential algorithm. It can also seen that, although the run time of OpenMP algorithm is smaller than the rest of the algorithms, OpenMP algorithm run time increases more than those of the Cuda based algorithms.

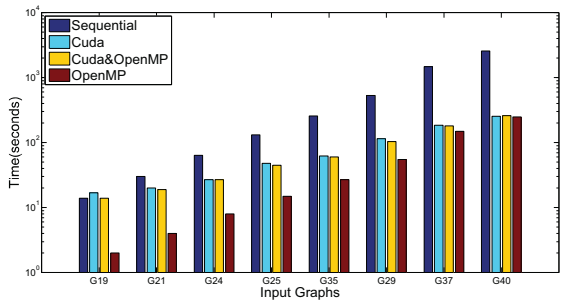


Fig. 5: Algorithm Run Times for Small and Medium Sized Graphs.

In Figure 6, we try to observe the algorithms' run times on large graphs where the number of edges vary between 4M to 24M. As can be seen in Figure 6, OpenMP algorithm runs slightly slower than two other parallel algorithms. On the other hand, Cuda and hybrid(Cuda and OpenMP) algorithms' run times are almost the same. The difference between these two algorithms is that hybrid algorithm uses OpenMP to parallelize the neighbour update process after finding the minimum distanced unvisited node. If the number of neighbours to update is small for a node, the OpenMP parallel instructions cannot achieve a good efficiency due to small parallelism and thread management overhead.

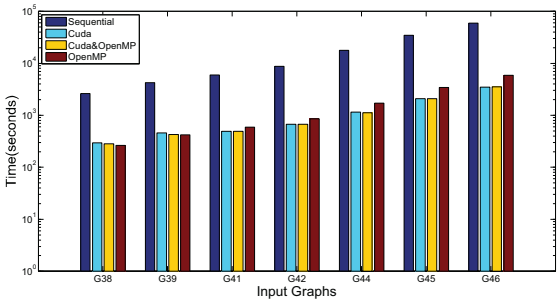


Fig. 6: Algorithm Run Times for Large Graphs

3) Speedup: When evaluating speedup performance, we consider all graphs ($G17 - G46$) used in our tests given in Table I. First, we order the graphs with respect to their sequential algorithm run times. The graph with the smallest run time is the first element of the list. The ordered list of graphs with respect to the sequential algorithm run times for the set of $G17 - G46$ are: $G17$, $G18$, $G19$, $G20$, $G32$, $G21$, $G22$, $G23$, $G33$, $G24$, $G25$, $G34$, $G26$, $G27$, $G35$, $G28$, $G29$, $G36$, $G30$, $G31$, $G37$, $G40$, $G38$, $G39$, $G41$, $G42$, $G44$, $G43$, $G45$ and $G46$ respectively.

In Figure 7, speedups of the parallel algorithm are plotted. The ratio of the number of edges divided by the number of vertices for each graph is also presented in this figure. OpenMP algorithm starts with a speedup value of around 7 on small graphs. After the graph size increases, OpenMP algorithm reaches its maximum speedup which is just above 10. On the other hand, it can be seen from this figure that Cuda and hybrid algorithm starts with a speedup of 1 and ends up with a speedup of 17. As the graph size increases the ratio of run times between sequential and the Cuda algorithms increase. Also, note that speedups of the the hybrid algorithm and the Cuda algorithm are close to each other. This is because the difference between those two algorithms is the updating of the arrival times of the neighbours after finding the node with the smallest arrival time. Therefore, the difference between speedups of these two Cuda algorithms increases when there are more neighbors to update, which also means when the ratio of the number of edges to the number of vertices increases.

In Figure 8, speedups of our three parallel implementations are shown with respect to the number of vertices and the number of edges in graphs using the same graphs in Figure 7. In Figures 7 and 8, fluctuations in speedups of Cuda and hybrid results are because of the ratio of number of edges to the number of vertices in the graph changes. An increase in this ratio causes the speedups of Cuda and hybrid implementations to decrease as the graph size increased, because the Cuda implementations mostly benefit from finding the minimum distanced unvisited node. So, an increase in number of vertices while the number of edges does not change has

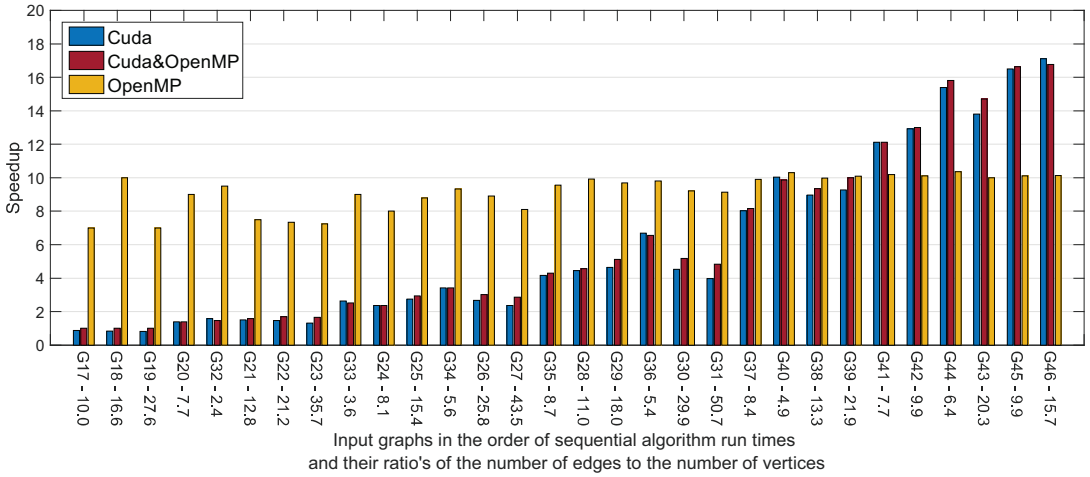


Fig. 7: Speedup of Parallel Algorithms Against Sequential Algorithm.

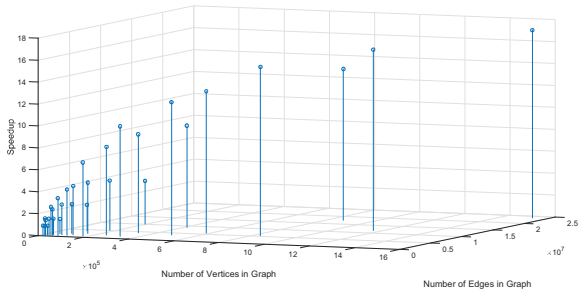
more speedup effect on Cuda algorithms than an increase in the number of edges while the number of vertices does not change in a graph. Note that, hybrid implementation uses OpenMP when updating neighbours of the minimum distanced node, therefore, the hybrid implementation's speedup does not decrease as much as the the Cuda implementation speedup while the ratio of number of edges to the number of vertices increasing. However, in the hybrid implementation there is a copy overhead from host vector to device vector after the update process, which also results in decreased speedup. Note that, OpenMP implementation speedup also decreases when the ratio of the number of edges to the number of nodes increases, but, because it has no copy overhead as in the hybrid implementation, the effect is smaller than those in other parallel algorithms.

4) Scalability: Scalability measures the performance of the program while the input and the processor size increases at the same ratio. Because the graphs we are testing are the graphs with unreachable nodes removed, it is hard to produce those graphs which have the same ratio of number of nodes to number of edges as in Graph500.

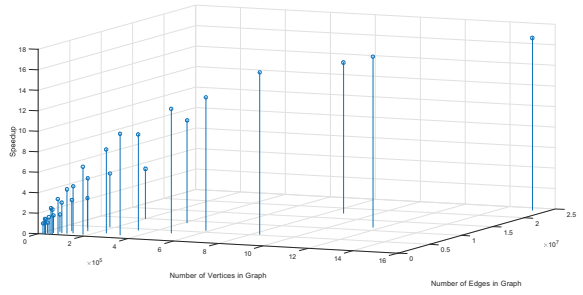
Therefore, we use 3 graphs, namely $G20$, $G24$ and $G35$, to examine our OpenMP implementation's scalability. Those graphs have twice as many number of reachable nodes and edges in the given order above. Then, we compare the efficiency of these graphs under 2, 4 and 8 threads respectively. Therefore, as we increase the graph size twice, we increase the number of threads twice as many too.

The results are shown in the Table II. In this table, T_s is the sequential algorithm run time in seconds, T_p is the OpenMP algorithm run time in seconds with given number of threads and η is the efficiency.

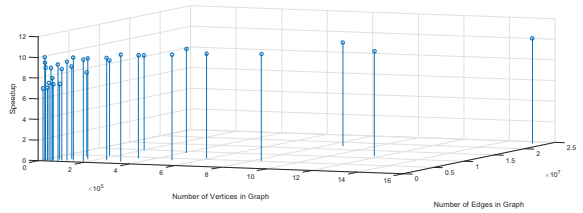
Results show that, as we increase the graph size and the number of processor by 2 and efficiency of algorithm is almost the same (range in between 0,81 and 0,84).



(a) Cuda



(b) Cuda&OpenMP



(c) OpenMP

Fig. 8: Speedups of Parallel Algorithms with Respect to the Number of Vertices and the Number of Edges in Graphs.

TABLE II: Scalability of OpenMP Algorithm.

Graph	$\|E_R\|/\|V_R\|$	No. of threads	T_s	T_p	η
G20	7.7	2	18	11	0.81
G24	8.1	4	64	19	0.84
G35	8.7	8	258	39	0.82

V. Discussion

OpenMP algorithm runs faster than sequential algorithm without taking into consideration of graph size. However, there is an increase in the algorithm speedup from 7 up to 10 while we increase the input graph size. This is because increasing graph size makes the program run more efficiently so that the total idle time of the processors with respect to the total run time will decrease.

Cuda and hybrid algorithms mostly benefit from finding the unvisited node with minimum arrival time faster than the OpenMP algorithm. As we can see in the figures at Section IV-C, Cuda algorithms increase the speedup ratios when graph size increases up to 17. On the other hand, as the graph size increases the speedup gap between Cuda and Cuda&OpenMP algorithms also increases. This is because, there is an additional parallelization section in the hybrid algorithm so that arrival times of the neighbors of the unvisited minimum arrival timed nodes are also updated in parallel. When the number of neighbors is small, the effect of parallelization cannot be perceived because there is also an overhead of the copying of all of the updated nodes to the $thrust :: device\_vector$.

VI. Conclusion

In this work, we present efficient parallel algorithms for shortest path problem from one node to all other nodes on time dependent graphs. We implement parallel algorithms in Cuda, OpenMP and hybrid Cuda and OpenMP. We evaluate these algorithms in terms of algorithm run times with respect to the sequential algorithm. For testing, we use an extensive set of graphs which have number of edges between 200K and 24M, and have number of nodes between 22K and 1.5M. We achieve up to 10-fold speedup with 12 cores in our OpenMP implementation, and 17-fold speedup in Cuda implementations. It is observed that OpenMP implementation has better performance when the number of nodes in the graph is around or smaller than 400K. On the other hand, Cuda and hybrid algorithms are good choices when the number of nodes in the graph is larger than 450K. Also, we can say that as the ratio between number of edges and number of nodes in a graph increases the hybrid algorithm becomes a little more efficient than the Cuda algorithm. If the memory constraint is not an issue, then hybrid algorithm can be a preferred solution on dense graphs.

Acknowledgment

The authors have used Bogazici University Polymer Research Center's cluster (funded by DPT project 2009K120520).

References

[1] S. ø. Larsen, H. Koren, and R. Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 7, pp. 859–869, 2009.

[2] B. Ding, J. X. Yu, and L. Qin, "Finding time-dependent shortest paths over large graphs," in *Proceedings of the 11Th International Conference on Extending Database Technology: Advances in Database Technology*. Acm, 2008, pp. 205–216.

[3] G. S. Brodal and R. Jacob, "Time-dependent networks as models to achieve fast exact time-table queries," *Electronic Notes in Theoretical Computer Science*, vol. 92, pp. 3–15, 2004.

[4] S. E. Dreyfus, "An appraisal of some shortest-path algorithms," *Operations Research*, vol. 17, no. 3, pp. 395–412, 1969.

[5] T. Ohshima, "A landmark algorithm for the time-dependent shortest path problem," Ph.D. dissertation, Citeseer, 2008.

[6] E. Kanoulas, Y. Du, T. Xia, and D. Zhang, "Finding fastest paths on a road network with speed patterns," in *Data Engineering, 2006. Icde'06. Proceedings of the 22Nd International Conference On*. Ieee, 2006, pp. 10–10.

[7] I. Chabini, "Discrete dynamic shortest path problems in transportation applications: Complexity and algorithms with optimal run time," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1645, pp. 170–175, 1998.

[8] K. Sung, M. G. Bell, M. Seong, and S. Park, "Shortest paths in a network with time-dependent flow speeds," *European Journal of Operational Research*, vol. 121, no. 1, pp. 32–39, 2000.

[9] B. C. Dean, "Shortest paths in fifo time-dependent networks: Theory and algorithms," *Rapport Technique, Massachusetts Institute of Technology*, 2004.

[10] H. D. Sherali, K. Ozbay, and S. Subramanian, "The time-dependent shortest pair of disjoint paths problem: Complexity, models, and algorithms," *Networks*, vol. 31, no. 4, pp. 259–272, 1998.

[11] A. V. Goldberg and C. Harrelson, "Computing the shortest path: A search meets graph theory," in *Proceedings of the Sixteenth Annual Acm-Siam Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2005, pp. 156–165.

[12] I. Chabini, "A new algorithm for shortest paths in discrete dynamic networks," Tech. Rep., 1997.

[13] V. Kumar and V. Singh, "Scalability of parallel algorithms for the all-pairs shortest-path problem," *Journal of Parallel and Distributed Computing*, vol. 13, no. 2, pp. 124–138, 1991.

[14] I. Chabini and S. Ganugapati, "Parallel algorithms for dynamic shortest path problems," *International Transactions in Operational Research*, vol. 9, no. 3, pp. 279–302, 2002.

[15] N. Tremblay and M. Florian, "Temporal shortest paths: Parallel computing implementations," *Parallel Computing*, vol. 27, no. 12, pp. 1569–1609, 2001.

[16] A. Ziliaskopoulos, D. Kotzinos, and H. S. Mahmassani, "Design and implementation of parallel time-dependent least time path algorithms for intelligent transportation systems applications," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 2, pp. 95–107, 1997.

[17] A. K. Ziliaskopoulos and H. S. Mahmassani, "Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications," *Transportation Research Record*, pp. 94–94, 1993.

[18] R. Mitton, *Very Sleepy*, 2014, Nov 2015. [Online]. Available: http://www.codersnotes.com/sleepy/

Statistical Data Analysis for Network Infrastructure Monitoring to Recognize Aberrant Behavior of System Local Segments

Mikhail Sukhoparov, Alexander Davydov
Dept. of Secure Information Technologies
Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University)
Saint-Petersburg, Russia
sukhoparovm@gmail.com

Ilya Lebedev, Nurzhan Bazhayev
Dept. of Secure Information Technologies
Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University)
Saint-Petersburg, Russian Federation
lebedev@cit.ifmo.ru

***Abstract*—Wireless network of low-power devices "Smart Home", "Internet of things" has been considered. A number of signs of security attacks on behalf of potential information interloper have been identified. We have analyzed the characteristics of a system based on wireless technologies that are obtained as a result of passive surveillance and active polling of devices comprising the network infrastructure. A model is presented for the information security state analysis based on identifying, quantitative, frequency, timing characteristics. In view of the peculiarities of the devices providing network infrastructure, estimation of the information security state is focused on the analysis of the system normal functioning profile, rather than on search of signatures and features of anomalies during various kinds of attacks. An experiment has been disclosed that provides obtaining statistical information about operation of wireless network remote devices where data acquisition for decision-making purposes occurs by comparing statistical signal messages from the leaf nodes in passive and active modes. Experimental results of information onslaught on the standard system have been presented. The proposed model may be used to determine technical characteristics of WLAN ad hoc network devices and to draw recommendations for IS state analysis.**

***Index Terms*—Information security, "soft" space wireless networks, vulnerability, device availability, information security model**

I. Introduction

The emergence of a large number of mobile devices connected to the Internet and the implementation of the processes of reception, transmission, processing of information incoming from them outside the controlled area necessitate providing information security.

Wireless technologies used in the APCS, "smart cities", "smart homes", "Internet of things", "soft spaces" become particularly vulnerable. And whereas there are typical tailored remedies in the first two areas, with regard to the remaining applications due to the lack of standardization, to date the developers pay little attention to solutions in the field of security [1,2,3].

User familiar household "smart" objects ("smart" microwave ovens, coffee makers, washing machines), being the components of the network segments of the Internet of Things or Smart Home, provide an opportunity to detect and identify themselves by generating information messages and can be exposed to hacking by the attacker. Unification interacting facilities by the individual producers of "smart" appliances, enabling users to operate simple setup processes without appropriate qualifications predetermines a variety of attempts to control such devices externally [4,5].

There are a number of signs that attackers are trying to get external access to the system. Detection of abnormal network traffic parameters, wrong or situationally unsuitable commands, identification of a large number of repeated events may be harbingers of security attacks [6,7].

Thus, the protection of information flows in order to ensure the integrity of the transmitted data is one of the urgent tasks for various low-power systems.

This results in a number of tasks aimed at implementation of the external monitoring of information security events for "smart" devices.

II. Problem description

Wireless network comprising multiple nodes is a typical solution for the organization of interaction between the devices. Network infrastructure is a set of physical and logical components that provide connectivity, security, routing, control, access and other required properties of the network [8].

The considered devices providing connection of various appliances and "soft space" devices on the whole are low power ones and allow for the reception, processing and transmission of the limited types of messages.

To detect anomalous behavior, it is possible to use data reflecting the system states which can be used in the statistical analysis [9,10]. At the initial stage of operation after deployment, one can appreciate various characteristics of the intensity of information and signal messages, response time to service requests, frequency of unrecognized and missed

messages. Such characteristics can be obtained by passive surveillance and active polling devices.

Thus, it is necessary to determine the characteristics of the multiple elements of the controlled system enabling to identify its abnormal condition relative to “normal” operation with the specified probability.

III. Description of the Approach

Traditionally, detection of intruder’s activities involves identifying unforeseen parameters of network packets (wrong addresses, simultaneously cocked message flags and connection requests, network traffic analysis). However, analysis of these events at the low levels of networking with the purpose to detect information security incidents is a challenge even for professionals; it requires knowledge of networking protocols that are ad-hoc developed for devices of specific manufacturers.

One of the possible ways to analyze the system states may be carried out based on statistical data of application-layer protocols for interacting nodes of low-power devices by passive and active monitoring. In the first case the network devices listening and statistical analysis of transmission and reception events for various types of messages is performed; in the second case the monitoring system is requested in the form of various commands and temporal delays, load changes of separate computing resources are analyzed and the identification information and settings are compared.

Therefore, detection methods that can be implemented on the external independent devices are the most attractive for the average user.

Figure 1 shows a typical system, where the information from the sensors by passive listening accumulates in the processing node.

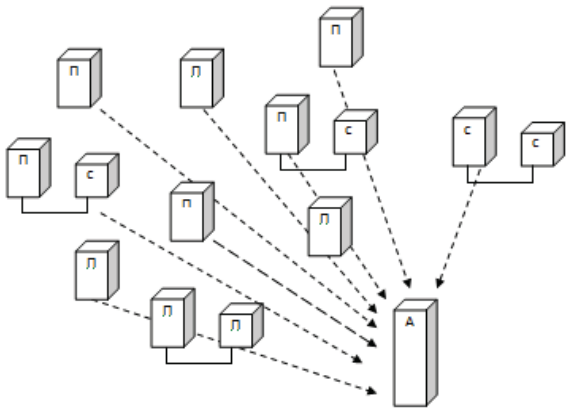


Fig.1. Interaction of network nodes.

Tracking a number of events associated with the increase in the network enables to consider the model for information security state analysis in the form of identifying (I), quantitative (N), frequency (ϕ), and timing (T) characteristics. These events include:

- emergence of unrecognizable messages;
- emergence of repeated messages;
- increase in the number of device error and fault messages;
- increase in the number of broadcast and signal messages;
- occurrence of delays resulting in statistical change of information and signal message traffic;
- change in delays of device response to service broadcast requests for various operation modes;
- increase in the number of lost messages;
- frequency variation of information and signal messages.

The profile of system functioning will be determined by a tuple of attributes:

$$\Phi = < I, N, \phi, T > \quad (1)$$

Each of the attributes is a time varying vector of values. Depending on the operation mode the change in the statistical portrait of networks and devices functioning is observed.

In view of the low-power of devices providing the network infrastructure, it is easier to evaluate the state of information security based on the normal system functioning profiles. This makes it possible to implement the basic methods of control over the endpoints based on the criteria of "who", "what", "how", "where", "when". There is the possibility of storing the feature space in the form of relational database tables related events of these devices, which is filled with at least the functioning of the system. Template for assessing the state of information security applied at the node is shown in Figure 2.

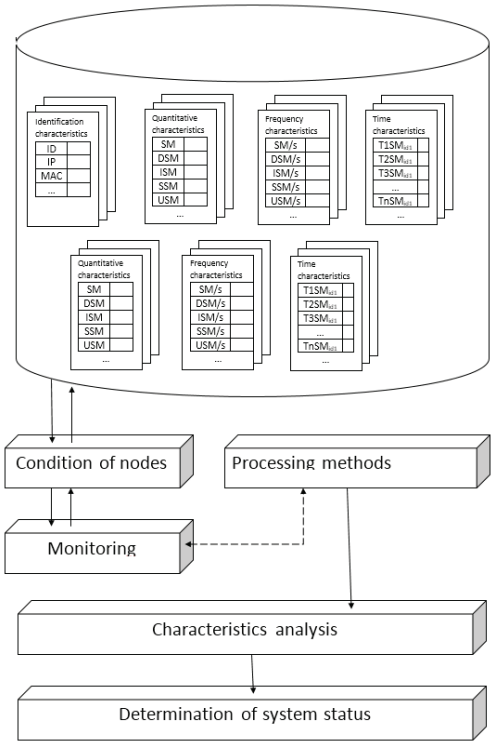


Fig.2. Template for assessing the state of information security systems.

IV. Experiment

Figures 3 shows the analyzed system. Information and signal messages circulate between the nodes A and B, the device C is designed to gather information. In the first case, the device C listens to the network and generates sampled statistical data; in the second case, it additionally sends requests to the devices and measures various characteristics.

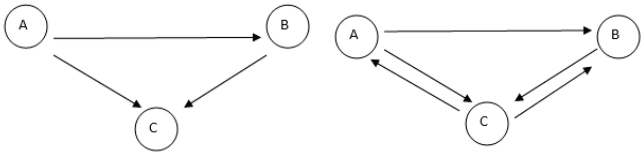


Fig. 3. The system scheme

In various system operating modes anomalies can be observed that require a more detailed study regarding the possibility of unauthorized access. Determination of quantitative and frequency data indicating unrecognized, skipped messages, obtaining information on the final state of the nodes on the basis of statistical data of application-layer protocols for interacting nodes of low-power devices by means of passive and active monitoring make it possible to construct the classifier.

At the same time the enumerated attributes may be a result of not only the attacks, but also of random errors and equipment failures.

The naive Bayes classifier was implemented in the framework of the experiment. Its advantage is a small amount of training data needed to assess the parameters required for the classification.

$$C = argmax_{h \in H} \frac{p(X|h)p(h)}{p(X)} \quad (2)$$

where h, X are predicted and preceding events, and function P – the probabilities of these events and their consequences ($P=m/n$, where m – number of the events occurred, n – total number of all events).

To form a decision rule the data obtained in conditions of active and passive monitoring were used:

- relative frequencies of classes of the system states;
- total number of attribute characteristics in the classes defined for the analysis;
- relative frequencies of attributes within each class;
- the number of sampling attributes

In this experiment, attacker's hacking at the application layer is reduced to changing the intensity of message arrivals (interception and reproduction of information messages, over-filling with signal messages causing the process of association, authentication, dissociation, deauthentication, connection requests). The sequence of commands sent to the operating devices was chosen randomly.

The data are accumulated by comparing the statistical information of service messages from leaf nodes in the passive and active modes.

Figures 4,5,6 show a number of analyzed states of the system considered in the experiment, where the functioning results are grouped into statistical data based on the types of messages.

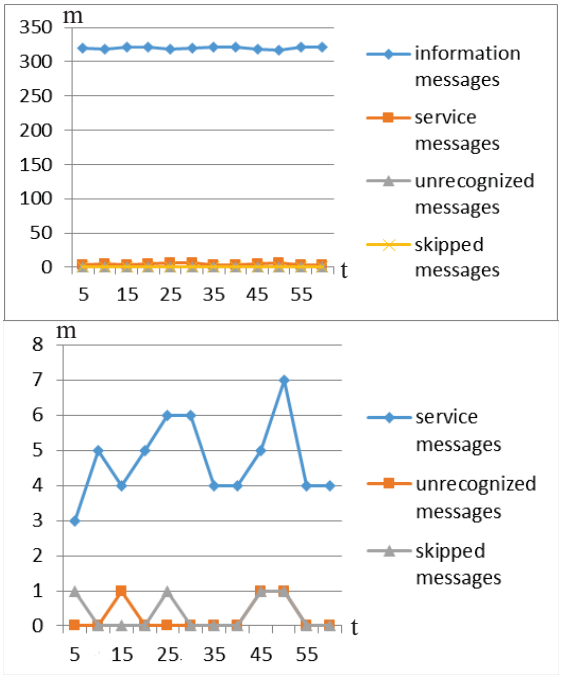


Fig. 4. The system in the usual state. m - number of the events occurred, t – time in seconds.

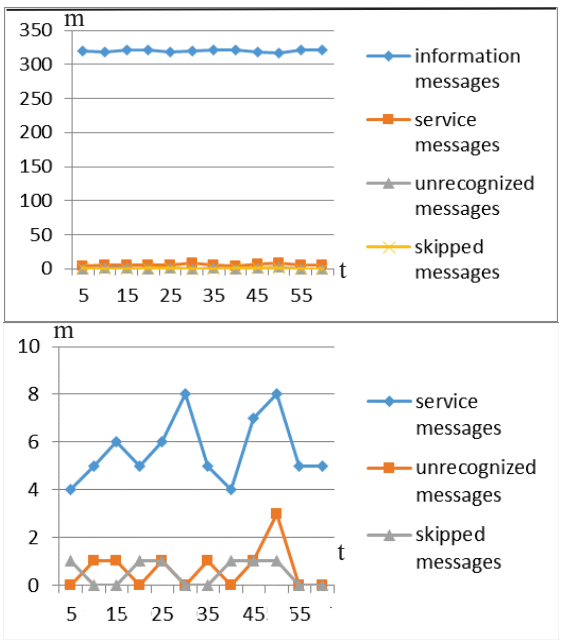


Fig. 5. The system during active monitoring. m - number of the events occurred, t – time in seconds.

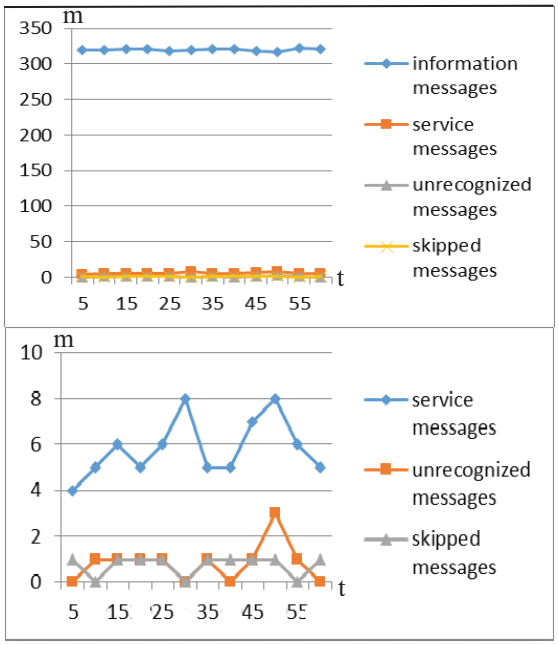


Fig. 6. The attacked system during active monitoring. m - number of the events occurred, t – time in seconds.

A reference tuple shown in Fig. 7 is constructed based on the statistical information obtained at regular time intervals, associated with the operation of devices.

	5 sec	10 sec	…	T sec
information messages	320	319	…	321
signal messages	3	5	…	5
unrecognized messages	0	0	…	0
skipped messages	1	0	…	0
. . .				
time delay for message sending	0.001	0.001	…	0.002

Fig. 7. The system in the usual state

While obtaining and analyzing the dependence of qualitative parameters of the system information signal messages for various modes a number of activities was performed:

- system switching to the required operation mode for learning;
- analysis of the system state attributes;
- formation of the analyzed attribute tuple;
- reception of messages from the devices, accumulation of statistics, database creation for the studies parameters;
- processing of the accumulated statistical data, comparing them with the delivered data, the current state assessing.

Figures 8-9 show the results of the classifier operation for various states of the system.

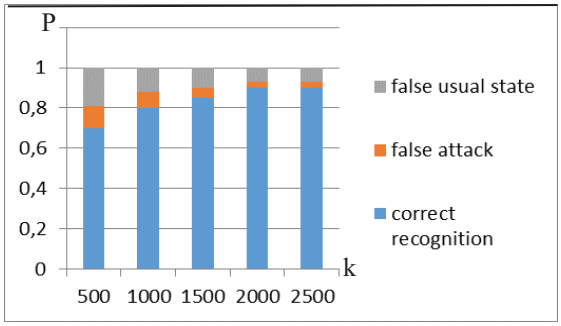


Fig. 8. Statistical data tuples of the system. P – probability, k – training sample size.

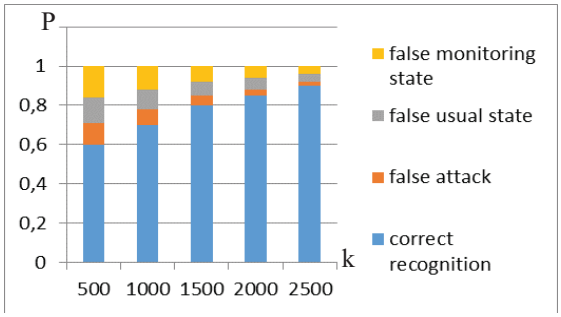


Fig. 9. Statistical data tuples of the system. P – probability, k – training sample size.

To assess the qualitative characteristics it is required to select various parameters and their groups. However, even the statistics obtained on the basis of the conducted experiment shows a different kind of the analyzed system responses sufficient for probabilistic determination of the state.

V. Conclusion

A method is proposed for monitoring the information security state of low-power device network segments based on the statistical data of interaction between receiving-transmitting devices used to control the functioning of the local information system application layer, which allows obtaining the probabilistic characteristics of the information security state. A feature of the approach is the ability to quickly adapt the decisions of the Internet of things and smart home to the local networks of low-power devices made by different manufacturers.

To implement this kind of monitoring it is not necessary to develop complicated system applications.

The proposed method can be used when searching for anomalies segments of smart home networks and the Internet of things. At the same time feature space requires additional analysis and definition of informativeness of individual characteristics to improve the accuracy and adaptation to the system real-time intrusion detection.

References

[1] Kumar P., Ylianttila M., Gurtov A., Lee S.-G., Lee H.-J. “An Efficient and Adaptive Mutual Authentication Framework for

Heterogeneous Wireless Sensor Networks-based Applications," MDPI Sensors, 14(2), pp. 2732-2755, 2014.

[2] Sridhar P., Sheikh-Bahaei S., Xia S., Jamshidi Mo, "Multi agent simulation using discrete event and soft-computing methodologies," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics 2, pp. 1711-1716, 2003.

[3] Page J., Zaslavsky A., Indrawan M., "Countering security vulnerabilities using a shared security buddy model schema in mobile agent communities," Proc. of the First International Workshop on Safety and Security in Multi-Agent Systems (SASEMAS 2004), pp. 85–101, 2004.

[4] Zikratov I., Lebedev I., Gurtov A., "Trust and Reputation Mechanisms for Multi-agent Robotic Systems," Lecture Notes in Computer Science Volume 8638, pp. 106-120. (Proc. of the Internet of Things, Smart Spaces, and Next Generation Networks and Systems), 2014.

[5] Wyglinski A.M., Huang X., Padir T., Lai L., Eisenbarth T.R., Venkatasubramanian K., "Security of autonomous systems employing embedded computing and sensors," IEEE Micro 33 (1), art. no. 6504448, pp. 80-86, 2013.

[6] Lebedev I.S., Korzhuk V.M., "The Monitoring of Information Security of Remote Devices of Wireless Networks," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9247, pp. 3–10, 2015.

[7] Prabhakar M., Singh J.N., Mahadevan G., "Nash equilibrium and Marcov chains to enhance game theoretic approach for vanet security," International Conference on Advances in Computing, ICAdC 2012; Bangalore, Karnataka; India; 4 July 2012 through 6 July 2012, Volume 174 AISC, pp. 191-199, 2013.

[8] Bazhayev, N., Lebedev, I., Korzhuk, V., Zikratov, I., "Monitoring of the Information Security of Wireless remote devices," Source of the Document 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings 7338553, pp. 233-236, 2015.

[9] Nikolaevskiy I., Lukyanenko A., Polishchuk T., Polishchuk V.M., Gurtov A.V., "isBF: Scalable In-Packet Bloom Filter Based Multicast," Computer Communications. Vol. 70, pp. 79–85, 2015.

[10] Al-Naggar Y., Koucheryavy A., "Fuzzy Logic and Voronoi Diagram Using for Cluster Head Selection in Ubiquitous Sensor Networks," Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 14th International Conference, NEW2AN 2014 and 7th Conference, ruSMART 2014 Saint Petersburg, August 27–29, Proceedings. Springer, LNCS 8638, pp. 319–330, 2014.

[11] Chehri A.,Hussein T., "Moutah Survivable and Scalable Wireless Solution for E-health and Emergency Applications In EICS4MED 2011," Proceedings of the 1st International Workshop on Engineering Interactive Computing Systems for Medicine and Health Care, Pisa, Italy, pp. 25–29, 2011.

Text Big Data Analytics: exploring API opportunity

Internet as Global storage – how to get the situation awareness from Dark Data

Olga Kolesnichenko
Security Analysis Bulletin
Moscow, Russia
oykolesnichenko@list.ru

Gennady Smorodin
EMC Academic Alliance Russia & CIS
St. Petersburg, Russia
gennady.smorodin@emc.com

Dariya Yakovleva
Vladivostok State University of Economics and Service
Vladivostok, Russia
darya.yakovleva15@vvsu.ru

Lev Mazelis
Vladivostok State University of Economics and Service
Vladivostok, Russia
lev.mazelis@vvsu.ru

Oleg Zhurenkov
Altai Academy of Economics and Law
Barnaul, Russia
zhur@pie-aael.ru

Yuriy Kolesnichenko
Security Analysis Bulletin
Moscow, Russia
green-apple_2000@mtu-net.ru

***Abstract* — Text Big Data Analytics Study «Third Wave» is described. Morphological matrix of several Keywords Phrases was collected from Internet's open textual resources using API. The results are analyzed from the point of view that global Internet's audience forms «people-to-IT» system through that we can study the three levels of behaviour of Global society. Level 1 – countries' activity on the Internet. Level 2 – technological evolution in countries. Level 3 – countries' reaction to destructive stress factors. Cluster analysis was used for investigation of Level 1. Simon Wardley's Value Chain Map was used for investigation of Level 2. And diagrams with percentage ratio of different sets of Keywords Phrases were used for investigation of Level 3. Article is focused on regions of Eurasia and North Africa. The Study findings are following: among post-Soviet countries Russia shows the highest activity on Internet and has advantages of economic development, as well as Georgia; Azerbaijan among some countries from Eurasia and North Africa shows the lowest risk to become fragile country under Global warming.**

***Index Terms* — Big Data Analytics, Google, Yandex, API, Data Mining, K-Means Cluster Analysis.**

I. MULTICENTER STUDY «THIRD WAVE»

EMC Academic Alliance Russia & CIS has launched this study in 2015. Big Data Analytics Study «Third Wave» has been named after Alvin Toffler's «Third Wave» concept about information era [1]. Morphological matrix of several Keywords Phrases (KP) was collected from Internet's open textual resources. Data Mining was implemented as High Performance Big Data Analytics (HPDA) [2]. API-access (Application Programming Interface) to Google and Yandex data storages was used. Google and Yandex are considered as non-classical hybrid supercomputers with API-access «as-a-Service». KP were counted in million, in conjunction with the countries names and year of publication or mention of the year in the text. K-Means Cluster Analysis was used for metadata. Graphic data processing performed using Microsoft Excel.

In all the results of Text Big Data Analytics for 49 countries analyzed in the study «Third Wave». Following KP were selected for Data Mining from Google and Yandex: Computer; Cloud computing; Electric cars; Mobile phone; Drip irrigation; Gas supplies; Oil; Oil price; Nuclear power plant; Solar panel. The languages of KP are English and Russian (words were counted together). The results of the study were considered from the point of Value Chain Map that has been presented by Simon Wardley at O'Reilly Conference «Open Source Convention» [3; 4]. Value Chain Map was modified to be applied to analyze the results of Text Big Data Analytics in this study. Value Chain Maps with number of Keywords Phrases in million during 2015 were created for some groups of countries associated with appropriate geographic region. In this article the region of Commonwealth of Independent States (CIS) and post-Soviet countries is described. Also the group of countries from Eurasia and North Africa is analyzed using enlarged set of KP.

Analyzing text arrays from the Internet we don't study the Global society separately from the information environment, but we can make an assessment of Global society reflection in information systems. The Internet's environment is some kind of global audience discussion about various processes. In this

regard the identified KP patterns are not a direct indication to processes in society, economics and politics, but KP patterns reflect the interest in various economics or politics processes. KP patterns reflect the society as a dynamic flexible system with adaptive spatiotemporal changes. Global Internet's audience forms «people-to-IT» system through that we can study the three levels of behaviour of Global society:

Level 1 – activity on the Internet that is different for different countries and depends on number of computers in the country; level of literacy and IT education; number of IP connections and Internet accessibility; level of economic development and number of Internet services; amount of country's population; processes in politics and society, which are actively discussed; citing of country.

Level 2 – Internet's open textual resources reflect the different features of technological evolution in different countries. People write about technological and economic processes which have place in their life, business, country's agenda etc.

Level 3 – reaction to stress, people write more about that problems which concern them more.

II. Society behaviour, Internet activity

Different countries are very varied. Cluster analysis helps to divide countries into groups with different activity and different levels of citing of countries on the Internet. In this study five clusters are found (Fig. 1).

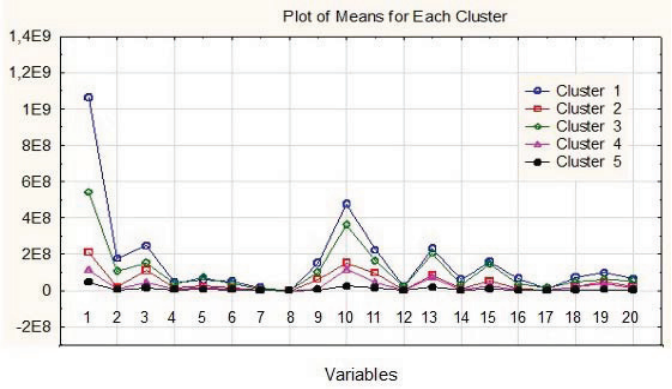


Fig. 1. Cluster analysis, k-means algorithm. 49 countries are divided into five clusters. Variables – matrix of 20 Keywords Phrases, 2015: 1 – Computer; 2 – Cloud computing; 3 – Mobile phone; 4 – Dollar exchange rate; 5 – Euro exchange rate; 6 – Electric cars; 7 – Solar panel; 8 – Drip irrigation; 9 – Gas supplies; 10 – Oil; 11 – Oil price; 12 – Nuclear power plant; 13 – Crisis; 14 – Inflation; 15 – Price increase; 16 – Unemployment; 17 – Dismissal; 18 – Migrants; 19 – Refugees; 20 – Muslims.

Cluster № 1 includes United States, Canada, Germany, China – the most active countries on Internet. Cluster № 3 includes United Kingdom, Australia, Japan, Russia. These countries also have high level of activity and citing on the Internet.

Cluster № 2 includes Hong Kong, Singapore, New Zealand, Ireland, Malaysia, Ukraine, Georgia, Turkey, India. Cluster № 4 includes Switzerland, Chile, Sweden, Denmark, Norway, Colombia, Taiwan, South Korea, Austria, Israel, Afghanistan, Pakistan, Egypt, Iraq, Iran, Syria. Both clusters № 2 and № 4 contain countries with middle level of activity and citing on the Internet.

Cluster № 5 includes Luxembourg, Estonia, Lithuania, Finland, Macau, Saint Lucia, Barbados, Peru, Libya, Bahrain, Tunisia, Yemen, Palestine, Azerbaijan, Kyrgyzstan, Uzbekistan. This cluster contains countries with low activity and citing on the Internet.

As was mentioned above, it is hard to take into account all the factors affecting the activity on the Internet. At this stage of our study we chose the clusterization to arrange countries with different number of keywords on the Internet. In this article following countries are described: Russia, Estonia, Lithuania, Ukraine, Georgia, Azerbaijan, Kyrgyzstan, Uzbekistan, China, Turkey, Syria, Iraq, Afghanistan, Tunisia, Libya, Yemen. These 16 countries belong to different clusters.

High level of Internet activity and citing: Cluster № 1 – China; Cluster № 3 – Russia.

Middle level of Internet activity and citing: Cluster № 2 – Ukraine, Georgia, Turkey; Cluster № 4 – Afghanistan, Iraq, Syria.

Low level of Internet activity and citing: Cluster № 5 – Estonia, Lithuania, Libya, Tunisia, Yemen, Azerbaijan, Kyrgyzstan, Uzbekistan.

These clusters allocation should be taken into account as some common denominator while considering next figures. But at the same time we should understand that more or less activity and citing of countries on the Internet point out the trends of economic development in countries, as well as interest in countries regional market and countries' activity at Global market.

III. Society behaviour, Technological evolution

We used Simon Wardley's approach for assess the economic situation in terms of competitiveness and shaping the future consumer market. Wardley's map was modified and KP from our study were added (Fig. 2). Making map we can look to KP data in dynamics.

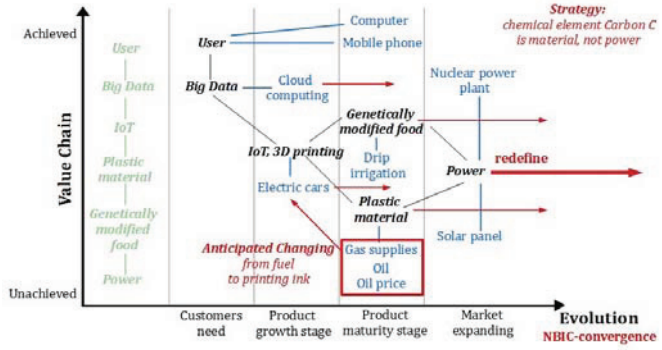


Fig. 2. The modified Value Chain Map for results of Text Big Data Analytics in study «Third Wave». All the components of the Value Chain Map are evolving from left to right due to demand competition. Blue font on this scheme – the selected Keywords Phrases.

All the components of the Value Chain Map are evolving from left to right due to demand competition. Value Chain Map allows to put KP in needed sequence to determine what is already cover previous stage of development, and what is the innovative stage that determines future success.

The chosen KP for Data Mining in this study reflect the main modern economic trends. The technological evolution can be investigated using KP and Value Chain Map. Kondratiev's waves are the measure of the evolution of technological order [5]. Kondratiev's waves endure 30-50 years. Let's look to Value Chain Map with KP for region of Commonwealth of Independent States (CIS) and post-Soviet countries (Fig. 3; Fig. 4). Map helps to look to diagram not as static that shows more or less number of KP, but in dynamics, with understanding past and future processes.

Fig. 3. Google map of Commonwealth of Independent States (CIS) and post-Soviet countries with indication to chosen countries for study «Third Wave».

Start of Cybernetic revolution had coincided with the Fourth Kondratiev's wave (period from 1947 to 1991). The KP *«Computer»* is the sign of Fourth Kondratiev's wave. This keyword has the highest frequency among all ten chosen KP. Today Fourth Kondratiev's wave is completely finished and we can see a clear dominance of KP «Computer» in the study as widely distributed products.

KP *«Mobile phone»* can be attributed to the period of the Fifth Kondratiev's wave (period from 1991 to 2020). Only in the last decade mobile phones have become widespread in the world as distributed products.

KP *«Cloud computing»* is the sign of upcoming Sixth Kondratiev's wave (period from 2020 to 2070), during this wave will be completing of Cybernetic revolution. Today Cloud computing cannot yet be called as distributed product, it is still looking for ways of development and places at the market.

KP *«Electric cars»* is the sign of Sixth Kondratiev's wave. Electric cars are the real turning point in technological order. The most interesting trend is the emergence of 3D-printing technology for production of electric cars. 3D-printing material for these cars are carbon fibre composite tubes. This means that oil and gas are no longer needed as fuel for new cars, but they are very needed as oil and gas chemistry products for 3D-printing of cars. The process of 3D-printing is based on Cloud computing [6]. More Cloud computing will expand, more 3D-printing cars will be at the market. The new dependency will appear during Sixth Kondratiev's wave: more Cloud computing, more common electricity grid, more electric cars, more 3D-printing, more needed oil and gas chemistry products as material for cars 3D-printing.

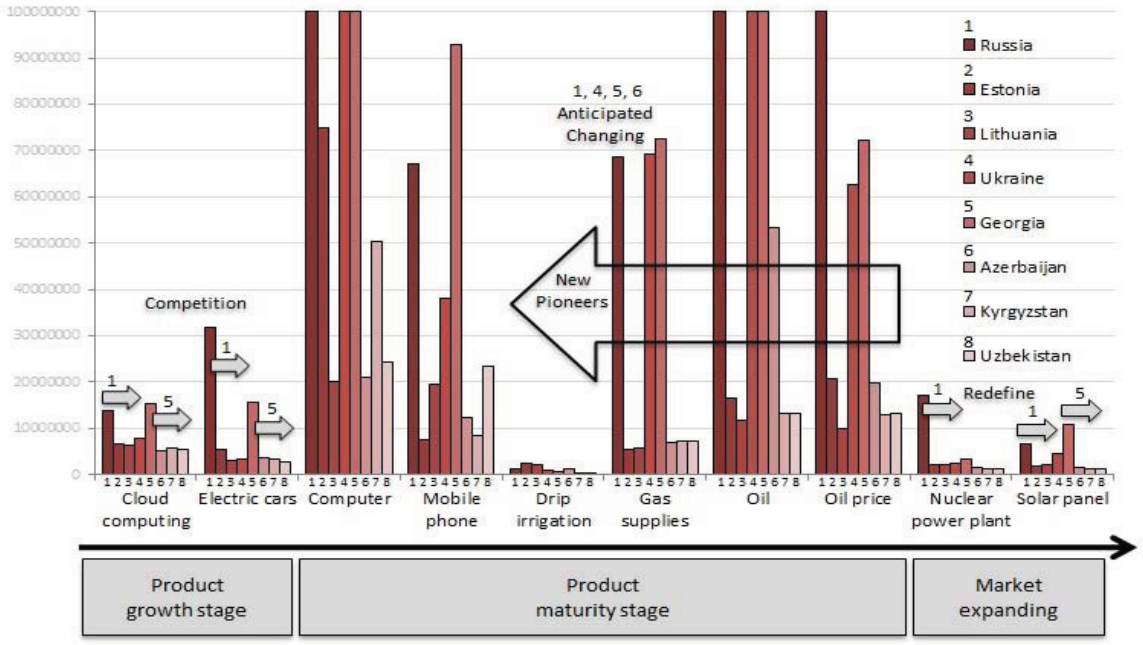


Fig. 4. The number of Keywords Phrases in million during 2015 in conjunction with the countries names. Level above 100 million isn't shown.

KP *«Gas supplies»; «Oil»; «Oil price»* are the signs of previous hydrocarbons' age. But during Sixth Kondratiev's wave hydrocarbons will change the function from fuel to «printing ink» for 3D-printing mass consumption products, including cars.

KP *«Drip irrigation»* reflect the important trends – agricultural IoT, technology involvement into agriculture, food security. Adequate water supply is topical problem in terms of Global warming for regions with growing population in Africa and Asia.

KP *«Solar panel»* reflect the most important future oriented trend of transition to renewable energy sources.

KP *«Nuclear power plant»* reflect the traditional energy type for 36 countries around the world. In this article among mentioned countries China, Russia, Ukraine and Lithuania have the Nuclear power plants.

Looking to Value Chain Map (Fig. 2) we can emphasize that the aim of map is to show a moment of advantage, a moment of competition. In this chain the moment of competition, that will guarantee leadership, is the energy power, that will determine the acceleration of development of new technology, IoT. Value Chain Map for CIS and post-Soviet countries (Fig. 4) shows that Russia and Georgia have increased level of KP «Cloud computing», «Electric cars» and «Solar panel». Despite the difference of clusters (Russia is in the High level cluster; Georgia is in the Middle level cluster, as well as Ukraine; and other countries from diagram are in the Low level cluster), other KP (indicated at Product maturity stage) are more comparable among countries from different clusters.

We can notice that Russia and Georgia in this group of countries show the growing interest in two crucial trends of development – Cloud computing and Electric cars, and also they show the growing interest in renewable energy (for example Russia has Orsk and Abakan Solar power plants with capacity 25 MW and 5,2 MW; Georgia has Solar Thermal systems and Solar photovoltaic panels in Tusheti, Solar electricity generation system in Tbilisi Shota Rustaveli International Airport, and schedules the Solar power plant in Gardabani, Georgia's Kvemo Kartli region). Both new emerging technologies and unlimited energy supply will ensure advantages of economic development for Russia and Georgia among presented group of countries. Should be mentioned that word «Georgia» on Internet is also related to U.S. state, which has two times more territory than country Georgia. For example U.S. state Georgia has two Nuclear power plants: Plant Vogtle in Augusta and Plant Hatch in Baxley. But as we see on diagrams (Fig. 4; Fig. 5) on the countries level this additional regional data do not affect whole picture and do not exceed the level of countries without Nuclear power plants.

Russia, Georgia, Ukraine and Azerbaijan have the growing interest in hydrocarbons (Fig. 4). We can notice that these countries will be in the common row of pioneers at the new market of carbon materials, «printing ink» for 3D-printing, that will replace at hydrocarbons market fuel to materials.

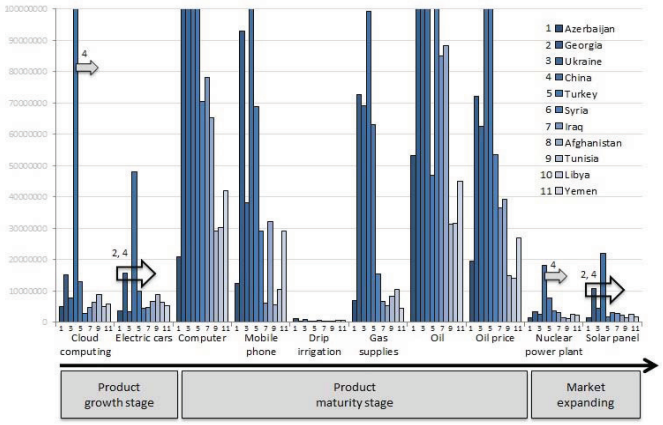


Fig. 5. The number of Keywords Phrases in million during 2015 in conjunction with the countries names. Level above 100 million isn't shown. Keywords Phrases: Cloud computing, Electric cars, Computer, Mobile phone, Drip irrigation, Gas supplies, Oil, Oil price, Nuclear power plant, Solar panel.

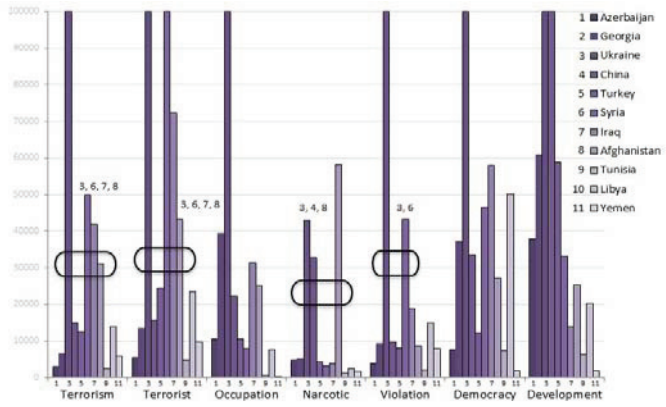


Fig. 6. The number of Keywords Phrases in million during 2000-2015 summarily, in conjunction with the countries names. Level above 100 million isn't shown. Keywords Phrases: Terrorism, Terrorist, Occupation, Narcotic, Violation, Democracy, Development.

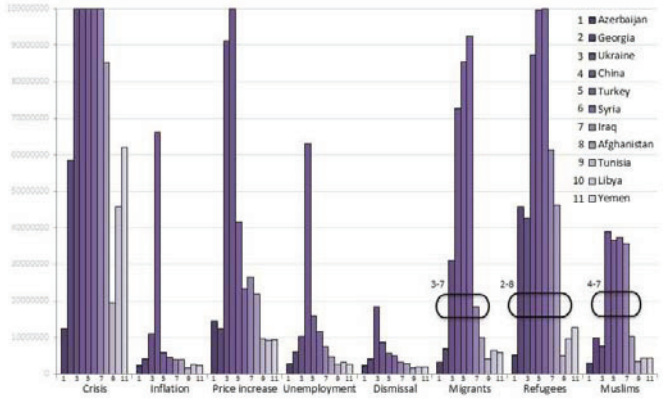


Fig. 7. The number of Keywords Phrases in million during 2015 in conjunction with the countries names. Level above 100 million isn't shown. Keywords Phrases: Crisis, Inflation, Price increase, Unemployment, Dismissal, Migrants, Refugees, Muslims.

IV. Society behaviour, Reaction to stress

Looking to three diagrams (Fig. 5; Fig. 6; Fig. 7) we can investigate the impact of stress factors on the countries interest in innovative development. There are three sets of KP in conjunction with countries names: Azerbaijan, Georgia, Ukraine, China, Turkey, Syria, Iraq, Afghanistan, Tunisia, Libya, Yemen (China is in the High level cluster; Georgia, Ukraine, Turkey, Afghanistan, Iraq, Syria are in the Middle level cluster; other are in the Low level cluster). On Value Chain Map (Fig. 5) we see two countries that show the growing interest in new emerging technology and unlimited energy supply – China and Georgia. As on previous diagram (Fig. 4) the leaders belong to High and Middle level clusters, but other KP (indicated at Product maturity stage) are more comparable among countries from different clusters.

Ukraine, Syria, Iraq, Afghanistan have the higher level of KP frequency «Terrorism», «Terrorist», «Narcotic», «Violation» (Fig. 6). Also these countries have the higher level of KP frequency «Migrants», «Refugees» (Fig. 7). Thus countries with indication to innovative development don't have indication to destructive stress factors that are causing damage to country's development. And countries with indication to destructive stress factors don't have indication to innovative development. But we can notice that Georgia has higher level of KP «Refugees» and China has higher level of KP «Narcotic». This is the indication to inhibitory factor for each of two countries.

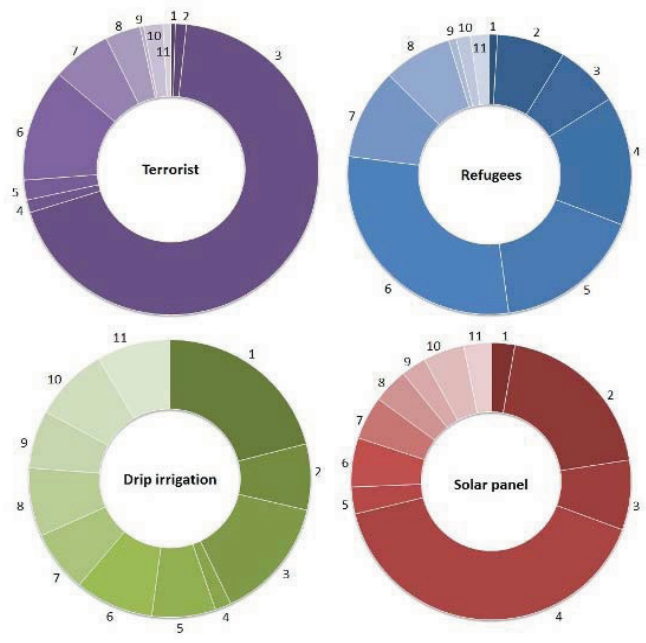


Fig. 8. The percentage ratio of countries interest in four Keywords Phrases (Terrorist, Refugees, Drip irrigation, Solar panel) during 2015 by count of Keywords Phrases. Countries: 1 – Azerbaijan; 2 – Georgia; 3 – Ukraine; 4 – China; 5 – Turkey; 6 – Syria; 7 – Iraq; 8 – Afghanistan; 9 – Tunisia; 10 – Libya; 11 – Yemen.

Another four round diagrams (Fig. 8) show that Azerbaijan (1) has the lowest indication to destructive stress factors among presented countries, and also Azerbaijan has predominant interest in drip irrigation. Azerbaijan as well as all eleven presented countries will suffer from impending Global warming. Deficit of water for agriculture and drought will cause hunger and instability in the countries. To prevent falling countries need to develop a broad network of drip irrigation, which requires a lot of electricity. Solar energy is more appropriate to regional environment. Looking to four diagrams (Fig. 8) and making correction that China (4) belongs to the High level cluster, and data about Georgia (2) include data about U.S. state Georgia (which has 400 MW capacity Solar panels park), we can recognize Azerbaijan as country with lowest risk to become fragile country under Global warming. Now Azerbaijan has Surakhani Solar Power Station with 2,8 MW capacity, and plans to open new Solar power plants in Sumgayit, Samukh, Sangachal and Pirallahi [7].

V. Conclusion

This type of Text Big Data Analytics is the starting point. API-access to data storages is the modern and growing industry. We should investigate this opportunity and elaborate varied analytical methods for extraction the new knowledge from Dark Data (data that before was not recognized as data). In this article we show that morphological matrix of several Keywords Phrases can open the knowledge about countries activity on the Internet, interest in technological trends and reaction to destructive stress factors. The understanding of different words frequency on the global Internet is the important scientific task that leads to connection of Artificial Intelligence with Internet for the rapid analysis of situation and forecast.

References

[1] E. Toffler, The Third Wave. Moscow, Russia: AST, 2004.

[2] D. Dietrich, B. Heller, B. Yang, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, USA: Wiley, 2015.

[3] S. Wardley. Open Source Convention – OSCON 2014. Anticipating the Future: An Introduction to Value Chain Mapping. Keynote lecture. Web: http://www.youtube.com/watch?v=NnFeIt-uaEc.

[4] S. Wardley. Open Source Convention – OSCON 2015. Situation Normal, Everything Must Change. Keynote lecture. Web: http://www.youtube.com/watch?v=Ty6pOVEc3bA.

[5] Kondratiev's waves: Heritage and Modernity. Edited by L.E. Grinin, A.V. Korotaev, V.M. Bondarenko. Volgolgrad, Russia: Uchitel, 2015.

[6] What is the future of Cloud computing? Dell official website, Web: https://powermore.dell.com/micro-sites/future-of-cloud-computing/.

[7] Ilham Aliyev attended the opening of the Surakhani solar power station. Official website of the President of the Republic of Azerbaijan. 16 July 2014. Web: http://en.president.az/articles/12402.

Determination of Loss of Information During Data Anonymization Procedure

Alexey Pervushin
ITMO University
St. Petersburg, Russia
172609@niuitmo.ru

Viktoria Ermachkova
ITMO University
St. Petersburg, Russia
buka-e@ya.ru

Anton Spivak
ITMO University
St. Petersburg, Russia
Anton.Spivak@niuitmo.ru

***Abstract*—One of the effective approaches to the protection of personal data is their depersonalization, as it reduces the requirements for the level of data protection. Therefore anonymization procedure is widely used in practice. As a result of the application of such methods discards some of the information, which entails a loss of information content of the personal data.The problem of determining the loss of information during data anonymization procedure. It is often easier to allocate a group of similar objects, to analyze the loss for each group, than analyze all the data at once. A large amount of data and the lack of pre-known cluster, prompted the decision of one of the tasks of Data Mining - Data Clustering. Realize the data by identifying the cluster structure allows us to understand how the data looked at the beginning, and how have they changed after anonymization procedure. Analyzing the structure of the cluster can be seen on the data identity.**

***IndexTerms*—Privacy-Preserving Data Publishing, Anonymity, Algorithms, Machine Learning**

I. Introduction

Currently, the volume of data generated grows exponentially[1]. Among these data, there is information that directly or indirectly relates to individuals. Ensuring the protection of information including personal data is a priority and the most important task in ensuring the information security of an organization. It is impossible to imagine the organization without processing information about a person. They are stored and processed data on the members of the management bodies and employees, partners, and shareholders of the persons attending the organization.

Breach of confidentiality to ensure the organization database of personal data security can be a serious incident in information security, which may result in irreparable damage to financial or reputation risks.

The need for personal data protection measures due to the increased technical possibilities also for copying and dissemination of information. Modern people are not physically able to escape from the whole variety of explicitly or implicitly used against it collect technical devices and people data processing technologies.

With the development of electronic commerce and the available means of mass communication have also increased, and the possibility of abuse related to the use of collected and stored information about the person. There were also effectively used by attacker means of integrating and quick processing of personal data, creating a threat to the rights and legitimate interests of the person.

Realizing the importance and value of the information about the person, as well as taking care of respecting the rights of its citizens, the state requires organizations and individuals to provide reliable protection of personal data.

One approach to the protection of personal data is their depersonalization. It can reduce the requirements for the level of data protection and, accordingly, reduce the costs of defense. Therefore anonymization procedure is widely used in practice. For the implementation of these procedures use the data anonymization algorithms, such as k-anonymity, l-diversity, and etc. After application of these algorithms it becomes impossible to identify the entity that owns the personal data.

Anonymity is used to preserve the confidentiality of data.Depersonalization of personal data must ensure not only the protection from unauthorized use, but also the ability to handle them. For this anonymous data should have properties which keep the basic characteristics of depersonalizing of personal data:

- fullness (save all the information about specific subjects or groups of subjects, which had to depersonalization);
- the applicability of (the possibility of solving the personal data processing tasks without prior personalizationthe total volume of the subjects of the records);
- anonymity (inability to uniquely identify the data subjects, derived from depersonalization, without additional information).

One of the most advanced and user-friendly methods of anonymization of personal data is a method of changing the composition or the semantics of data. It is realized by replacing the statistical treatment of results, generalization or deletion of the information.

As a result of this method discards some of the information, which entails a loss of information content of the personal data. The ideal option would be to hold the data without loss of information content of depersonalization, unfortunately, using the above method generalization or deletion part of information is not possible to avoid data loss. So try to reduce (minimize)

the informativeness of losses and increase (maximize) the level of personal data anonymous.Applicability of anonymous data in order to achieve these goals depends on the rest of the information and the preservation of the original structure in the data set after replacing, removing or generalization of the information induring anonymization algorithm works. It is possible to anonymize data is good enough, from the viewpoint of privacy, but they did not apply for further processing, as will lose too much information or too much structure will be distorted. In this regard will be used parameter as the informative. With this option you can compare the initial data set and treated, and on this basis to draw a conclusion on how the information content decreased. Just a lot of it is important to establish the identity of anonymous data source, or will use incorrect data, which will lead to incorrect results and conclusions.

A. Tasks

Next target definition information loss was set in the procedure data anonymization. Having some kind of assessment of loss of information can be compared to the data with a given level of confidentiality and choose anonymous data with minimum losses for future use for statistical or other research purposes. Just such an assessment will select the optimum solution between the loss of information due to confidentiality and depersonalization.

From this goal the following tasks:

- develop a method for determining the loss for the assessment anonymization algorithms;
- comparing baseline data with impersonal to establish equivalence

The first task will allow to compare with each other clustering algorithms. It will be possible to assess how much information is lost in depersonalization of this algorithm, and choose the algorithm with minimal losses for a given level of data privacy.

The second task will enable comparison between the original and anonymous data, and to assess how much data had been distorted as a result of depersonalization. Using false information will lead to incorrect conclusions, so it is important to establish the equivalence of data

B. Relatedworks

Generalized Information Loss (GenILoss). This metric captures the penalty incurred when generalizing a specific attribute, by quantifying the fraction of the domain values that have been generalized. This metric is based on the concept that data cell values that represent a larger range of values are less precise than the ones that represent a smaller range of values. The downside is that we can not evaluate the data structure. Is similar to the original data preserved. [2]

ILoss is a data metric proposed in Xiao and Tao to capture the information loss of generalizing a specific value to a general value vg:

$$ILoss(vg) = (|vg| - 1)/|DA| \quad (1)$$

where | vg | is the number of domain values that are descendants of vg, and | DA | is the number of domain values in the attribute A of vg. This method of estimating the loss of information content also does not allows you to track changes to data structures, and requires the parameter wi which is an indication of the weight of the penalty attributes Ai of vg. The value of the fine can be set by expert assessment that the large amounts of data is difficult computable. [3]

Entropy approach can be used to estimate the amount of information in a data set. With the help of entropy can be estimated descriptiveness of each feature and find all the information content of the total data set. A distance of Kullback–Leibler can be interpreted as the magnitude of the loss of information when replacing the true distribution of the distribution of anonymous data. [4]

II. CLUSTERING DATA

Clustering data - Grouping similar objects - is one of the in the field of data analysis and Data Mining fundamental problems. List of application areas where it is applied, wide. We are interested in clustering algorithms from the point of view of their application in Data Mining.

Clustering in Data Mining acquires value when it acts as one of the data analysis stage, the construction of a complete analytical solution. Analysts often easier to identify groups of similar objects, examine their characteristics and build a separate group for each model, you can create a common model to all data.

Clustering is designed to partition a set of objects into homogeneous groups (clusters or classes). If the sample data is represented as a point in the feature space, the clustering problem is reduced to the definition of "condensations points." The task of clustering refers to a broad class of problems of learning without a teacher. This means that the studied set of data classes are not predetermined in advance, as opposed to classification problems.

The following objectives before clustering:

- data compression;
- novelty detection;
- understanding of data by identifying the cluster structure.

If the original sample is excessively large, it is possible to reduce it, by leaving some of the most typical representatives from each cluster. By reducing the amount of data being processed, reduces the number of tools and resources necessary for their treatment. The saved resources can be spent on the protection of personal data.

Novelty detection. Objects that fail to connect to any of the clusters are allocated. Such atypical objects are potentially vulnerable point for data personalization. Therefore, these objects should be treated in a special way and try to get rid of them.

With output some data structure, we get a generalized view of the data thereby reducing the amount of information that must be analyzed to identify the difference between the original data and impersonal. Comparing the structure of the data processing algorithms to anonymization and after, you can

see whether the data sets are identical, and decide on the further use of anonymous data. After all, the use of anonymous data, which are very different from the original, will give incorrect results and conclusions. In the end we will not achieve the purposes of the assembly and processing of personal data, which were placed before the initial data set. The identification and subsequent analysis of the cluster structures, and will be the main objective of this work.

Often you have to depersonalize high-dimensional data containing hundreds of thousands of records (objects), which in turn describes numeric and categorical attributes. All this complicates the analysis of the data. Analytics is very difficult process. It hard to compare a large amount of data besides dissimilar. And while judging the loss of information content as a result of anonymization algorithms.

In this paper will be use the cluster analysis in order to get an idea of the distribution of objects to form an understanding of data by identifying the cluster structure. Splitting the sample into groups of similar objects, we simplify the further analysis of the data. It is easier to work with these groups to study their features than working with the whole set.

III. Evaluation Informative Value Loss

In this work it was mentioned that how data anonymization algorithms work, and found that the lost or generalized piece of information as a result of their work. This leads to a loss of information content. It is evaluated these losses, not to get the output of useless data anonymization algorithms, or comparing several sets of anonymized data to select the most informative data for a given level of anonymity.

To identify a cluster data structure was chosen algorithm DBSCAN because it is able to identify clusters of arbitrary shape. Its working principle is well-suited for clustering of personal data. The advantage of this algorithm is that clusters can allocate arbitrary shape. The algorithm evaluates the intracluster and intercluster distance, and on the basis of the allocated clusters.[6]

A. Algorithm

The essence of the algorithm is as follows, using Machine learning for clustering data. Clustering algorithm assigns each object class label, to which it carried that. If the object could not be attributed to any class, the algorithm marks it as noise. Then look at the distribution of objects in classes, but rather how many objects are contained in a particular class. Then subjected to data anonymization procedure. The resulting data set is subjected to clustering again. Then see the distribution of objects in classes. Check distribution to depersonalization and after the existence of the correlation between them.[7]

This evaluation gives us an idea on how similar these datasets. Based on this can be seen the loss of informativeness. Decide whether you can use anonymous data for further processing.

Algorithm 1 EVALUATION INFORMATIVE VALUE LOSS

```
Initialize Data {reads documents with Personal Information of objects}
Initialize X1:= Data[feature selection] {These data source}
Initialize X2:= X1 {These data will be anonymized}
Normalization (X1)
Clustering (X1) {Fit DBSCAN algorithm train dataset}
Initialize lable1 {cluster structure of X1}
Depersonalized (X2)
Normalization (X2)
Clustering (X2)
Initialize lable2 {cluster structure of X2}
Count the number of instances of each class
Calculate the correlation between the distribution of classes
```

IV. Experiment

In the work repository with data from the following algorithm have been selected[5]. Will not be used to investigate all the attributes, but only the following feature: sex,age,race,marital_stat,education,native_country,workclass,occupation. They can be classified as personal data. With them it is possible to identify a person, if it was still the object name. Consequently, these data can depersonalize, and used for further processing.

Since this set of data has numeric and categorical attributes, they should be pre-treated. Categorical attributes must be coded by assigning to each possible value of the numeric indication own label. Then, should conduct the normalization calculating the mean and standard deviation. After normalization can begin to cluster.

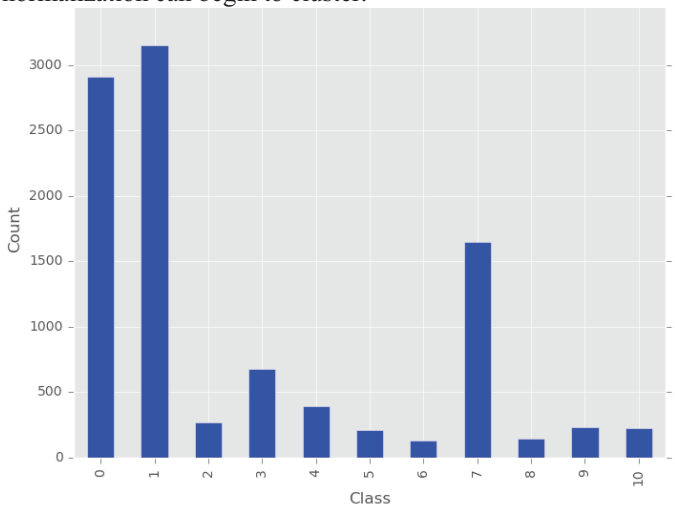


Fig. 1. The cluster structure of the training data

Now it needs to depersonalize the data and to cluster them. To compare the loss depersonalization need multiple sets of data with varying degrees of generality. Make 3 sets of data, where the attribute "age" will be divided into 2 range [0; 50) [50; 100); 8 ranges [0; 12) [12; 25) [25; 37) [37; 50) [50; 62) [62; 75) [75; 87) [87; 100); 16 ranges.

After clustering, we obtained the following data distribution after depersonalization classes.

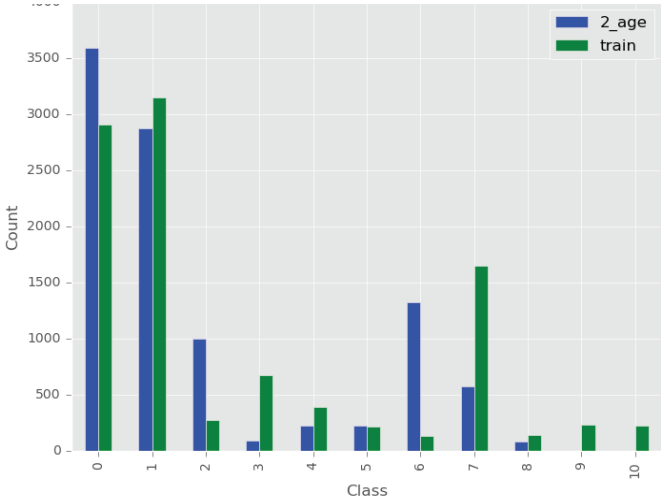


Fig. 2. The cluster structure of the training data and 2_age anonymised data

The figure shows that how the number of classes fell after anonymization. The figure shows that clustering structure before and after the depersonalization differs. In the future, we try to estimate how much different these structures and determine the loss of data during depersonalization.

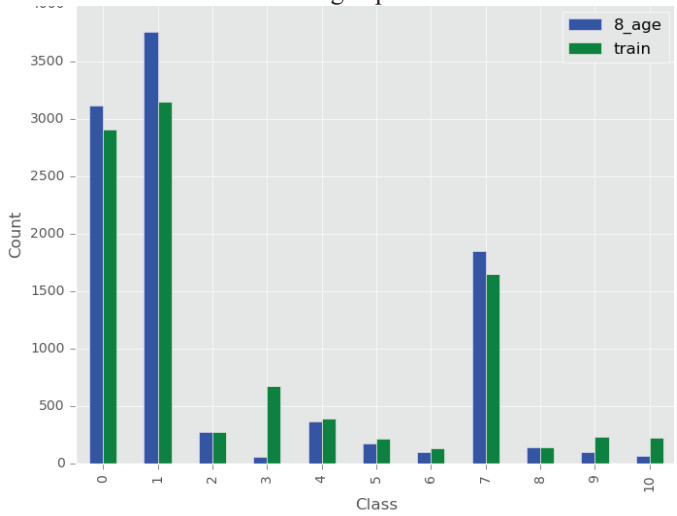


Fig. 3. The cluster structure of the training data and 8_age anonymised data

After clustering anonymous data whose age was divided into 8 ranges can be seen that the number of clusters has become equal to the number of input data clusters. This means that the data have a great similarity between them than before It is to be expected because it is working with more detailed data than the partition age only two ranges.Also, the distribution of the number of objects in each cluster has a similar appearance

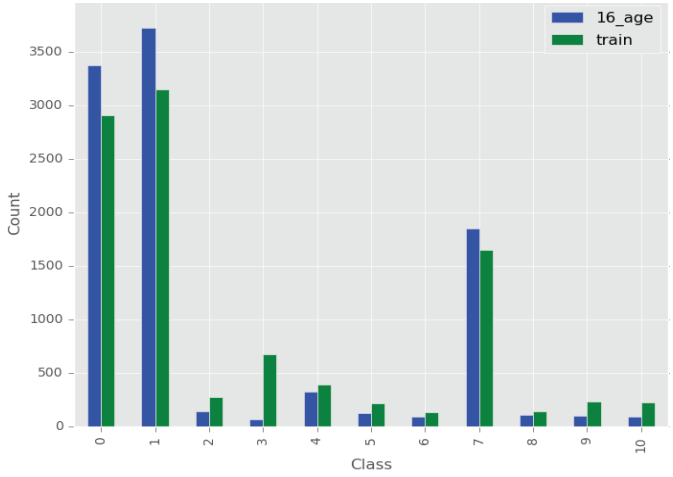


Fig. 4. The cluster structure of the training data and 16_age anonymised data

A. Methods for estimating the loss of information

Now we have to determine how the data will be compared to further losses in the process of depersonalization. We distinguish three ways of evaluating information loss:

- Comparison of the number of clusters
- Transitions between clusters of objects
- Pearson's correlation coefficient of cluster structures

Consider each of the methods in detail. The first and easiest is to compare the number of clusters before and after anonymization. If we generalize the depersonalization of the information about the object, because this reduces the distance of n-dimensional Euclidean space. In this regard, fusion can occur in one of several clusters. Thus it is possible to assess the loss of information, the less is left after anonymization of clusters, the more we lose information. This method is based on the fact that in the beginning there was information about objects that have been allocated, let to three clusters, and output objects have become so similar that were owned by the same cluster. It is reasonable to note with three different classes of objects, we have more information than when we have only one class of objects.

The second method contains a rule: when clustering each object is assigned a label of some class to which it belongs, then it is divided the set of objects to depersonalization in the subset. Each subset contains only one class of objects. As a result, we obtain a partition of the original objects. Later, after the depersonalization and clusteringeach object is assigned a new label of the new class. Again, make partition but subsets that we received in the previous step. And because of this partition, select the subset with the largest capacity. This power is divided by the number of objects in a subset of the original data. And so it goes for each subset, we summarize obtained relationship, and this sum is divided by the number of objects in the original data. As a result, we get a ratio α which lies in the range from 0 to 1, and shows how the structure of similar anonymous data from the original structure. Losses are expressed as in the depersonalization of data$(1–\alpha)*100\%$

$$\alpha = \sum_{i \in A} \frac{max\left(\{|Y_j| : Y_j \in X_i\}_{j \in B_i}\right)}{|A_i||B_j|}, \text{ where } \alpha \in [0; 1]$$

$$A - the\ set\ of\ indices\ of\ the\ set\ X,$$

$$B_i - a\ subset\ of\ the\ set\ of\ indices\ X_i, i \in A$$

$$X = \bigcup_{n \in A} X_n \qquad X_n = \bigcup_{k \in B_n} Y_k, n \in A$$

$$X_n \cap X_m = \emptyset, \forall n, m \in A : n \neq m$$

$$Y_k \cap Y_p = \emptyset, \forall k, p \in B_n : k \neq p, n \in A$$

Fig. 5. The calculation of the remaining objects in a class

TABLE I. SIMILARITY COEFFICIENT DATA

Data	2_age	8_age	16_age
α	0.8069	0.8753	0.9156

It is not difficult to notice that the less impersonal data, the smaller than the loss of information. The biggest losses, while depersonalization of all ages into two groups and family status.

The third method is based on a correlation between cluster data structures before and after anonymization. In this study, the Pearson correlation coefficient was chosen to estimate the correlation. When do you need an objective measure, indicating the presence or absence of a relationship between the variables and measuring the intensity of this connection, using correlation coefficients. The correlation coefficient was proposed as a tool with which you can test the hypothesis of dependence and measure the strength of dependence of two variables. In this case acts as a variable number of objects of a cluster before and after anonymization procedure.

This evaluation gives us an idea about how much data are similar to each other. Based on this can be seen the loss of information content, the smaller the correlation coefficient, the greater the loss of information. You can depersonalize data in several ways, and selecting those data that have the highest correlation coefficient. And use anonymous data for further processing.

We calculate the correlation coefficient between the original data set and impersonal data. We assume the following coefficient for 2_age anonymised data, 8_age, 16_age

TABLE II. CORRELATION COEFFICIENT

Data	Correlation coefficient
2_age	0.841769
8_age	0.987626
16_age	0.990681

As seen from the table, the less impersonal data, the higher the correlation coefficient. The higher the correlation coefficient, the less the loss of information, which are defined as (1 - the correlation coefficient) * 100%. In the most heavily anonymity, in this case «2_age», the loss is about 25%. And at least of anonymity, in this case «16_age», loss of about 1%.

V. CONCLUSION

In this paper, the existing methods were analyzing for assessing loss of information during data depersonalization, as a result of the analysis of significant disadvantage of these methods has been revealed. The above method does not allow to evaluate the structure of the data, thereby to establish equivalence between the original and anonymized data. We used machine learning to identify the structure of the data for later analysis.

Analysis of cluster structure significantly simplifies the understanding and comparison of large data sets. The advantage of analyzing the cluster structure to a simple estimate of the amount of data loss that primarily evaluated the equivalence of the original set of impersonal data.

Using the method described in this paper, it is possible to assess the similarity of the two sets of data in three ways, and to assess the loss of information during the procedure data anonymization. Using this method, we can evaluate the performance of anonymization algorithm for data loss during data depersonalization and choose the algorithm with minimal losses. With the evaluation of information loss can be to find the optimal solution between privacy and the remainder in the data information.

REFERENCES

[1] Big Data, for better or worse: 90% of world's data generated over last two years Web:https://www.sciencedaily.com/releases/2013/05/130522085217.htm

[2] A Systematic Comparison and Evaluation of k-Anonymization Algorithms for Practitioners, Transactions on data privacy 7 (2014)

[3] Information metric sacm computing surveys, vol. 42, No. 4, Article 14, Publication date: June 2010.

[4] Kullback S., Leibler R.A. On information and sufficiency, The Annals of Mathematical Statistics., 79-86, 1951

[5] UC Irvine Machine Learning Repository Web:http://archive.ics.uci.edu/ml/datasets/Adult

[6] Han J.. Efficient and Effective Clustering Methods for Spatial Data Mining, Proc. 20th Int. Conf. on Very Large Data Bases, 144-155, 1994

[7] Nikolić, D; Muresan, RC; Feng, W; Singer, W. "Scaled correlation analysis: a better way to compute a cross-correlogram" European Journal of Neuroscience, 1–21, 2012

[8] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu Institute for Computer Science, University of Munich, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, 1996

Big Data: Potential and Indicators

Makrufa S. Hajirahimova
Institute of Information Technology of ANAS
Baku, Azerbaijan
makrufa@science.az

Aybeniz S. Aliyeva
Institute of Information Technology of ANAS
Baku, Azerbaijan
aybeniz63@rambler.ru

Goychak F. Mahmudlu
Qafqaz University
Baku, Azerbaijan
goychak_mahmudlu@yahoo.com

***Abstract*— In this article the issue of big data is dedicated, the essence of big data is described which is considered key motivating force of the information society. Big data incorporates along the volume of collected information, as well as storage, computing technologies, combines and services. Big data has become one of the rapidest developing spheres of the information technologies and has great potential and value in many fields. Philosophy and economic potential of big data has been explained briefly. In order to assessment of big data current situation have been analyzed according to some indicators.**

K*ey Words* — big data; big data indicators; big data potential; big data philosophy

I. INTRODUCTION

The end of the twentieth century and the beginning of the third millennium is characterized by the emergence of the information society. In this society information has become an important resource as a particular type of goods from the perspective of interests of commercial, private and state. In the past decade the increase of information and communication facilities and ease of accessibility of information has led the increasing of data stream, the volume of data volume has growthed in geometric progression every year. Such "explosion" of information has become the serious problem one of the main challenges of the world of the twenty-first century- named Big data.

The rapid growth of digital information and its value has put new demands to available processing technologies of data and withdrawal of useful information. In response to these challenges exposed to a big information flood and big data projects have been processed Google, Oracle, IBM, Microsoft, Amazon, Facebook organization and so on discovered useful information with the help of different algorithms, technologies. Thus, since 2004 processed large scale data in real-time mode and creation of technological platforms such as MapReduce, Hadoop, NoSQL were organized from more depth analytical tools have risen the value of data a bit [1].

Currently, Big data has become one of the driving force of the development of information technology. Already, in the USA, many other western countries, some developed and developing countries this field began to develop rapidly. In the near it is inevitable future economic and political competition between the countries will be based on using of Big data's potential opportunity. In other words, the investigation and application of large scale data will be necessary in order to increase of the competitiveness ability of any country [2]. That's why, the distribution of economic potential of Big data according to available technologies, fields, countries and measuring of value is an actual issue.

II. THE PHILOSOPHY OF BIG DATA

Big data is increasing as a field of information technology, service, and science, and so too is the need for its intellectual comprehension and explanation from a theoretical, philosophical, and societal perspective.

The Philosophy of Big Data is the section of philosophy concerned with the foundations, methods, and implications of big data. the definitions, meaning, conceptualization, knowledge possibilities, truth standards, and practices in situations involving very-large data sets that are big in volume, velocity, variety, veracity, and variability. Three aspects are considered: what might compile a Philosophy of Big Data, how the disciplines of the Philosophy of Information and the Philosophy of Big Data are growing, and an example of the Philosophy of Big Data in application in the data-intensive science. Data-intensive Science is computationally-intensive science involving very-large data sets that may require data science computing techniques for comprehensive high-dimensional modeling, observation, and experimentation [3].

The Philosophy of Big Data is evolving into a discipline at two levels. One is internal to an area as a generalized articulation of the concepts, theory, and systems that include the overall conduct of big data science. Data science contains big data, and is described as a broader discipline that applies techniques and theories from mathematics, statistics, computing, and information technology, for instance machine learning, to uncover patterns in data from which predictive models can be improved. The other is external to an area; as a ocassion of the influence of big data science more broadly on individuals, society, and the world. Methods, tools, and concepts are estimated at both the level of industry practice theory and social affect.

Overall a Philosophy of Big Data might profitable in conceptualizing and apprehending big data science as a service practice, and also in transitioning to data-rich futures with human and data entities more effectively co-existing in reciprocal surplus and cooperation [3].

III. BIG DATA ECONOMIC POTENTIAL

Growing of volume, speed, diversity and value has resulted to significant changes in transition direction to the model that managed with data in socio-economic activity paradigm. So, Big Data began to play an important role in the creation of social impact, competitive advantage and innovation field. Therefore, Big Data can be regarded as an underlying asset for all sectors, organizations, countries an regions.

Big Data has mainly great value and potential in 5 fields: health care in the United States; public sector administration in the European Union; retail in the United States; global manufacturing; and global personal location data. In 2010 40% of GDP according to the world has been created by these 5 field [4-5]. In 2020 the volume of useful data is expected to be 16 zetabite. In society, its citizens and all business sectors it is reality using of this information for creating of value.

Now, big data has become a large volume invested field all the world. Compared to 2013 in 2014 the number of companies which performed projects relate to big data increased by 125%, the market volume 45%. According to valuation of SNS Research organization only in 2015 $ 40 billion was invested to Big Data. The growth rate of this investment is expected to be 14% in the next 5 years [6-7].

In 2020 in Europe the GDP of big data and open data is expected to increase 1.9 %(206 billion euros) [4]. It is equal to the overall economic growth in one year of all EU countries. In 2020 is anticipated that Big data will organize the contribute of trade to GDP 23%, manufacturing sector 22%, finance and insurance sector 13%, state administration 12%, , information and communication sector 6%, healtcare an social welfare sector 5% [5].

The expansion of using of Big Data will be postive impact on employment and will lead to the creation of many new jobs. According to source of Gartner in 2015 in this field in the accordance with IT at the international level additional 4.4 million jobs will be created. Near to the 1.9 million jobs will be belong to the USA [8-9]. In 2017 is projected 3.75 million jobs will be created in EU [10].

Big Data is already making a significant contribution to the UK economy. The Centre for Economic and Business Research estimated that by 2017 Big Data could contribute £216 billion and generate 58,000 new jobs in the UK and Ireland [11].

IV. BIG DATA MEASUREMENT AND INDICATORS

Initially, big data have occured as the terms described large-scale data that cannot be processed by conventional methods. In recent years, big data incorporates along the volume of collected information, as well as storage, computing technologies, combines and sevice.

According to evaluation of Cisco company in 2015 the volume of mobile data traffic was 3.7 exabytes per month. In 2020 it is expected will be 30.6 exabytes [12].

According to research of EMC company in 2020 the volume of data will be increase up to 40 trillion Gbytes, digital data will be 5,200 gigabytes per person in the world. [13].

Although, the new emergency of Big Data it is already stabilized technologies sphere. It has spreaded in many areas of the business and began to play an important role in the development of companies. According to evaluating of IDC company the growth rate of big data is more than 6 times as a whole Big Data market. During the years 2013-2017 the compound annual growth rate (CAGR) of Big Data market is expected 27% and total cost will be $ 50 billion [14].

Big Data Indicators can be useful in assessing the progress and impact of big data-related growth and performance in the information society (and innovation economy). Generally accepted indicators system for measurement of big data is not still available.

In 2014 initial indicators have been proposed for measurement of big data in Europe market by IDC company. These indicators are as follows [5]:

1. Number of data workers
2. Number of data related companies
3. Revenue of data related companies
4. Data market size
5. Data workers skill gaps
6. Citizen's data

The market value of Big Data is measured in accordance with revenues from the sale of hardware, software and ICT services.

The Innovation Institute at MassTech (Massachusetts Technology Collaborative) has identified eight key indicators that summarize Massachusetts' competitive position in big data and the expansion of the big data ecosystem. These indicators are [15]:

1. Number of big data and data-driven related companies
2. Volume investments in data-driven and big data companies
3. Number of data-driven research centers
4. Number of big data related meetup groups
5. Number of big data related Patents
6. Data science related programs
7. Big Data Projects received federal investment
8. Number of data-related STEM (Science, Technology, Engineering and Mathematics) fields graduate

In scientific sources the number of published documents related to big data can be viewed as an indicator of the big data measurement. Big data research has begun in 2001 with one published document in Scopus [16]. In recent years, published big data documents have increased, in 2014 only in this base the number of published big data documents have reached to 3472. In order to number of these documents the USA, China, Germany, Great Britain and India are the first place [17].

Let us, consider the measurement of Big Data market according to following indicators.

A. Number of data workers and companies

In order to leverage the potential of Big Data, a key challenge is to ensure the availability of highly and rightly skilled people.

According to investigation result of IDC company in 2014 compared to other countries in USA a number of data workers (10457 people) and a number of data related companies (277821) is quite high. In EU a number of data workers

reached 6102 people, a number of data related companies were 243610, in Japan a number of data workers were 3344, a number of data related companies were 95919, in Brasil a number of data workers were 1031, a number of data related companies were 34840. At the same time the value of data market was €103,935 million in the USA, in EU was €50,454 million pounds, in Brasil was €5,289 million, in Japan was €22,228 million pounds [18].

B. Data workers skill gaps

The digital skills gap is one of the most urgent policy challenges facing the developed, and also developing countries. Unlike conventional analytics, mining big data requires an extremely diverse set of skills like data visualization, statistics, machine learning, deep business insights and computer programming. Estimates shows that the US alone faces a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million analysts and managers to analyze big data and make decisions based on those findings [11].

The digital skills gap is one of the most critical challenges for UK too. With recent estimates the UK is already losing a potential £2bn per year 28 from unfilled roles requiring digital skills. The techUK survey of tech companies conducted in October 2014 suggested that 93% of tech companies believe the digital skills gap affects their commercial operations and talent acquisition. Tech Partnership have predicted that by 2020 there will be 56,000 Big Data jobs in the UK, a rise from 21,400 in 2013 [19].

C. The size of Big data market according to regions and countries

Big data is increasingly attracting attention of states which is regarded as a strategic resource in the fields such as science, health, industry, business and etc. Today states have been aimed at allowing the use of large volumes of data to increase the efficiency of decision-making processes, activity of organizations, creation of new services, generation of new ideas and etc [20].

Recently, Big data was applied in USA compaines, now, the other countries of the world interested in this field. In a number of countries are given full support to applying of big data technologies and solving problems in this field at the state level. According to the information of IDC in 2014 according to software, services and facilities of Big Data sphere 45% of the market fell into Europe, Near East, Asia (excluding Japan) and African countries' share.

Also (according to the survey of CIO's) Asia and Pacific ocean countries has a high rate of development rate in the field of analysis, secure storage and cloud technologies. Only in 2014 in the Asia and Pacific ocean regions the revenues of big data services amounted 290 billions relate to market [21].

Latin America countries have decided in the 2 nd places in order to the volume of investment for the development of Big Data.

According to evaluation of Cisco company in 2014 the majority of data was accounted in Northern America (1,32 zb.) and Asia (1,36 zb.) [21]. Middle East and Africa, Eastern and Central Europe will be the greatest growth rate of big data market (Tabl. 1).

TABLE I. GROWTH RATE OF BIG DATA ACCORDING TO REGIONS

Regions	The volume of Big Data		Growth rate
	2014	2020	
Northern America	1,32 zb	3,89 zb	20%
Western Europe	0,60 zb	1,87 zb	20%
Eastern and Central Europe	0,24 zb	1,02 zb	28%
Latin America	0,23 zb	0,82 zb	23%
Middle East and Africa	0,09 zb	0,70 zb	40%
Asia	1,36 zb	4,77 zb	24%

In order to IDC investigation in 2012 the volume of the information was 898 exabytes in USA. In 2020 it is predicted that the volume of information will be 6.6 zetabyte. According to research of EMC company in 2014 the majority of data was accounted China, India, Germany and Japan (Fig. 2) [22]. Digital data was 664,5 gigabytes per person in China, 255,5 gigabytes in India, 3893 gigabytes in Japan, 1045 gigabytes in Brasil, 2839 gigabytes in Germany, 1059 gigabytes in Russia, 789,5 gigabytes in Mexico.

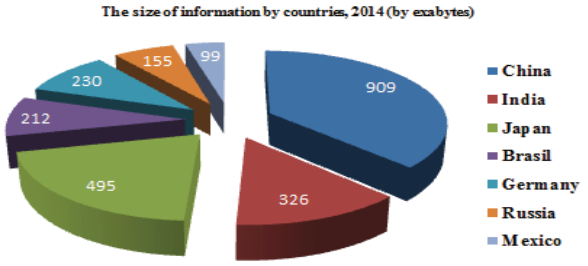


Figure 1. The volume of information by country

According to analyts' prognosis total amount of big data market will be grow result of wide application of big data technologies in developing countries recently. For example, in 2014 the volume of the collected information amounted 40% in developing countries. According to forecast of EMC company in 2017 the structure of Big Data market will be change and in 2020 the share of developing countries will be more than 60%. In China (8060 exabytes), India (2800 *exab*ytes), Brasil (1600 exabytes) and Russia (980 exabytes) the volume of information will be grown rapidly and will account for the majority of data by 2020. [23].

Thus, analysts' forecasts of CIO and EMC company in recent years the developing countries will be active develop market of Big Data technologies.

D. Distribution in accordance with technologies

The technologies that used for the collection and processing of Big Data can be divide to 3 places: software (NoSQL, SQL, Infrastructure Software), hardware (compute, storage and networking) and services(professional and cloud services).

The most widely technologies for software contain SQL, NoSQL, MapReduce, Hadoop, SAP HANA.

In order to questionnaires; results of T-Systems 30% of companies that participated in the survey preferred to SAP, HANA, Oracle's "in-memory" platforms, 18% to NoSQL platforms, 15% to Splunk və Dell companies' analytical platforms. 11% of companies preferred to Hadoop/MapReduce products fo solving the problems of big data [24].

According to information of IDC the volume of big data world market amounted to 27.3 billion dollars in 2014 and 21.3 billion dollars in 2015. The great part of Big Data revenues were obtained from infrastructure segments (Fig. 2). Their share in total revenue was 51,6%, the share insoftware infrastructure segments was 23,7%, the share of service segments was 24,7% [9,14].

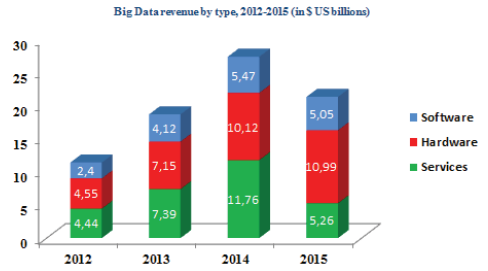


Figure 2. Distribution of Big Data according to segments

According to information of IDC company related to application of different fields of the industry, in 2019 the volume of Big Data technologies and services will be increase to $ 48.6 billion in the world [9,14].

E. Distribution according to fields

Approach to Big Data began to rise in order to sources as mobile device, internet, social media, sensors, log file and so on, in most fields (commercial, banking, management, science, healthcare), etc began to apply widely.

According to estimation of IDC consulting company in 2014 discrete production associate with big data amounted to $ 2.1 billions, $ 1.8 billion ito bank operations, $ 1.5 billion to production processes. Fields as securities and investment services (average annual growth rate of 26%), banking (average annual growth rate of 26%) and media (average annual growth rate of 25%) has a higher growth rate [9,14].

According to results of survey of Tech Pro Research company, Big Data has broadly spreaded in the telecommunications sector (58% of requests), IT engineering (surveys, 45%), manufacturing, transportation, financial and state-owned enterprises. In order to the survey healthcare and education are the least spread areas of Big data [25].

According to the forecast of Bain & Company's Insights Analysis company in 2020 the companies of business sector will be the greatest sector, energy sector will be the highest growth rate of Big Data market.

F. According to distribution of vendors

Now Big Data has become an integral part of developing of many companies as one of the main directions of IT. According to poll results of Accenture company in 2014 more than 50% of many companies that used Big Data technologies incurred expenditures for Big Data amounted 21% to 30% in the budget [24].

In 2014 Dell company earned $ 582 million from the sale of big data hardware. At the same time original design manufacturers gained $5.8 billion from big data hardware [24].

In 2014 according to the services segments revenues of Big Data of IBM company were $ 624 million [9].

According to the information of Wikibon in 2015 in accordance with the revenues of Big Data services of IBM company, software and hardware segments increased 19% to nearly 2104 million dollars compared to 2014. In SAP company revenues increased 24% to $ 890 million, in Oracle revenues increased 19% to $ 745 million. Compare with 2014 in 2015 the revenues of Teredata company rose 50%, the revenues of Splunk company rose 43%, the revenues of Microsoft company rose 37% (Fig. 3) [26].

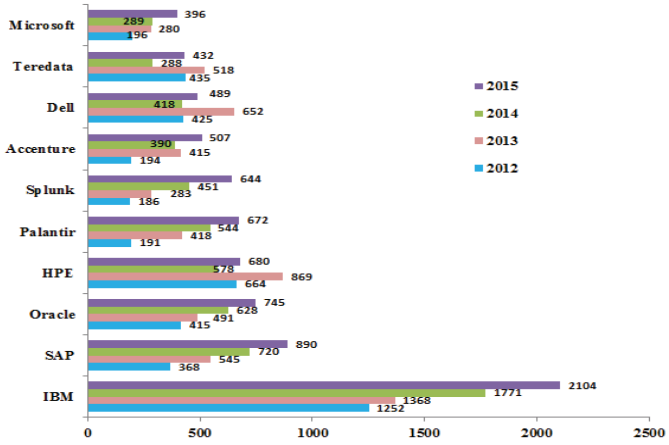


Figure 3. Big data vendors revenue by 2012-2015 ($US millions)

CONCLUSION

Investigations according to the topic give reason to say that big data has great economic potential. It is no coincidence that big data estimated as a new economic asset at the World Economic Forum in 2012 [27]. There are optimistic forecasts according to growing share in GDP, job creation, increasing investment and so on. Research shows that, indicators system of big data according to the world is not yet completely formed. Therefore, the evaluation of data still remains a problem.

REFERENCES

[1] R. M. Alguliyev, M. S. Hajirahimova, ""Big Data" phenomenon: Challenges and Opportunities", Problems of Information Technology, 2014, no 2, pp. 3-16.

[2] X. Jina, W. W Benjamin, X, Chenga, Y. Wanga, "Significance and Challenges of Big Data Research", Big Data Research, 2015, vol. 2, no. 2, pp. 59–64.

[3] M. Swan, "Philosophy of Big Data: Expanding the Human-Data Relation with Big Data Science Services", 2015 IEEE First International Conference on Big Data Computing Service and Applications (BigDataService), 2015, pp. 468-477.

[4] S. Buchholtz, M. Bukowski, A. Śniegocki, Big and open data in Europe: A growth engine or a missed opportunity? www.microsoft.com/global/eu/Report.pdf

[5] European Big Data Value Strategic Research & Innovation Agenda - http://www.nessi-europe.eu/Files/SRIA__v099%20v4.pdf

[7] The Big Data Market: 2015 - 2030 - Opportunities, Challenges, Strategies, Industry Verticals and Forecasts, http://www.researchandmarkets.com/reports/

[8] Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015, http://www.gartner.com/newsroom/id/2207915

[9] http://www.statista.com/topics/1464/big-data/, 2016.

[10] J. M. Cavanillas, E. Curry, W. Wahlster, "The Big Data Value Opportunity", http://download.springer.com/static/pdf/

[11] A.Gupta, "Making Big Data Something More than the "Next Big Thing"", The Global Information Technology Report 2014: Rewards and Risks of Big Data, pp. 87-93, http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf

[12] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper http://www.cisco.com/c/en/us/solutions/collateral/service-provider/

[13] By 2020, there will be 5,200 GB of data for every person on Earth- http://www.computerworld.com/article/2493701/data-center/

[14] http://www.idc.com/getdoc.jsp?containerId=prUS40560115

[15] Massachusetts Big Data Indicators 2015, http://massbigdata.org/assets/Uploads/Final-Big-Data-Report-2015.pdf)

[16] G. Halevi, "The Evolution of Big Data as a Research and Scientific Topic Overview of the Literature", Research Trends, Issue 30, September 2012, https://www.researchtrends.com/wp-content/

[6] https://www.reportbuyer.com/product/2164339/

[17] R.Aliguliyev, N. Ismayilova, Bibliometric analysis of Big data Research, "Big data: opportunities, multidisciplinary problems and perspectives", I Republican scientific-practical conference, Baku, 25 February, 2016, pp.58-60.

[18] F. Lupianez-Villanueva, "Big data in Europe: new environment, new opportunities", http://cdn.bdigital.org/PDF/BigDataCongress15/

[19] Submission to the inquiry on 'The Big Data Dilemma' House of Commons Science and Technology Select Committee, September 2015, https://www.techuk.org/

[20] M.S. Hajirahimova, A.S. Aliyeva, Big data strategies of the world countries, National Supercomputer Forum (NSKF), 2015, Russia, Pereslavl.

[21] Revenue from big data services market in the Asia-Pacific region, http://www.statista.com/statistics/496266/

[22] EMC Digital Universe Study: How Large is the Digital Universe? How Fast is It Growing?, April, 2014, www.infodocket.com/2014/04/16/

[23] http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

[24] Analytical overview of the market Big Data, 2015, https://habrahabr.ru/company/moex/blog/256747/

[25] http://www.techproresearch.com/downloads/the-power-of-big-data-and-iot/#ftag=YHF87e0214)

[26] http://www.theregister.co.uk/2016/04/08/wikibon_goes_big_on_big_data/.

[27] Big Data, Big Impact: New Possibilities for International Development, 2012, http://www.weforum.org/reports

Architecture and security tools in distributed information systems with Big Data

Rinat Pavlikov, Roza Beisembekova
Department Computer and Software Engineering
Kazakh National Research Technical University named after K.I. Satpayev
Almaty, Kazakhstan
rinatp@icloud.com, beisembekova.r@gmail.com

Abstract: Nowadays the main areas of application of information systems with large data is very different, but the wide range of applying now limited due to the unresolved many of the problems of information security big data technologies. The solution of security will enable to increase consumer in the information system of large data and on the other hand, to reduce the risks of unwanted retrieve information by combining the technology of data mining and internet intelligence. This article posed the problem of ensuring information security for big data technologies outlined the main approaches to its solution, and considered their specifics also briefly described infrastructure security for big data related safety of information system.

Keywords: architecture, information security, information systems, big data, model of threats to information security.

I. INTRODUCTION

Modern corporations of different sizes, subordination and activities receive a huge amount of data about the present state of their IT-infrastructure. These data need to be processed correctly and promptly to detect information security incidents and to highlight the areas of the IT-infrastructure, that most often exposed to high-risk, with a view to its immediate elimination. The data is generated from the information considered in a particular context, coming from the separate domain controllers, the proxy servers, the DNS servers, information security tools; this information describes the current configuration of network devices, the characteristics of the network traffic, application operation and network services operation, the activity and the specific acts of individual final users as well as this information contains postal correspondence, web content, digital audio and video, business processes data, internal documents and Corporation analytical data over the years of its life.

The volume and heterogeneity of data and related with it activity to be further close monitored and analyzed are very high. The question of their structured, consolidated and visual representation to make well-timed and reasonable decisions in the field of information security management by all assets of the corporation's IT- infrastructure rises very sharply. Constantly increasing volumes of information about information security events, the assets of the IT-infrastructure, their vulnerabilities, users, threats and related information, as well as the need for more rapid receipt of a systematic and in a certain manner analyzed "raw" diverse information for faster understanding of the current situation to guarantee information security has given rise to a known issue of Big Data problem [1].

The totality of the existing technology in this area is called Big Data. It is generally agreed that support technologies of distributed computing environments, including Big Data, are designed without taking into account the information security requirements for them. For this reason, this problem is particularly urgent, but is still far from a complete solution.

The subject of the investigation is the problem of information security for IT Big Data.

The aim of this study is to develop a security architecture in distributed information systems with Big Data.

II. NEED FOR INFORMATION SECURITY OF THE SYSTEMS WITH BIG DATA

A. Requirements for the security of information systems with Big Data

The traditional security mechanisms such as firewalls and anti-virus software installed on the computer are insufficient to protect effectively Big Data. The problem is that all this was created to protect small volumes of static data, i.e. files saved on the hard disks rather than a large stream of information coming from the cloud environment. Security measures should be sufficiently flexible and responsive, which will ensure the uninterrupted operation of data receiving and safety of multiple "access points".

Cloud Security Alliance [2] members named the following main problems of modern information systems, arising from the storage of large volumes of data:

- security calculations in distributed software systems;
- security of non-relational databases;
- safe data storage;
- validating;
- security monitoring in real time;
- data mining and analytics, which maintain confidentiality;

- access control encryption and security of links;
- fragmented access control;
- detailed audits;
- the origin of the data. Security solutions must be certainly focused on monitoring and tracking the sources from which the data come.

B. Statement of the problem of Big Data protection

The lack as of today a systematic approach to ensure the information security of the Big Data processing in real-time constrains the use of these IT.

The task of ensuring the information security of Big Data can be correctly formulated in the following way: to create the complex of models and methods that will make it possible to ensure the security of the Big Data processing in the "hard" real-time in the processing nodes and in the modules of information systems (IS).

Thus, the IS functions are described through the graph, whose vertices are the basic operations, and ribs are the streams between them. The specific content of the executed basic operations and configuration of the streams between them are defined through the application tasks which are carried out by the IS. The intensity of the output streams does not necessarily coincide with the intensity of the input streams.

Under these conditions it is required, firstly, to ensure security of information stream transmitted over the IS, including both conventional key aspects of information security (confidentiality, integrity and authenticity of data), and specific new requirements arising in the context of the data stream characteristics and its processing procedures.

The second group of requirements relates to the algorithmic part of the processing operation: it is required to ensure the confidence in the IS modules implementing separate basic operations, and if a sufficient level of confidence in them is not possible, then the verification of the operation results and the correction of random errors or deliberately introduced tampering.

Finally, the third group of requirements refers to the security of IS infrastructure. The most common current approach to the Big Data processing, regardless of the relevancy of processing requirements in real-time – is creating of a computing resources cluster and software implementation of massively parallel processing of the very large array elements [3]. In the consumer has no his own ITI he can use the service of cloud computing providers.

In this case, the cloud computing is provided by the "infrastructure as a service" (IaaS) model, as the other models are irrational or even impossible in the presence of real-time requirements.

The solved problem in this formulation is one of the key problems that lead to the solution at a satisfactory level of the information security problems of Big Data technologies.

Scientific novelty of the problem to solve is determined by the following factors:

1) previously known problem solving on information security of processing procedures of very large scale arrays of semi-structured data are obtained for the processing conditions in the near real time: all of these solutions are mainly focused on the methodology of "displaying and reduction" (MapReduce) [4] – there are only a few solutions that are not connected into a single cycle of work and devoted to the analysis of problems of information security in real-time or discussing the differences between them;

2) it is necessary to create the set of models of the objects to be protected, taking into account, unlike that formerly known, the operation of tools and mechanisms of information security as the elements of queuing system networks, representing the stream processing in IS;

3) it is necessary to create and use for research the formal model of information security threat, that takes into account not only the operational but also the computational capabilities of the attacker, allowing to formulate correct statements about the security algorithms properties and to obtain their evidence;

4) it is necessary to obtain solutions for task complex on security information, implemented as algorithms of incoming stream processing and internal streams between the basic IS operations.

III. Analysis Of The Distributed Information Systems Architectures With Big Data

The studies presented in [5,6] papers were selected for analysis of the distributed information systems architectures with Big Data.

The architecture of distributed information systems for processing Big Data in education is presented in the work [5]. The distributed IS, which allows the operational management of the process was designed for the automated control system of the educational process of higher educational institution.

The paper [6] presents the development of functional software and architecture of the information system of monitoring. The proposed concept of monitoring of educational institutions on the basis of intelligent technologies is implemented in the form of the information system. The architecture of this information system of monitoring which includes 4 subsystems: subsystem «Surveillance»; subsystem «Estimation and Control»; subsystem «Analysis»; subsystem «Management».

Analyzing the above mentioned architectures it can be concluded that the problems of information security are justified by the following features of the infrastructure of distributed information systems for processing data bulks:

- a plurality of distributed nodes, "Moving of the computing is cheaper than moving of the data" is the key paradigm for "Big Data" processing;

- sharding data: data in large clusters are redundant, with multiple copies, moved between different nodes to increase the performance and fail-safe feature guarantee;

- data access / authentication and authorization: Role-Based Access occupies a central position in the most database security schemes;

- the interaction between the nodes: Hadoop and the vast majority of the decisions (Cassandra, MongoDB, Couchbase, etc.) do not provide the proper level of security while the interaction – they use RPC over TCP/IP;

- the lack of built-in security tools: "Big Data" stack processing almost do not have built-in security tools. In addition to the authorization on the level of service and support of Web proxy, there are no means to protect data storage.

The aforementioned problems are not specific for "Big Data", however, the solutions are subject to them per their nature due to the distributed architecture. In order to add safety assurance aids into the "Big Data" environment, they shall also be easily adjusted to scale along with the distributed carcass.

Any Big Data solution is directed to implementation of a specific task and is unique in its own way. However, irrespective of whether the solution is constructed on free of charge components or commercial technologies, its architecture shown in Figure 1.

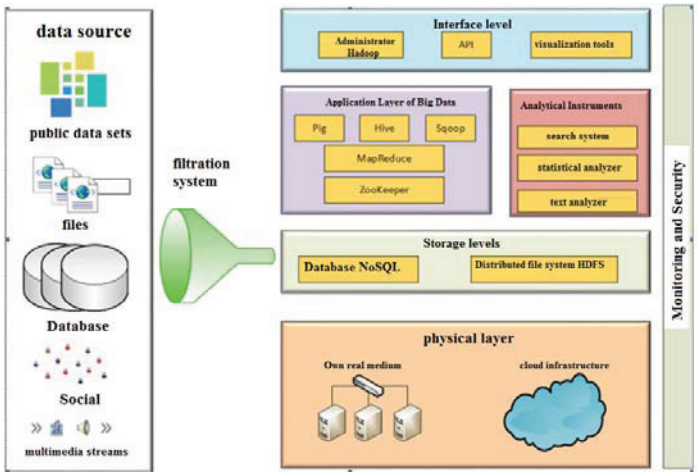


Fig. 1. Generalized architecture software solutions Big Data

The information can enter into the "Big Data" storage system from various internal and external sources. The sources may differ per their types, capacity and data submission rates. Through the agency of the "Big Data" technology stack, the incoming data are consolidated into big data sets, as well called the data lakes.

The filtration subsystem, which is identified as the Ingestion Layer in the English technical literature, serves as gateways for an enormous amount of data entering the "Big Data" system from various sources.

Organization of an effective protection system with the "Big Data" distributed system is a considerably more complex task compared to the conventional systems. The focal difficulty consists in tracking and monitoring of the entire system when log files are formed on multiple locales (nodes), and formation of the final result occurs without any capability on the part of a user.

As minimal safety measures, it is recommended to apply verification of validity at addressing to the system locales:

- protected communication between locales;
- encryption of files in storage;
- service of trusted certificates;
- distributed logging at all architecture levels [7].

IV. Model Of Threats To Information Security

Management by information risks consists in coordinated impact onto objects and subjects of information-computational and telecommunication system infrastructure of big data processing in the direction of elimination of threats, brittleness and risk factors with the purpose of prevention and minimization of possible consequences from their implementation.

The model of information risks can be presented as a weighted dynamic hypergraph (Figure 2) (1):

$$G = (V(S_i, Q_j), P_k(U_k)) \quad (1)$$

where subset of top elements Si = $\{s_i | i = 1, ..., Ns\}$ models threats, subset of top elements $Q_j = \{q_j | j = 1, ..., Nq\}$ – brittleness, multitude of hyperedges $P_k = \{p_k | k = 1, ..., M\}$ – attacks dynamically emerging in time moment t.

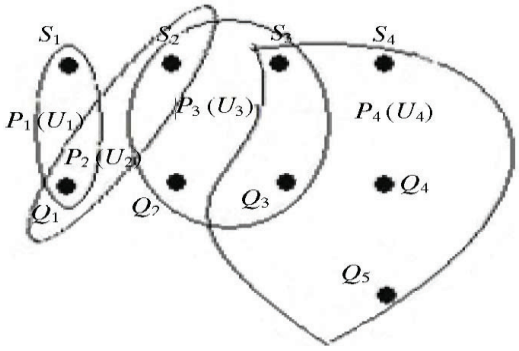


Fig. 2. Hypergraph model of information risks

The dynamic hyperedges combine a subset of top elements-threats with a subset of top elements-brittleness and build attacks as a result of which the information risk R_k emerge and is implemented involving consequences, which can be assessed by the brittleness amount U_k. In the model, the amount of possible brittleness U_k, d, information risk implementation case, is represented by the dynamic hyperedge weight. The assessment of information risk builds definition of the risk level, as well as its comparison with the permissible level at specific conditions of its implementation. For assessment of a risk, it is possible to apply such parameters, as the criticality level K_i of emergence of i-th threat, %, and attack implementation probability S_i, %, due to brittleness of system Q_i.

The criticality level of information risk can be assessed as a ratio of risk implementation number to risk emergence number during the time interval ΔT (2):

$$K_r = \frac{N_{real}}{N_{possible}} \times \Delta T \quad (2)$$

The threat can be assessed as (3):

$$S_i(Q_j) = K_i \times P_a \times 0{,}1 \quad (3)$$

where coefficient 0.1 is applied for reduction of dimensionality at multiplication of variables in percentage.

The information risk R can be assessed as (4):

$$R_i = P_i(VF) \times U_i = (P_i(V) \times P_i(F/V)) \times U_i \quad (4)$$

where $P_i(VF)$ – probability of joint risk emergence *i*-th threat (event *V*) and its implementation (event *F*); $P_i(V)$ – probability of risk emergence due to *i*-th threat; $P_i(F/V)$ – risk implementation conditional probability (event *F*) at occurrence of event *V*; U_i – level of damage from risk implementation (event *F*).

The probability of information implementation can be determined taking into consideration the brittleness factor $P_i(V) = P_j^u \times Q_i$, P_j^u – probability of implementation of *i*-th threat relative to j-th active; Q_i – i-th brittleness magnitude.

Risk magnitude R we can determine as (5):

$$R = \sum_{i=1}^{n} (P_i^u \times P_i(F/V) \times Q_i \times U_i), \quad (5)$$

Management by information risks in automated monitoring systems and dispatching control shall be organized in compliance with the information safety strategy of an eneterprise [8].

V. Architectural Provision Of Information Safety In Systems With Big Data

A. Development of safety provision architecture in systems with big data

One of the options in development of information safety architecture and strategy – to begin from protection of critically important business processes, which affect a business entity. It is necessary to highlight the essential factors of success. If they are known, then one can proceed with development of the information safety architecture from addressing the protection tools of key business processes.

The main structures of safety provision architecture in the systems with big data include the following [9]:

- information and events acquisition unit of tasks;
- unit of modeling of events, processes and attacks;
- unit of tasks assuring decision making.

The unit of tasks for acquisition (gathering) of information and events includes: highly scalable, reliable and multilevel acquisition of information; formation (selection) of languages for internal submittal and processing of data; multilevel correlation of events.

The unit of tasks for modeling of events, processes and attacks includes the following objectives: prediction analysis of safety; multilevel modeling of safety events; modeling of processes and attacks.

The unit of tasks assuring decision making comprises tasks in detection of safety incidents, selection of countermeasures and visualization of safety events in a comfortable for an operator format.

Figure 3 depicts the generic schematics of data streams (flow of data) in an informational system developed based on safety assurance in the systems with big data.

Generic schematics of streams in information system with information safety technology

This schematic diagram presents the overall processing cycle of safety events and information in a developed system. The data stream begins from external sensors building events in various formats. Afterwards, the data on events enter to the data bus via external (protected) informational commutators (ICs – multiplexers). Further on, safety events enter the correlation and processing system via internal ICs, and then in the events, processes and attacks, and decision making modeling systems.

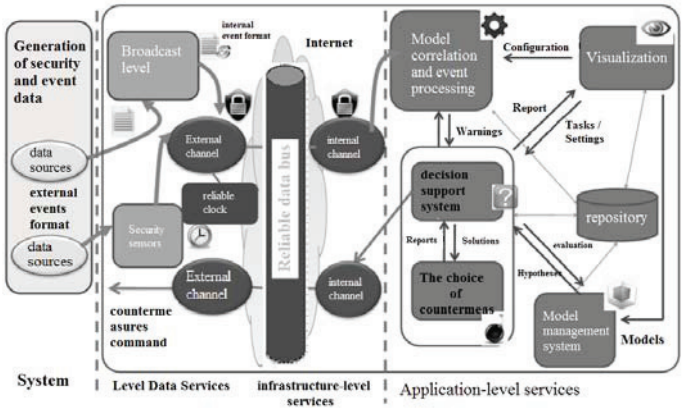


Fig. 3. Generic schematic of data streams in an informational system developed based on safety assurance

Let us consider below examples and current status of studies into two tasks: analysis of safety events based on application of network attack modeling system and construction of system repository.

B. Analysis of safety events based on application of modeling system of network attacks

In the project framework, it is proposed to introduce supplementary functionality into the existing systems – modeling subsystem of attacks, which will enable to expand capabilities and enhance the accuracy in detection of incidents associated with information security.

Since the results of operation of the attacks modeling system cannot be frequently received in the real-time scale, their application in the real-time scale processes is complicated. However, the constructed graphs of attacks maintain their applicability for a sufficient period of time.

Thus, with the framework of events analysis system, it is proposed to use the preliminary constructed attack graphs. These attack graphs can be applied for solution of two basic types of tasks – for prediction of subsequent actions of an

illegal intruder, and for analysis and detection of his previous actions that lead the system to its current status.

It is also necessary to note that for enhancement the efficiency within the framework of attack modeling, there are used not separately current events, but incidents taken into account by means of correlation of separate events.

Thus, the modeling subsystem will analyze not separate events of "host C received package on 80 port from host B" type, but incidents like "scanning of host C by host B is in progress" that will allow more efficient detection of attack graphs including in itself such incidents.

Prediction of subsequent actions of an illegal intruder is performed based on the analysis of the following elements:

- possible targets specified in attack graphs;
- illegal intruder's models, based on which the closest to the real attack graphs were constructed [10];
- attack classes and brittleness used by the illegal intruder [11].

Besides, the results of attack modeling subsystem operation can be the following characteristics:

- "weak" places in the network topology (hosts, through which the biggest number of attack graphs passes [12];
- selected countermeasures enabling lessening of the maximal number of attack graphs probability [13];
- possible consequences of countermeasures implementation taking into consideration dependences of servers [14].

At the present time, theoretical studies concerning application of various standards proceed, including methods of construction of attack graphs that take into consideration the existing brittleness and zero day brittleness, safety strategy, dependences of services, dynamic modeling [15,16], etc., and development of the attack modeling subsystem is accomplished as a basic component of the generic system.

C. Construction of information system repository

The core component of the big data information system is repository or information storage, in which data on events, rules and safety incidents are stored [17].

The task for construction of the repository is an essential one for an information system. The particular significance this task acquires in a critical informational infrastructure, where not only traditional computer infrastructure safety events are taken into consideration, but also safety parameters of a physical level.

For development of the information system repository architecture, there was accomplished the analysis of standards in the domain of events management (Common Event Expression, Common Base Event, XDAS, CIM and others), as well as in regard to the requirements in terms of content and structure of stored data and their processing mechanisms.

As a result, there was delivered a conclusion that the most acceptable standard is the CIM standard.

For construction of the repository architecture, there is proposed a method based on application of the service-oriented architecture (SOA) [18]. All put together, these data bases form the depository data base storage layer. Another functional layer builds a set of various services, which are performed relative to the stored data.

For selection of software and instrumentation tools in construction of a repository, there was performed the DBMS analysis of the following classes: relational DBMS traditional class, XML-oriented DBSM (Base X, Apache X Indice and others) and triplet storages (4store, Big Data, Big Owlim, TDB and Virtuoso). As a result, the selection was made in favor of the Virtuoso [19] program system server, which supportively associates capabilities of all DBMS three classes.

CONCLUSION

Development of methods and models in the presentation domain, information acquisition, storage and processing associated with safety events and which allow implementation of the requirements made to the information systems of new generation, is the pressing scientific task having great state and national economy importance and defining new trends of scientific studies in the information safety domain.

REFERENCES

[1] I.A.T. Hashem et al. / Journal of Information Systems, Vol.47, 2015, pp. 98–115.

[2] Cloud Security Alliance – CSA, Top Threats Working Group (2013). The Notorious Nine - Cloud Computing Top Threats in 2013. http://www.cloudsecurityalliance.org/topthreats.

[3] Meeker D, et al. A system to build distributed multivariate models and manage disparate data sharing policies: implementation in the scalable national network for effectiveness research // Journal of the American Medical Informatics Association, 2015, pp. 1187-1195.

[4] R.Logapriya, J.Preethi, Efficient Methods in Wireless Sensor Network for Error Detection, Correction and Recovery of Data // International Journal of Novel Research in Computer Science and Software Engineering, Vol. 3, Issue 2, pp. 47-54, Month: May - August 2016,

[5] Uvalieva I., Turganbayev E., Tarifa F., Development of information system for monitoring of objects of education on the basis of intelligent technology: a case study of Kazakhstan, Fifteenth International conference on Sciences and Techniques of Automatic Control & computer engineering (STA2014). - Tunis. 2014. - pp. 909-914.

[6] Uvalieva I., Garifullina Z., Soltan G., Kumargazhanova S., Distributed information-analytical system of educational statistics, 6th International Conference on Modelling, Simulation and Applied Optimization, Istanbul, 2015. pp. 1-6.

[7] V. Shchelkanov, "Architecture Big Data", [Online]. Available: http://www.dataved.ru/2014/06/big-data-architecture.html

[8] Zavgorodnii V. Informatsionnye riski i ekonomicheskaia bezopasnost' predpriiatiia // Information risks and economic safety of an enterprise, Moscow: Finakademiia Publ., 2008.

[9] C. Wang et al., "Architecture Support for Task Out-of-Order Execution in MPSoCs," in IEEE Transactions on Computers, vol. 64, Issue 5, 2015., pp. 1296-1310.

[10] Kotenko I., Stepashkin M., Doynikova Ye., Security Analysis of Computer–aided Systems taking into account Social Engineering Attacks // Proceedings of the 19th Euromicro International Conference on Parallel, Distributed and network–based Processing. IEEE Computer Society. 2011. pp.611–618.

[11] Kotenko I., Stepashkin M., Doynikova Ye., Analiz zashchishchennosti avtomatizirovannykh sistem s uchetom sotsio - inzhenernykh atak // Problemy informatsionnoy bezopasnosti. Komp'yuternyye sistemy, 2011, № 3, pp. 40-57.

[12] Kotenko I., Stepashkin M., Kotenko D., Doynikova Ye., Otsenka zashchishchennosti informatsionnykh sistem na osnove postroyeniya derev'yev sotsio - inzhenernykh atak // Izv. vuzov. Priborostroyeniye, Vol. 54, Issue 12, 2011, pp.5-9.

[13] Sommestad T., Sandström F., (2015) "An empirical test of the accuracy of an attack graph analysis tool", Information & Computer Security, Vol. 23, Iss.5, pp. 516-531.

[14] Ingols K., Chu M., Lippmann R., Webster S., Boyer S. Modeling modern network attacks and countermeasures using attack graphs // Proceedings of the 2009 Annual Computer Security Applications Conference (ACSAC'09), Washington, D.C., USA, IEEE Computer Society, 2009, pp.117–126.

[15] Shameli-Sendi A., Dagenais M. // Journal of Network and Computer Applications, Vol.55, 2015, pp. 89–107.

[16] K. A. Farris, S. R. McNamara, A. Goldstein, G. Cybenko, «A preliminary analysis of quantifying computer security vulnerability data in the wild» // Proc. SPIE, vol. 9825, 2016.

[17] Malhotra N., Implementation of Data Marts in Data ware house // International Journal of Advance research, Ideas and Innovations in Technology, vol. 1, no. 12, 2015.

[18] R. T. Tiburski, L. A. Amaral, E. D. Matos and F. Hessel, The importance of a standard security architecture for SOA-basediot middleware, in IEEE Communications Magazine, vol. 53, no. 12, pp. 20-26, Dec. 2015.

[19] Virtuoso, [Online]. Available: http://virtuoso.openlinksw.com

Distributed Information and Control System for Emergencies in Critical Infrastructures

Sagit Valeev, Natalya Kondratyeva
Dept. of Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
vss2000@mail.ru, knv24@mail.ru

***Abstract* — The architecture of distributed information and control system for emergencies in critical infrastructures is presented. The hierarchical structure of data processing is proposed. Application of cloud computing for real time systems is discussed.**

***Index Terms* — large scale critical infrastructures, evacuation, information and control system, large data sets, cloud computing.**

I. Introduction

Modern airports, shopping malls and other similar distributed objects can be considered as critical infrastructures requiring application of wide range of advanced information technologies [1-5]. For example, a system for ensuring the evacuation process in critical situations includes technical means of detection and alarm response, a variety of sensors, computing, telecommunications, and human resources as engineering personnel and security service of the building. In addition, in a critical situation there is a variety of uncertainties that need to be taken into account in the decision to emergencies. Currently, the solution to reduce risks in the emergency is an important task.

New sensor technologies, more advanced real-time models and faster computers are defining the current state of the art in control of industrial processes. In particularly, development of wireless sensor networks for monitoring the state of distributed objects, development of cluster computers with low power consumption for implementation of data processing and control tasks, the availability of cloud services for improving the reliability of data storage allows to realize the concepts of information support and control of distributed large-scale critical infrastructures [6].

One of the urgent tasks of managing large-scale systems in real time is to ensure efficient evacuation of people and property in the event of emergencies. At all stages of this task implementation there are different subtasks associated with the transmission, storage and processing large datasets that can be associated with Big Data [7-9].

Distributed information and control system (ICS) of evacuation usually includes different technical means of detection and alert, typical sets of primary sources of information (fire sensors, video cameras, etc.), means of processing and storage of information, telecommunication channels, and different organizational resources. Different unstructured data streams generated in real time in case of emergency are needed to assess the current state of an object and decision making. The problem of developing an effective evacuation management system is linked to a number of different factors: the heterogeneity of the data from different sources, the large volume of datasets, the possible loss of part of the information infrastructure in the event of effect of different destructive factors (fire, earthquake, etc.). The considered distributed ICS are based on network information technologies, and this is the possible loss of data packages in digital data transmission process, time delays due to the limited bandwidth of communication channels, etc.

It should be noted that a critical situation is usually accompanied by a variety of different uncertainties that need to be taken into account in the decision-making process, which complicates the task management solution in the class of distributed control systems with a reference model.

For example, one of the most common causes of the critical situation in the infrastructure facility is the occurrence of a fire. Commonly the route of evacuation of people and property is determined by a single approved evacuation plan. The plan includes generic path evacuation of people and movement of material resources without taking into account the current uncertainties that mainly influence on the choice of an effective evacuation route.

To increase the efficiency of evacuation management, distributed ICS must be developed in the class of hierarchical control systems [10, 11]. In this connection, it is advisable to develop the architecture of distributed evacuation ICS to ensure the effectiveness of the evacuation plan for the selection procedure, based on analysis of current information about the status of the parameters characterizing the critical situation and the results of modeling different variants of the fire spread and adoption of appropriate management decisions with the use of new information technologies on based cloud services and big data paradigm.

II. Architecture of Hierarchical Distributed ICS for People Evacuation

Architecture of multilevel distributed ICS for emergencies, for instance, evacuation of people in the large-scale infrastructure objects is shown in Fig. 1.

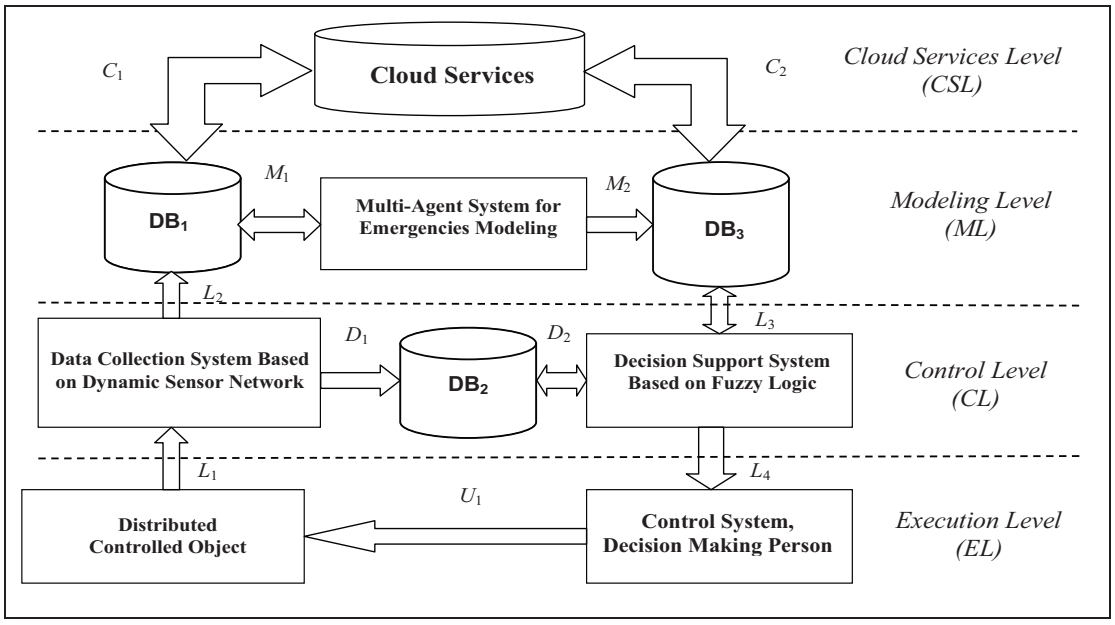


Fig. 1. Architecture of distributed ICS for people evacuation

The computational processes in the evacuation system should be provided by implementing cloud services for storing complex models, data sets and for executing algorithms of modeling, and also for providing decision making processes in real time on the basis of cloud computing.

Let us consider data flows more detailed. Data flows D_1 and L_2 come from data collection system. Data include of measurement of dynamic sensors system L_1 and followed to data bases DB_1 and DB_2. The power of these information flows depends on number and spectrum of sensors, as well as of sampling frequency. Data flow of executive control L_4 for devices, mechanisms and actuators applied during evacuation is transmitted from decision support system based on fuzzy logic. System of executive mechanisms (fire safety system, automated emergency exits, lighted paths etc.) transmits control information U_1.

The power of control information flow depends on number and variety of actuators, as well as of sampling frequency for sensors and frequency of transmitted commands. Information D_2 about the operation of the sensors is transmitted to the DB_2 database.

Decision-making in real time is based on a set of fuzzy logic rules L_3. These rules are stored in DB_3 and are formed in the multi-agent simulation system. This process uses the information M_1 from the database BD_1. Data sets of simulation results are transferred to the database DB_3 (Fig. 2, 3).

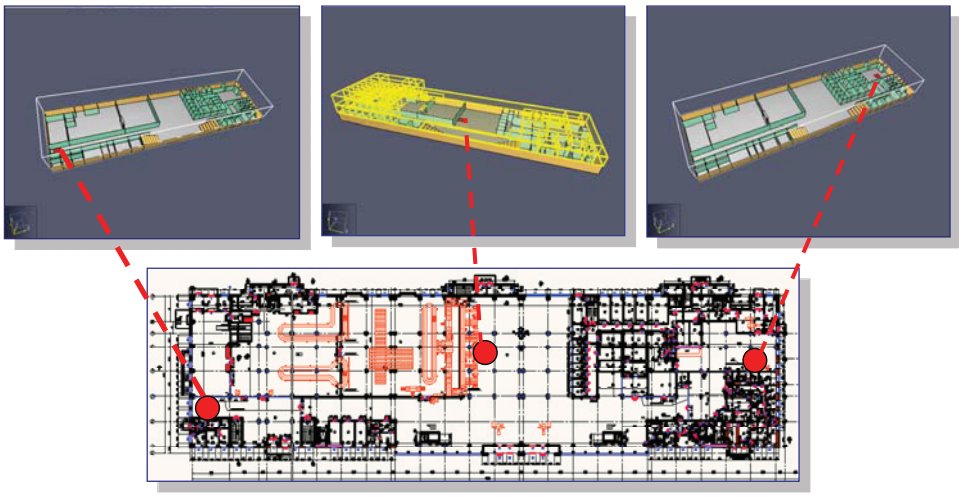


Fig. 2. 3D-models of infrastructure and possible origin of emergences

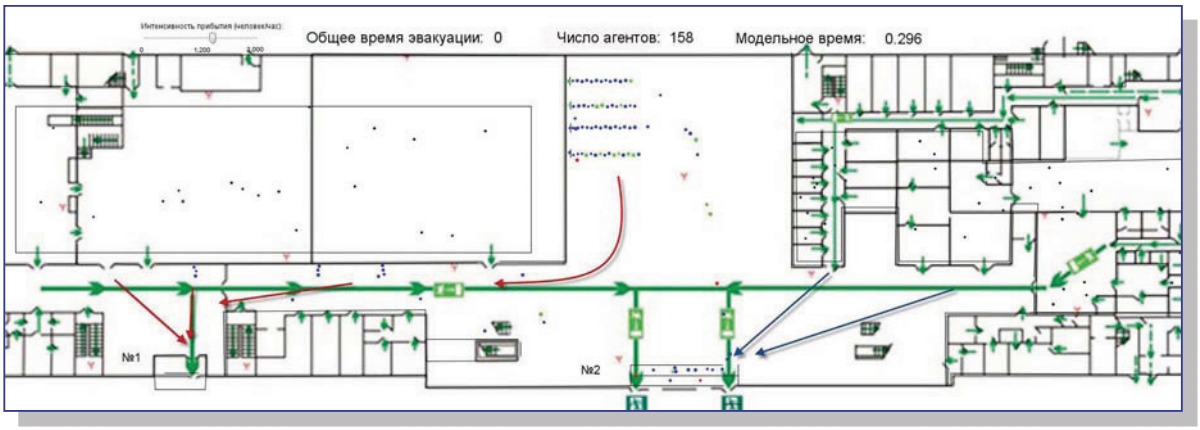


Fig. 3. Multi-agent simulation of suboptimal evacuation routes

The databases DB_1 and DB_3 may be implemented using cloud services C_1 and C_2. When selecting architectures for DB_1, DB_2 and DB_3 should take into account the features of these models and algorithms. Table 1 shows the results of data models analysis for discussed ICS as well as recommended software for database DB_1, DB_2 and DB_3.

TABLE I. TYPES OF APPLIED DATA MODELS FOR DISTRIBUTED ICS

Data Sets Complexity	Structure	Data Bases of ICS	Recommended software
Non-structured	Raw-oriented	DB_1, DB_3	Dynamo, PNUTS, Voldemort, SimpleDB
Semi-structured	Raw-oriented	DB_1, DB_2	CouchDB, MongoDB
	Hybrid-Oriented	DB_1, DB_2	HyperTable, BigTable, HBase, Cassandra
Structured	Raw-oriented	DB_2	Oracle RAC, MySQL Cluster, H-Store, VoltDB
	Column-Oriented	DB_2	Vertica, C-Store, MonetDB/X100, SyBase IQ
	Hybrid-Oriented	DB_2	Oracle ExaData, IBM Netezza, EMC soft

In considered case it is proposed to apply the relational type database for DB_2, as information in this case are from primary sources clearly structured. For databases DB_1 and DB_3 the non-relational database should be used, since it is necessary to handle semi-structured data. At the same time, to reduce the data processing time and the decision-making process big data technologies must be used, for example, Hadoop, MapReduce, and cloud computing services.

III. KEY ISSUES OF PRACTICAL IMPLEMENTATION OF DISTRIBUTED ICS

So, we need to solve the task of complex integration for heterogeneous hardware and software systems to implement the proposed architecture of ICS. For example, the subsystem at execution level is able to collect information about the sate of distributed object that includes a set of various sensors. These sensors inform about a state of the physical characteristics of external and internal environment, location of people on the basis of RFID-tags. The implementation of these tasks involves the execution of a number of the following steps:

- Assignment for each dynamic object a unique identifier (or code number).
- The application of specialized labels containing the object identifier.
- Reading the data from the digital tag.
- Transfer label data into electronic form.

It is necessary to retrieve data from systems that based on global navigation GLONAS/GPS. Large streams of the video data sets are obtained from video surveillance systems. Digitized information from these devices is supplied to a fire simulation subsystem. Then on the basis of the initial data, building plan and fire dynamics equations it can simulate the spread of fire and smoke (Fig .2). Subsystem for fire simulation takes into account many parameters, such as the evaporation of liquid fuels, solid fuel pyrolysis, etc. that requires massive computations. If there is sufficient data sources it can be directly simulated: spreading of smoke, warming of materials, flame propagation from object to object, etc. In the following simulation process critical data are obtained:

- The onset of fire hazards.
- Temperature fire mode.
- The response time of the detectors (smoke and heat sensors).
- The time of opening of sprinklers.
- Water flow rate and effectiveness of fire suppression, etc.

It is necessary to apply high-performance computing systems to solve proposed tasks in real time. It is possible to use the cloud service, and, if necessary, the cloud orchestra for the implementation of the simulation in real time.

For the processing of large data sets generated by the system, it is proposed to use the model of distributed computing MapReduce.

As for hardware for distributed processing it is proposed to apply single-board computers. Currently, there is a large range of single board computers from different manufacturers that

support the operating system Linux. While implementing proposed ICS the prototype of cluster computing system that can be applied at execution level is presented in Fig. 4.

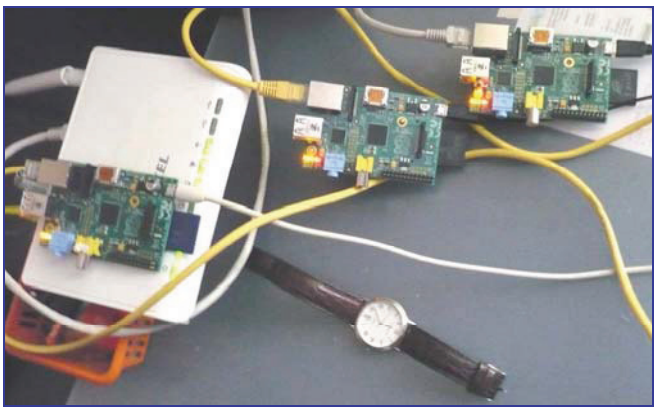


Fig. 4. Prototype of cluster computing system that can be applied at execution level of ICS

IV. Conclusion

The peculiarity of information support and control of critical infrastructures in emergency is complexity of observability and controllability of processes in it. Another problem is how to apply large data sets circulating in the system in real time to provide efficiency of proposed system. As a solution, the architecture of multilevel distributed information and control system is presented. Also application of hierarchical structure of information processing, possible data types, and application different simulation models in proposed system are discussed.

References

[1] R. J. Hansman, "The Impact of Information Technologies on Air Transportation," in Proc. of the 43rd AIAA Aerospace Sciences Meeting and Exhibit, Paper No. 01/2005; DOI: 10.2514/6.2005-1.

[2] The Industrial Internet // www.ge.com/digital/industrial-internet.

[3] S. S. Valeev, "Big Data technologies in aviation," in Proc. of the 2nd International Conference "Information Technologies for Intelligent Decision Making Support", 15-17 May, 2014, Ufa: USATU, 2014, pp. 150-152.

[4] S. S. Valeev, A. D. Kulakov, "Information technologies Big Data in Aviation," in Proc. of the 1st Russian Workshop "Big Data in aviation", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 27 April 2015. Ufa: USATU, 2015, pp. 21-23.

[5] S. S. Valeev, A. D. Kulakov, N. V. Kondratyeva, "The problems of Big Data processing in aviation," in Proc. of the 2nd Russian Conference "Free Flight-2015", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 22-24 December 2015. Ufa: USATU, 2016, pp. 9-15.

[6] S. S. Valeev, M. I. Taimurzin, N. V. Kondratyeva, "An adaptive data acquisition system in technical safety systems," in Automation and Remote Control. December 2013. Volume 74, Issue 12, pp 2137-2142.

[7] Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Eds. EMC Education Services. Indianapolis: Wiley & Sons, Inc., 2015, pp. 55-86.

[8] A. Adamov, "The Truth about Cloud Computing as new Paradigm in IT," in Proc. of the 3rd IEEE International Conference on Application of Information and Communication Technologies (AICT2009), Azerbaijan-Baku, 14-16 October, 2009.

[9] G. Smorodin, O. Kolesnichenko, "Big Data as the big game changer: Big Data-driven world needs Big Data driven ideology," in Proceedings of The 9th IEEE International Conference on Application of Information and Communication Technologies (AICT2015), Russia, Rostov-on-Don, 14-16 October, 2015, pp. 40-43.

[10] S. S. Valeev, N. V. Kondratyeva, "Intelligent Decision Support System for Critical Situations in Techno-Social Infrastructures," in Proc. of the 1st International Conference on Infrastructure Management, Assessment and Rehabilitation Techniques ICIMART'14. Sharjah, UAE. 4-6 March 2014. Paper No. 1569820957.

[11] S. S. Valeev, N.V. Kondratyeva, "Technical safety system with self-organizing sensor system and fuzzy decision support system," in Proc. of the IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2015. Istanbul, Turkey. August 2-5, 2015. Paper No. 15247.

Character-based Feature Extraction with LSTM Networks for POS-tagging Task

Zhandos Yessenbayev
National Laboratory Astana
Astana, Kazakhstan
zhyesenbayev@nu.edu.kz

Aibek Makazhanov
National Laboratory Astana
Astana, Kazakhstan
aibek.makazhanov@nu.edu.kz

***Abstract*—In this paper we describe a work in progress on designing the continuous vector space word representations able to map unseen data adequately. We propose a LSTM-based feature extraction layer that reads in a sequence of characters corresponding to a word and outputs a single fixed-length real-valued vector. We then test our model on a POS tagging task on four typologically different languages. The results of the experiments suggest that the model can offer a solution to the out-of-vocabulary words problem, as in a comparable setting its OOV accuracy improves over that of a state of the art tagger.**

***Index Terms*—LSTM networks, character-based features, continuous vector space word representations, POS-tagging, out-of-vocabulary words.**

I. Introduction

In the last decade deep learning with the neural networks has become the state-of-the-art approach in tasks such as image classification, speech recognition, natural language processing, and others [30]. Various neural network architectures are successfully applied to the common tasks in NLP, e.g. language modeling with feed-forward [2] and recurrent neural networks [23], POS tagging with LSTM networks [12], sentiment analysis with recursive neural networks [32], sentence classification with convolutional neural networks [13]. Moreover, Collobert and Weston proposed a unified architecture to solve several of these tasks simultaneously [4].

The key ingredient of these systems is the ability to learn word-to-vector mappings, so called word embeddings, on a large amount of unlabeled data. Representing words as dense vectors to a certain extent alleviates data sparsity problem. It also provides a good encoder of structure and semantics [24]. On the other hand, a careful review of the literature reveals that most of the approaches encode a sole word as an independent language unit not accounting for the difference between lexical forms and surface variations. Hence, if applied directly to morphologically complex languages, this approach will still suffer from the data sparsity and inability to properly model out-of-vocabulary words. Attempts have been made to incorporate morphological information into the systems [3, 17], raising another question - how to reliably obtain such information in the first place?

In this work, we attempted to build a LSTM neural network system that would implicitly learn morphology from the character streams. We use a LSTM network because this type of networks proved to be robust and successful in sequence labeling [7, 12]. Our approach handles seen and unseen tokens alike, it is language independent, requires no annotated data and assumes no preprocessing save for tokenization.

II. Related Work

The focus of this work is the morphology-aware continuous vector space word representations that able to map unseen data adequately. The problem of finding the distributed representations of words in a low-dimensional vector space has been studied since the 1980s [5, 8, 28]. However, due to the recent advancements of deep learning with neural networks this topic has been revived by the researchers. It has been shown that deep architectures together with unsupervised pre-training techniques can effectively learn various levels of representations in data [15]. The crucial component of such systems is the initial mapping of variable-length words to the fixed-length vectors and further propagation of thereof through the layers of the network.

Many approaches have been proposed to handle this task. The classic ones are "1-of-N" representation and a vector of co-occurrence counts for each word [16]. However, such representations are not scalable and will suffer from data sparsity with growth of the vocabulary size. To address this problem, the matrix factorization methods can be applied to reduce the dimensionality of a vector space. Alternatively, one can use projection matrices or lookup tables [2, 4, 23] and represent words as dense rows in such tables. Such word embeddings can be learned in an unsupervised fashion by training an autoencoder or a language model on unlabeled texts. Other popular word representations, such as skip-gram [23] and ivLBL [25] models are derived from predicting the context of a word and, thus, can capture the syntactic and semantic relationships. Possible alternatives are global vectors for representations [27], which capture both the statistical information via count-based methods and meaningful structures via the log-bilinear prediction-based methods.

While these and similar approaches perform very well, their obvious drawback is inability to handle out-of-vocabulary words in a generic fashion, for they treat a word as an atomic unit which cannot be divided. In fact, almost all of such representations are stemmed from "1-of-N" coding of a word, therefore they a priori limit the vocabulary. To overcome this issue, several morphology-aware approaches have been

proposed. The factored NLMs [1] provide a word or its stem with a fixed set of morphological features. The approach does not provide a generic solution for handling OOV words: they are mapped to a dedicated class. It also requires a morphologically annotated corpus which may not be readily available. Similarly, all the methods based on word compositionality [3, 14, 17] require data annotated for word segmentations. And, of course, such resources are either expensive to obtain (manual annotation) or error-prone (unsupervised segmentation).

Santos and Zadrozny proposed character level embedding, which are learned in an unsupervised fashion together with the word embedding [29]. Their approach does not use annotated corpora and is able to handle unseen words. However, it is not clear from the paper how exactly the OOV words are handled if there are no word embeddings for them. The character-based word encodings have been evaluated by other researchers for POS-tagging [36] and machine translation [37] using recurrent neural networks. All the works obtain state-of-of-the-art performance on the relevant tasks.

Motivated by the last works, we used LSTM neural networks to extract word representations from the character streams during a supervised POS-tagging task for an agglutinative language, namely, Kazakh, which is from the Turkic family of the languages. Although Kazakh is a low resource language, which has not been studied well, there are some papers related to morphological analysis [11, 20] and POS-tagging tasks for Kazakh [19]. The authors exploit either FST-based or HMM-based approaches to solve these tasks. The novelty of the proposed work is in its application of the neural networks to the stated problem. Moreover, the approach suggested can be applied to any language as it does not use any language specific information or assumptions for training and testing, except for the tokenization. In particular, we evaluated our method on four typologically different languages: English, Finnish, Kazakh and Russian.

In what follows, we describe the model description, the experimental setup and the results obtained. Finally, conclusions and future work are outlined in the last section.

III. Model Description

A. LSTM Feature Extraction

The long short-term memory (LSTM) architecture was proposed by Hochreiter and Schmidhuber [9] and it consists of four major components a self-connected memory cell and three multiplicative units – the input, output and forget gates [7]. It is this specially designed memory block that helps to overcome the vanishing gradient problem, a shortcoming of the standard RNNs [10]. This property of the LSTM networks enables effective learning of long-term dependencies, which serves as motivation for using them on character streams.

The key component of our model is the LSTM feature extraction layer. It sequentially reads in a stream of the input vectors corresponding to a word (token) and outputs a single d-dimensional dense vector associated with this word (token). Formally, let $x_1, x_2, \ldots, x_n$ be a sequence of the input vectors corresponding to a word, θ be a set of parameters of the LSTM neurons which are used to compute the LSTM forward pass equations:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ ht &= o_t \tanh(c_t); \end{aligned} \quad (1)$$

where x_t and h_t are the input and output (also hidden state) vectors at time t, respectively; i, f, o and c correspond to the input gate, the forget gate, the output gate, and cell vectors and indices of the relevant weight matrices (eg. W_{xi}) and bias vectors (eg. b_i); σ and $tanh$ are element-wise sigmoid and hyperbolic tangent activation functions, respectively.

If $g(x, h, c, \theta)$ is a function that implements the equations (1), then the final output h_n can be computed recursively as:

$$\begin{aligned} h_1 &= g(x_1, h_0, c_0, \theta) \\ h_t &= g(x_t, h_{t-1}, c_{t-1}, \theta);\ t = 2, 3, \ldots, n \end{aligned} \quad (2)$$

During this process the initial values of the hidden states h_0 as well as the memory cells c_0 are set to zero at the beginning of each word.

B. Input Vectors

The input to the network are character streams coded as a sequences of 10-dimensional binary vectors. Each such vector consists of an 8-bit representation of a character analogous to the ASCII code table, and 2 control bits to identify the beginning and the end of a word. The alphabet contains letters, digits, punctuation and other symbols occurring in text.

To mark the beginning of a word the first bit of a vector corresponding to the first character is set to 1. Similarly, the second bit corresponding to the last character is set to 1 to mark the end. Otherwise the control bits are set to 0. The control bits are used to set the initial values of the hidden states to zero and to let the neurons produce the final output. The control bits were incorporated due to the technical (for automatic processing of the vector streams) and conceptual (for applications operating on a word level, e.g. POS tagging) considerations. An example of such a coding is given in the Fig. 1. The control bits are highlighted.

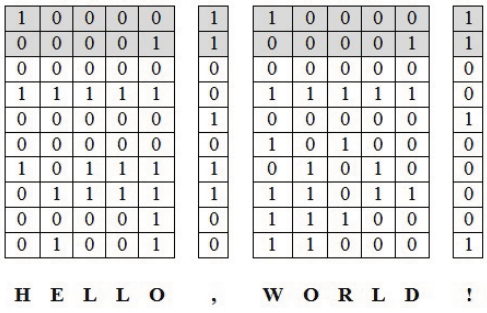


Fig. 1. The coding of the words (tokens)

Encoding words in such way allows generic treatment of the input. Thus, when applied to a specific task, the model will treat seen and unseen tokens alike without resorting to pre-defined classes or any other workarounds.

IV. Application to POS tagging

A. Model Implementation and Training

We test our feature extraction layer in the context of the POS tagging problem. Our model has two layers of LSTM neurons and one logistic layer for the output. The first LSTM layer, proposed as above, maps each token into a d-dimensional continuous vector space modeling the dynamics on a word level, i.e. (arguably) morphology. The second generic LSTM layer takes the tokens of a sentence as the d-dimensional vectors computed previously and further extracts higher order abstractions modeling the dynamics within a sentence, i.e. (arguably) syntax. The final layer of the neural network outputs a POS tag for each token using a softmax function.

To train the network, we optimize a negative log-likelihood objective function whose parameters are updated using a stochastic gradient descent algorithm. The weights of the model are initialized randomly depending on the neurons' activation functions as suggested by Glorot and Bengio [6].

There are only three hyper parameters of the network to setup. These are the sizes of two LSTM layers and a learning rate. In our experiments, we trained the network only for 6 epochs with the fixed set of learning rate values decreasing from 0.01 to 0.0001. The sizes of the first and the second layers are set to 200 and 100, respectively.

The system was implemented using Theano framework [31] in Python and run on one core of Intel Xeon E5620 2.40GHz processor together with NVIDIA Tesla K20 graphics card. The training time depends linearly on the number of input vectors.

B. Data and Experimental Setup

To assess the model's ability to learn various morpho-syntactic features we set up a POS tagging task on four typologically different languages. Namely we use: two agglutinative - Kazakh and Finnish; one fusional - Russian; and one analytic language - English.

In order to make the results more or less comparable across the languages we use the Universal Dependency [26] treebanks which are annotated following the common guidelines and using generic tag sets to ensure consistency across languages. Specifically, we used Russian, release 1.3; Finnish, release 1.2; English, release 1.2 [31]. Because the UD Kazakh treebank [18, 35] is very small (4.5K tokens), we use the Kazakh Language Corpus [21] sample which we semi-automatically re-annotated with the UD POS tag set. The basic statistics on each data set are shown in Tab. 1. The stats may differ from those provided in the original releases, because some test sets were too small, so we had to re-split the data. Also, for English we performed extensive cleaning of very lengthy (40+ chars) tokens, e.g. long URLs and symbol sequences commonly found in the Web-mined data.

TABLE I. Quantitative description of the data sets

Language	#tok., train	#tok., test	%OOV
English	193 344	22 737	7.67
Finnish	162 721	18 301	24.50
Kazakh	168 161	18 708	12.43
Russian	79 772	9 573	26.04

TABLE II. POS tagging accuracy

Language	SF	FB	L	OSF	OFB	OL
English	89.5	86.3	85.7	61.4	36.4	63.1
Finnish	88.2	85.3	86.6	64.8	52.3	73.9
Kazakh	91.8	88.9	89.0	67.2	44.7	75.2
Russian	87.2	82.2	88.1	58.5	39.6	77.3

For the comparison purposes we use the latest available version of the Stanford maximum entropy tagger (SFT) [34] and a frequency baseline that assigns a seen token its most frequent tag and an OOV token the overall most frequent tag.

For a relatively fair comparison, we deliberately set the parameters of the SFT in a way that emulates our model's behavior. As our model does not explicitly use context, tag sequences, and is not bidirectional, we limit the SFT to the extraction of surface features only from the five preceding tokens, and to the use of tag unigrams only.

C. Results

Table 2 contains the results of the experiment. As it can be seen our model (OL) consistently outperforms both the SFT (OSF) and the baseline (OFB) in terms of the OOV accuracy. For the overall accuracy, however, our model (L) outperforms the SFT only on Russian and the baseline (FB) - on Finnish and Russian. The English and Russian cases are particularly interesting, as for the former our model yields the lowest accuracy, and for the latter – the highest (relative to other methods). The English case is obvious: due to analytic nature of the language it is difficult to achieve good results without explicit use of context, which our model (for now) does not do. For Russian resource scarcity might be the culprit: there might be not enough data for the SFT to learn adequately[3] and the OOV rate is rather high for the baseline to perform well (see Tab. 1).

Lastly, we would like to emphasize the fact that the OOV accuracy of our model is much higher for morphologically rich languages. We speculate that our proposed representation implicitly captures morphology inherent in surface forms, and will investigate this behavior in the future.

V. Conclusion and Future Work

In this work we explored a possibility of using LSTM networks for learning the distributed word representations on a character level. By design our approach provides generic treatment for out-of-vocabulary words and is completely language independent, as it does not use any hand-crafted

features nor does it make any language-specific assumptions. We tested our approach on a POS tagging task, and, although the results were rather modest, our intention in this preliminary study was to demonstrate the applicability of the proposed representation not its excellence. Our plan is to further develop this idea and design an architecture for a character-based word generation. Some preliminary results have already been obtained which may appear in our future work. Also it is important to build a system with the state-of-the-art performance on NLP tasks using this framework as the presented model was oversimplified and did not take context into account..

ACKNOWLEDGMENT

This work was supported by the Nazarbayev University under the grant number 064-2016/013-2016 and the Ministry of Education of Science of the Republic of Kazakhstan under the research program number 349//049-2016.

REFERENCES

[1] Andrei Alexandrescu and Katrin Kirchhoff. 2006. Factored neural language models. In Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 4-9, 2006, New York, New York, USA.

[2] Yoshua Bengio, R´ejean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. J. Mach. Learn. Res., 3:1137–1155, March.

[3] Jan A. Botha and Phil Blunsom. 2014. Compositional morphology for word representations and language modelling. In ICML, volume 32 of JMLR Proceedings, pages 1899–1907. JMLR.org.

[4] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning, ICML '08, pages 160–167, New York, NY, USA. ACM.

[5] Jeffrey L. Elman. 1990. Finding structure in time. Cognitive Science, 14(2):179–211.

[6] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010, pages 249–256.

[7] Alex Graves. 2012. Supervised Sequence Labelling with Recurrent Neural Networks, volume 385 of Studies in Computational Intelligence. Springer.

[8] Geoffrey E. Hinton, James L. McClelland, and David E. Rumelhart. 1986. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Distributed Representations, pages 77–109. MIT Press, Cambridge, MA, USA.

[9] Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. Neural Computation, 9(8):1735–1780.

[10] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and Jurgen Schmidhuber. 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[11] Kessikbayeva, G., Cicekli, I.: Rule based morphological analyzer of kazakh language. In: Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM, Baltimore, Maryland, Association for Computational Linguistics (June 2014) 46–54

[12] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. CoRR, abs/1508.01991.

[13] Yoon Kim. 2014. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar, October. Association for Computational Linguistics.

[14] Angeliki Lazaridou, Marco Marelli, Roberto Zamparelli, and Marco Baroni. 2013. Compositional-ly derived representations of morphologically complex words in distributional semantics. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers, pages 1517–1526.

[15] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature, 521(7553):436–444, 5.

[16] Kevin Lund and Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical cooccurrence. Behavior Research Methods, Instruments, & Computers, 28(2):203–208.

[17] Thang Luong, Richard Socher, and Christopher D. Manning. 2013. Better word representations with recursive neural networks for morphology. In Proceedings of the Seventeenth Conference on Computational Natural Language Learning, CoNLL 2013, Sofia, Bulgaria, August 8-9, 2013, pages 104–113.

[18] Aibek Makazhanov, Aitolkyn Sultangazina, Olzhas Makhambetov, and Zhandos Yessenbayev. 2015. Syntactic annotation of kazakh: Following the universal dependencies guidelines. a report. In 3rd International Conference on Turkic Languages Processing (Turk-Lang 2015), pages 338–350, Kazan, Tatarstan.

[19] Makazhanov, A., Yessenbayev, Zh., Sabyrgaliyev, I., Sharafudinov, A., and Makhambetov O., On Certain Aspects of Kazakh Part-of-Speech Tagging. In the Proc. of 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT2014), 2014, pages 240-243.

[20] O. Makhambetov, A. Makazhanov, I. Sabyrgaliyev, Z. Yessenbayev, Data-Driven Morphological Analysis and Disambiguation for Kazakh. In the Proceedings of the 2015 Computational Linguistics and Intelligent Text Processing, Part I., Cairo, Egypt: Springer International Publishing, 2015, pages 151-163.

[21] Olzhas Makhambetov, Aibek Makazhanov, Zhandos Yessenbayev, Bakhyt Matkarimov, Islam Sabyrgaliyev, and Anuar Sharafudinov. 2013. Assembling the kazakh language corpus. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1022–1031, Seattle, Washington, USA, October. Association for Computational Linguistics.

[22] Tomas Mikolov and Geoffrey Zweig. 2012. Context dependent recurrent neural network language model. In 2012 IEEE Spoken Language Technology Workshop

(SLT), Miami, FL, USA, December 2-5, 2012, pages 234–239.

[23] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. CoRR, abs/1301.3781.

[24] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. In Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA, pages 746–751.

[25] Andriy Mnih and Koray Kavukcuoglu. 2013. Learning word embeddings efficiently with noise-contrastive estimation. In Advances in Neural Information Pro cessing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States., pages 2265–2273.

[26] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Paris, France, may. European Language Resources Association (ELRA).

[27] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar, October. Association for Computational Linguistics.

[28] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1986. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Learning Internal Representations by Error Propagation, pages 318–362. MIT Press, Cambridge, MA, USA.

[29] Cicero D. Santos and Bianca Zadrozny. 2014. Learning character-level representations for part-of-speech tagging. In Tony Jebara and Eric P. Xing, editors, Proceedings of the 31st International Conference on Machine Learning (ICML-14), pages 1818–1826. JMLR Workshop and Conference Proceedings.

[30] Jurgen Schmidhuber. 2015. Deep learning in neural networks: An overview. Neural Networks, 61:85–117. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].

[31] Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014).

[32] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In In EMNLP.

[33] Theano Development Team. 2016. Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688, May.

[34] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pages 173–180. Association for Computational Linguistics.

[35] Francis M. Tyers and Jonathan N.Washington. 2015. Towards a free/open-source universal-dependency treebank for kazakh. In 3rd International Conference on Turkic Languages Processing, (TurkLang 2015), pages 276–289.

[36] Wang Ling, Chris Dyer, Alan W. Black, Isabel Trancoso, Ramon Fermandez, Silvio Amir, Luís Marujo, Tiago Luís, Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation. In EMNLP 2015.

[37] Junyoung Chung, Kyunghyun Cho and Joshua Bengio. A Character-level Decoder without Explicit Segmentation for Neural Machine Translation. In ACL 2016.

Smart Home for Elderly Living Using Wireless Sensor Networks and an Android Application

Kashif Nisar[1, 2], Ag. Asri Ag. Ibrahim[1], Lifen Wu[3], Abzetdin Adamov[4], and M. Jamal Deen[2], *Fellow IEEE*

[1]Faculty of Computing and Informatics, Universiti Malaysia Sabah
Kota Kinabalu, Sabah, Malaysia (kashif@ums.edu.my, awgasri@ums.edu.my)

[2]Department of Electrical and Computer Engineering, McMaster University
Hamilton, ON L8S 4K1, CANADA (nisark1@mcmaster.ca, jamal@mcmaster.ca)

[3]Information Technology and Engineering Department
Jinzhong University, Jinzhong, Shanxi 030619, China (wulifenwlf@sina.com)

[4]Applied Research Center for Data Analytics and Web Insights (CeDAWI),
Qafqaz University, Baku, Azerbaijan (aadamov@qu.edu.az)

***Abstract*—A Smart Home (SH) is a house or an apartment equipped with advanced automation technologies to provide the occupants with intelligent monitoring and actionable information that can be situation specific. It allows for improvements in the way we live or work, and for improved energy efficiencies. The smart home is especially relevant for the elderly because the intelligent sensing systems would allow for remote monitoring and possibly control of health and environmental parameters according to their health status and living needs. The burgeoning elderly population and rising costs of elderly healthcare and home needs have led to a rapid evolution of the elderly smart living environment. However, the lack of proper eldercare smart home technologies has inspired us to develop Information and Communication Technologies (ICT) based smart home for the elderly using the Android platform. We have proposed a model for the elderly using Wireless Sensor Networks (WSNs) implemented as an Android application. Our smart home research potentially allows the elderly to continue to live in their own homes while being monitored non-invasively, seamlessly and economically according to their healthcare needs and status.**

***Index Terms*—SH, ICT, WSNs,**

I. Introduction

The Smart Home (SH) is a living environment incorporating sophisticated automation technologies for customized intelligent monitoring, feedback and action according to the needs of the home occupant(s). Wireless techniques in the SH are essential for non-invasive and non-intrusive implementation of the advanced automation systems. The advantages and limitations of available architectures and their implementations for heterogeneous sensing were examined in [1]. Heterogeneous sensors and actuator nodes with communications capabilities can be installed in the SH to create Wireless Sensor Networks (WSNs) that could allow the nodes to generate, acquire and integrate real-time data securely [2]. WSNs are becoming more important in recent years because of their ability to achieve real-time situational information from surveillance and monitoring using short-range low-power communication services [3] and [4].

The ease of remote access by state-of-the-art Information and Communication Technologies (ICT) based systems will improve our living and working environments by providing us with more convenience and customized environments to meet our needs. In fact, the SH concept has led to the idea of ageing in place in which elderly persons can continue to live in their own homes while still in control of their healthcare needs and living environment [5].

Recent data [6] highlights the alarming trend that the fraction of worldwide population 60 years and older is rising at a startling rate. By 2050, this specific group will have globally increased to over 50% of the world's population [7]. With such a dramatic increase and the resulting huge elderly population, the urgency of eldercare is being emphasized in many government policies and programs, and the elderly smart home smart is a good viable solution to this dilemma. One successful and cost-effective implementation of elderly healthcare is remotely monitoring so they can stay at home rather than in nursing homes, clinics, and hospitals that are very expensive and with limited spaces [8].

A. Problem Statement

The rapidly rising elderly population and the lack of economical and suitable elderly smart living home technologies have inspired us to develop information and communication technologies (ICT) based elderly smart living home in the Android platform. Many elderly persons encounter with various difficulties in their everyday life. These difficulties are especially acute when they are isolated at home without anyone to provide needed assistance. Therefore, a smart home system tailored to the needs of the elderly can make their life more comfortable, happy and help sustain their

good health and well-being. Based on these identified needs for the elderly smart home, the objectives of this paper are the following.

a. To design a smart home automation system that could assist the elderly in specific daily activities.
b. To implement a system that could minimize wastage of electricity, preserve human energy and increase the elderly's quality of life.
c. To periodically or continuously monitor the status of smart home devices and to remotely access home appliances.

II. Literature Review

A summary of the relevant literature review is shown in Table I.

TABLE I. Summary of Literature Review

Reference	Brief Description
S.M. Hemant Ghayvat [9]	Aimed to design and develop a sensor network which is reliable, efficiency, flexible, economical and realistic for smart home system that is WSN- and IoT- based.
Sumi Helal, [10]	Aimed to develop pervasive computing in smart home system by interconnecting sensors, actuators, computers and other smart devices.
D. Han and J. Lim, [11]	Designed a smart home energy management system using IEEE 802.15.4 and ZigBee.
H. Isilak [12]	Aimed to develop smart home application for disabled individuals through application of a wireless sensor network.

III. Methodology

Our methodology is divided into six phases.

Phase 1 - Problem identification and background: In this phase, the objectives of the research and the problem statement are formulated and defined. Both functional and non-functional requirements are identified and described.

Phase 2 - Literature review and survey: In this phase, a detailed literature review is carrier out. Relevant papers are analyzed, and the methods and concepts used in different types of smart home technologies are catalogued. In addition, methods used to connect sensors with computing systems are analyzed. Then, the most suitable server to be used in this system is proposed. The method of integration of Android application to the smart home technologies is analyzed.

Phase 3 - Design proposed model: In this phase, the Graphical User Interface (GUI) and database is designed based on the requirements identified in Phases 1 and 2.

Phase 4 - Develop prototype and simulation: In this phase, the prototype is developed and simulations are performed.

Phase 5 - Testing and Evaluation: Here, the developed system is tested to ensure the system objectives and requirements are met. The developed ***app*** undergoes alpha and beta test before deployment. If bugs or errors occur, refinements and improvements are conducted until the objectives are satisfied.

Phase 6 - Deployment: In this phase, the tested system is officially released for public use.

In the next three sub-sections *A* to *C*, the wireframe of the proposed model, the algorithm and the conceptual model are shown and described.

A. Wireframe

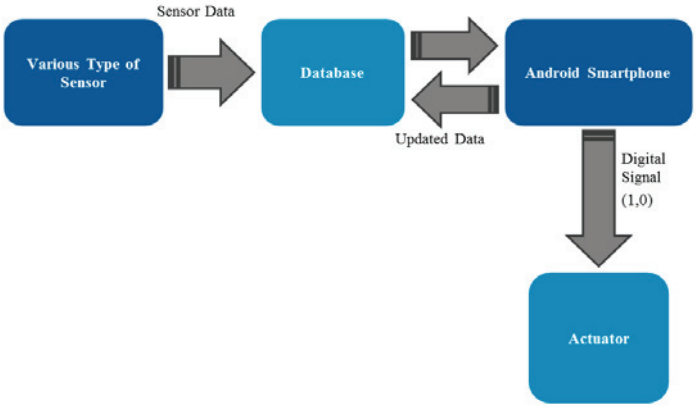


Fig.1. Wireframe

B. Algorithm

Here, the Observe, Learn, and Adapt (OLA) algorithm is applied in our smart home system. The OLA algorithm uses the results of the combination of wireless sensors and artificial intelligence concepts towards improving energy management in the smart home.

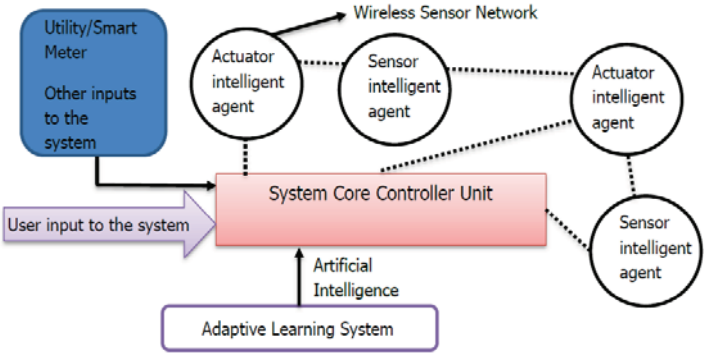


Fig. 2. Research Algorithm

Based on the OLA algorithm, a block diagram of the hybrid intelligent technique for the smart thermostat is schematically illustrated in Fig. 1 and Fig 2. The system includes a two way communication device connected to the smart meters. With this capability, the power consumption can be monitored remotely, and if needed, in devices like sensors, home appliances and switches.

C. Conceptual Model

A conceptual model of a home in which automated systems will be installed with labels, is shown in Fig.3.

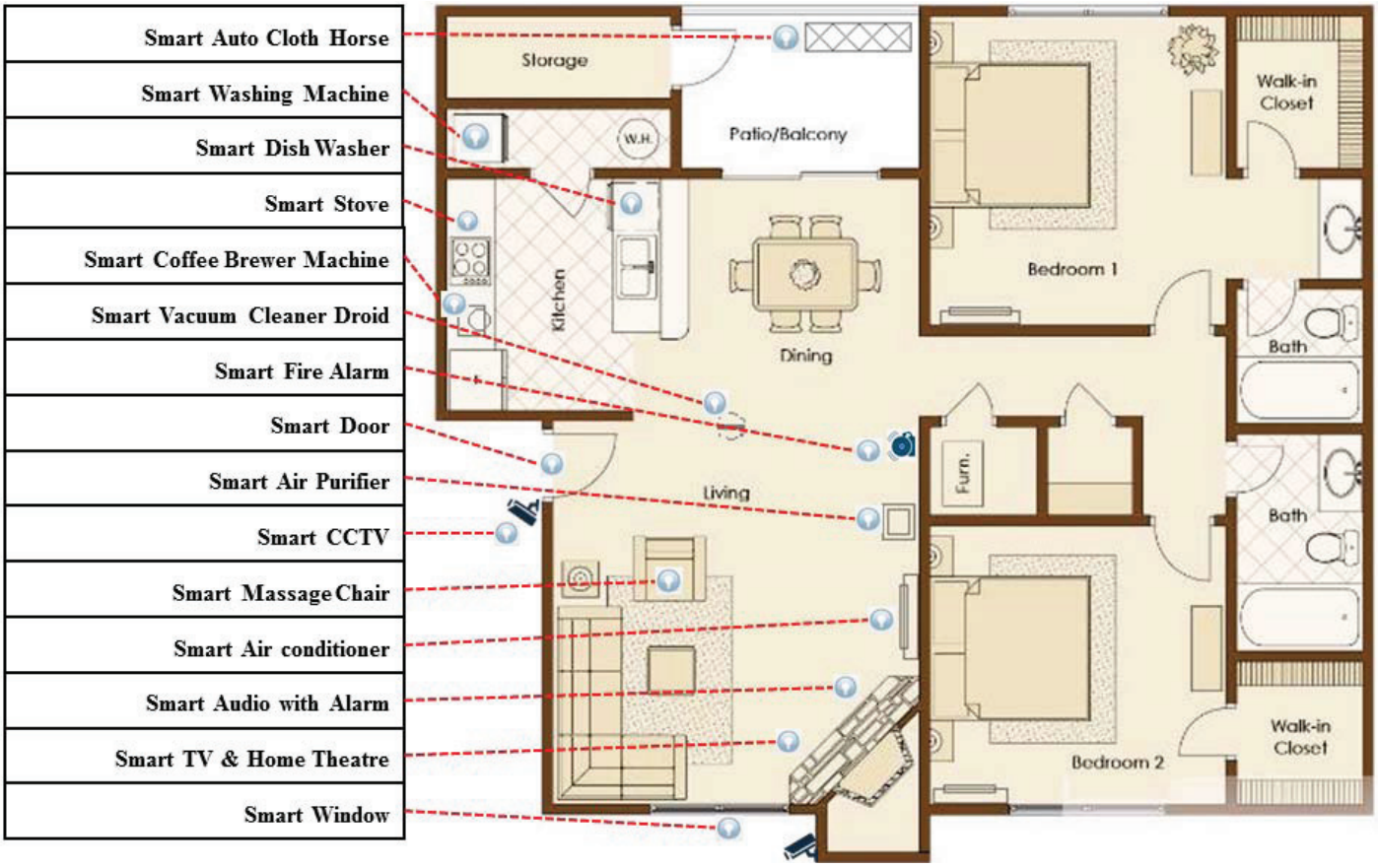


Fig. 3. Conceptual Model of Smart Home.

IV. Experimental Setup

In recent years, the demand for smartphone apps in the Android platform has grown rapidly. Therefore, for our research, we chose the Android platform for the development of our ICT-based smart home system. In this section, the reasons for using Android are explained.

First, Android is chosen because it is an open source platform. Android has a low barrier to entry and there are no costly licensing fees or development tools. Further, the Android Software Development Kit (SDK) is open-source, so it is easy to interact with the software development community and to be aware of new developments of Android mobile applications. Also, the Android SDK includes a comprehensive set of development tools like debugger, libraries, handset emulator, sample code, documentation and tutorials. This makes it easier to develop our own applications in the Android platform.

Second, Android is an integrated software platform. The Android platform is suitable for customization and it allows developers to design applications for users based on their preferences. Therefore, we can integrate and optimize the mobile app according to our needs. Android is an excellent mobile platform between the application and processes architecture.

Third, Android uses the consistent and open programming language Java and the emulator of the Android platform has a modern design and is easy to use. The Android applications are easily scripted in Java language with the help of a rich set of libraries. Therefore, anyone can build Android applications once they have a working knowledge of Java.

Fourth, compared to the other platforms, Android is an excellent environment for the development of information and communication technologies smart home system which is specially designed for the elderly. Details are shown in data flow diagram in Fig. 4.

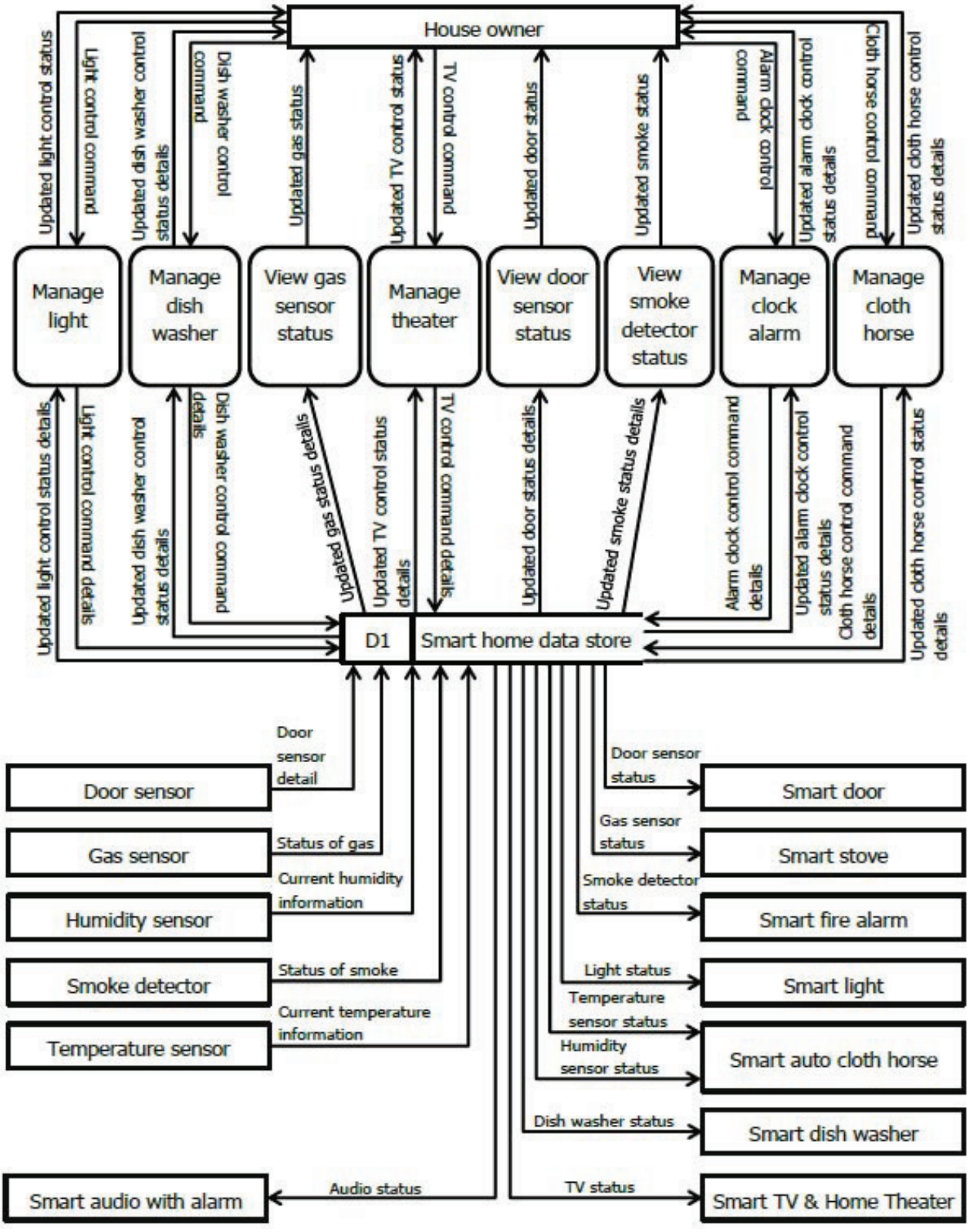


Fig. 4. Data flow diagram

A. Wireframe of Android Application Sensor

Generally, we have chosen to focus in our research scope on few different areas in the field of smart home systems. Smart door could possibly detect the presence of intruders and break-ins as soon as the door is opened without authorized permission. The siren will immediately turn on when the presence of intruders is confirmed. Smart fire alarm uses smoke detector to detect the presence of fire lighting and send immediate alert to the user when fire happens. Smart stove installed gas sensor used to detect the presence of a dangerous flammable gas leaking that caused by the forgotten action of extinguishing the fire. It gives a simple off signal when gas leaking is detected and alerts the user through mobile apps, is displayed in Fig. 5.

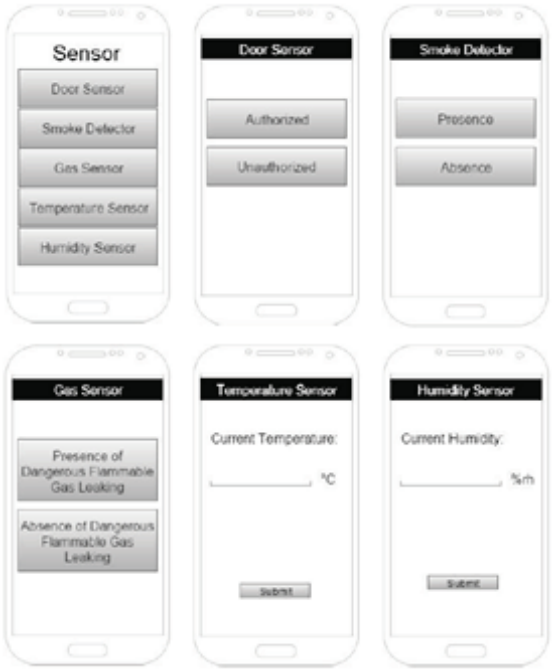


Fig. 5. Wireframe of Android Application

B. Wireframe of Smart Living Home's Android Application

The smart air purifier detects visible color change and air quality level by sensor technology and then it will remove dust, dander, mildew, fungi and allergens. Through the temperature sensor and the humidity sensor, the auto cloth horse can automatically protect the cloth from getting wet and optimize the drying process. Morning call the resident with the pleasant music. Remind resident to take medicine on schedule. Allows power saver mode that could be turn off within time set. TV set could be remotely accessed via mobile apps. Allows power saver mode that could be turn off within time set, is shown in Fig. 6.

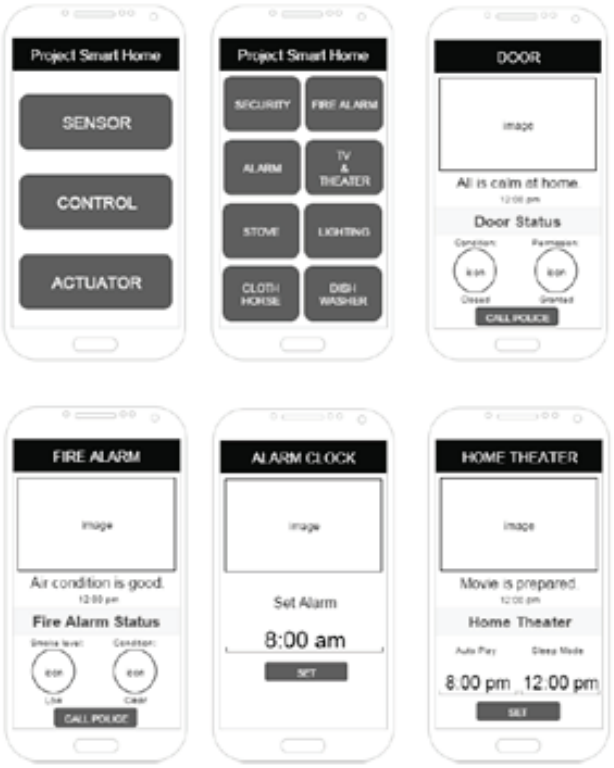


Fig. 6. Wireframe of Android Application

C. Wireframe of Actuator

Smart lighting system can practice flexible and freedom in creating different lighting scene around the home. The dimness could be adjusted through the apps according to various scenarios. User can receive notification after the wash process of dish is done. User can check the washing status in or outside home and also use all the functions that washing machine offer through smart phone. Besides, the washing machine can let the user to set the water temperature they want by using smart phone, explained in Fig. 7.

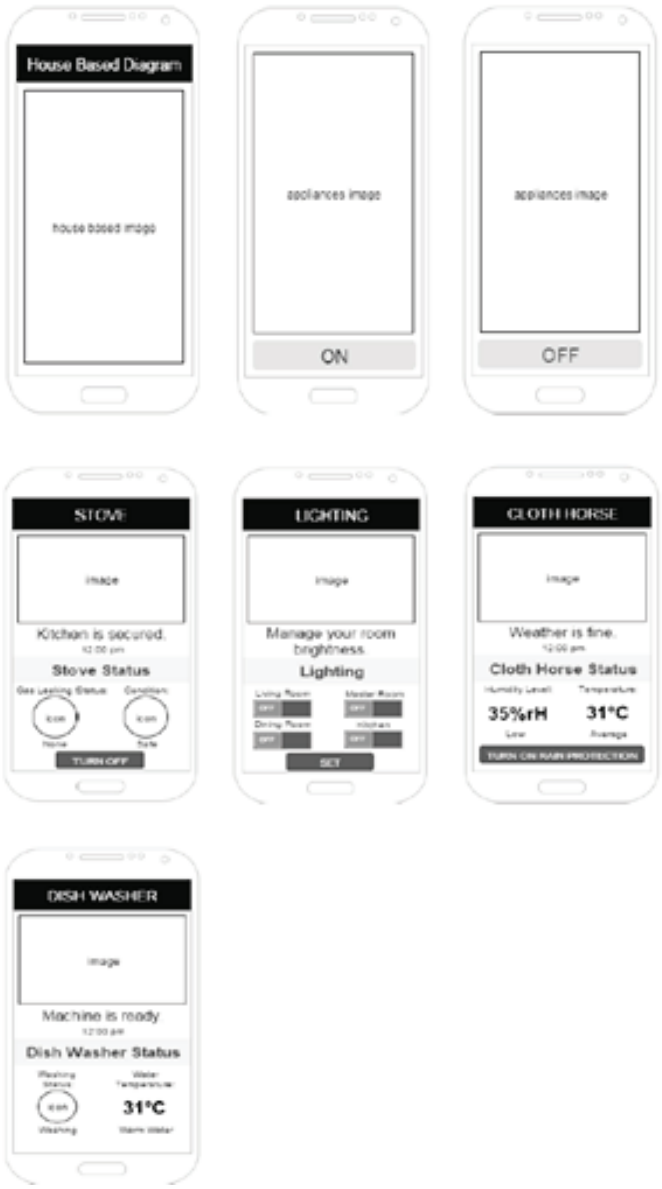


Fig. 7. Wireframe of actuator.

V. Results and Discussion

The prototype of our Android-based Smart Home (SH) application is implemented using Android Studio and Geny motion as the simulator. The application is divided into three main modules such as Sensor module, Control module and Actuator module. The sensor module functions as the data transmitter medium, and the sensor is triggered by the module's functionality. At present, the sensors deployed are Door Sensor, Smoke Detector, Gas Sensor, Temperature Sensor and Humidity Sensor. Sub-modules included in the Control module are Door module, Fire Alarm module, Alarm Clock module, Home Theatre module, Stove module, Light module, Cloth Horse module and Dishwasher module. These modules function as the main control of Smart Home, and each different sub-module allow for remote monitoring or controlling the particular appliances. The Actuator module shows the response of the appliances in the actual system as an illustration. The following sections show screenshots of the Smart Home Android Application's user interface.

Fig. 8. The user interface of Main module.

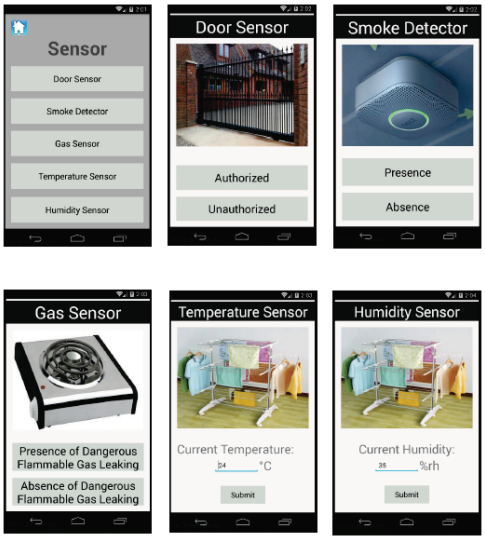


Fig. 9. The user interfaces of the sensor module and its sub-module.

Smart air conditioner integrates the air-conditioning and ventilation to a coherent and efficient climate control. The remote access allows optimum temperate and air quality is maintained all time long, is presented in Fig. 8 and Fig. 9.

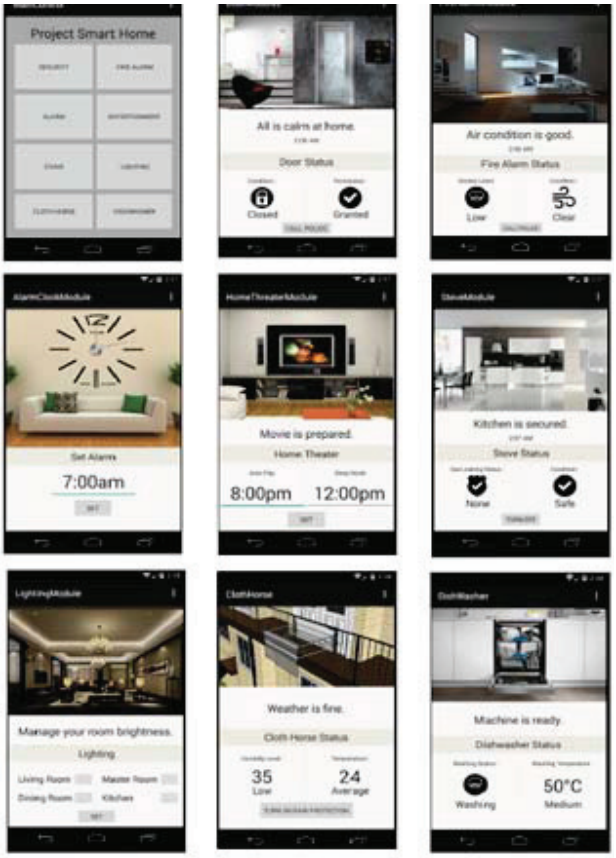


Fig. 10. The user interface of the control module and its sub-modules.

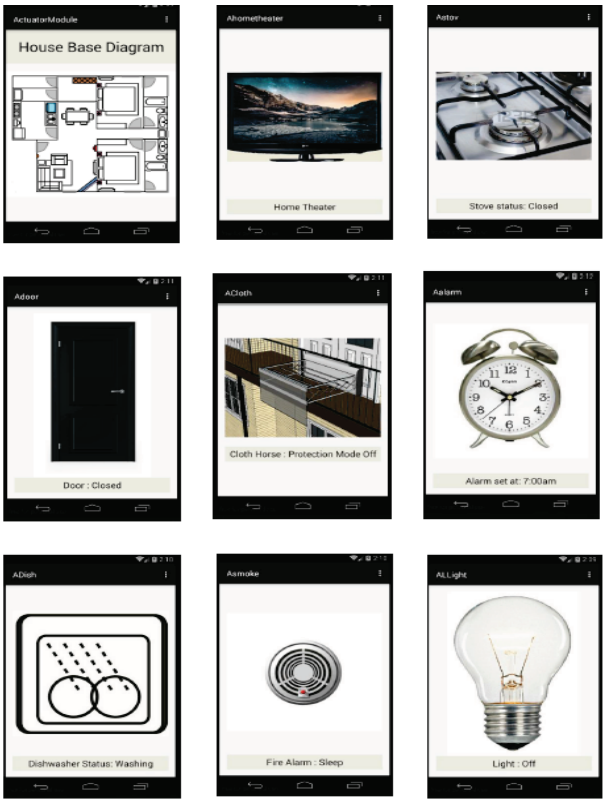


Fig. 11. The user interface of the actuator module and its sub-modules.

Smart coffee brewer provides remote access for setting up a weekly schedule and adjusting brew time if needed. New recipe could be founded and installed into the machine to brew new coffee flavor. The smart vacuum cleaner droid will constantly clean the floor according to the schedule that set by user, shown in Fig. 10 and Fig. 11.

VI. Conclusion And Future Work

We have successfully developed a system which can prevent wastage of electricity, preserve human energy and simultaneously improve the quality of life of elderly persons. Our system can help the home owner to continuously or periodically monitor the smart home devices and to conveniently take control on devices status and remotely access the home appliances. The ease of remote access with our system will eventually make our lives easier, especially for the elderly, as it could provide conveniences and safety guidance. Therefore, this smart home research potentially allows the elderly to continue to live in their own homes while being monitored non-invasively, seamlessly and economically according to their healthcare needs and status.

At present, our research only considers eight appliances in our information and communication technologies based smart living home system. However, the experience gained is easy to build on towards a functional smart home for elderly living and healthcare needs. Our immediate future plans include research in two areas. First, we plan to increase the number of the appliances included in the information and communication technologies smart living home system. Second, we will deploy the system to the many real-world situations so that it can be redesigned and redeveloped to increase its adaptability and use.

ACKNOWLEDGMENT

Authors are most grateful to Professor Dr. Yong-Jin Park from Department of Computer Science and Engineering, WASEDA University, Tokyo, JAPAN, who has contributed to various aspects of this research work described here. These works were performed in Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, CANADA and Faculty of Computing and Informatics, Universiti Malaysia Sabah (UMS), Kota Kinabalu Sabah, MALAYSIA. Furthermore, this work is supported by SGI research grant and its value in Malaysia Ringgit RM 20,000 from Centre for Research and Innovation UMS, Malaysia.

References

[1] Federico Viani. et al.” Wireless Architectures for Heterogeneous Sensing in Smart Home Applications: Concepts and Real Implementation”, Proceedings of the IEEE, Ternto, Italy, Vol 101, Issue: 11, pp. 2381-2396, 16, October 2013.

[2] Kamrul Islam, Weiming Shen and Xianbin Wang, “Security and privacy considerations for Wireless Sensor Networks in smart home environments,” IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD), London, Canada, pp. 626-633, 25, September 2012.

[3] Jinsung Byun et al., “An intelligent self-adjusting sensor for smart home services based on ZigBee communications,” IEEE

Transactions on Consumer Electronics Vol 58, Issue: 3, 794-802, 27, September 2012.

[4] P. Porras, D. S. Ghataoura, J. E. Mitchell and G. E. Matich, "Networking and application interface technology for wireless sensor network surveillance and monitoring", IEEE Communications Magazine, Vol. 49, No.10, pp. 90-97, 06, October 2011.

[5] J. Misic and V. B. Misic, "Bridge performance in a multitier wireless network for healthcare monitoring", IEEE Wireless Communications, Vol. 17, No. 1, pp. 90-95, 2010.

[6] H Wang, HS Choi, N Agoulmine, MJ Deen, and, J.W-K Hong, Information-based sensor tasking wireless body sensor networks in healthcare systems, proceedings of the 2010 IEEE International Conference On Network And Service Management (CNSM), Niagara Falls, Ontario, Canada, 25–29, October 2010.

[7] M. Jamal Deen, "Information and communications technologies for elderly ubiquitous healthcare in a smart home", Invited Paper, Pers Ubiquit Comput, Springer-Verlag, pp. 573–599, 26, June 2015.

[8] A.S. Saisakul Chernbumroong, "Perception of Smart Home Technologies to Assist," pp. 3-5, 2015.

[9] S.M. Hemant Ghayvat, "WSN- and IOT-Based Smart Homes and Their Extension to. Sensors", pp.10351-10375, 4, May 2015.

[10] Sumi Helal, "The Gator Tech Smart House: A Programmable Pervasive Space," IEEE Computer Society, Florida, Mar 2005.

[11] D. Han and J. Lim, "Smart Home Energy Management System using IEEE and ZigBee", IEEE Transactions on Consumer Electronics, Vol. 56, pp.1403-1409, 2010.

[12] H. Isilak, "Smart Home Applications for Disabled People by Using Wireless Sensor Network", Engineering project report, Yeditepe University, Faculty of Engineering and Architecture, Department of Computer Engineering, pp.1-46, October 2010.

Fatigue Test Optimization for Complex Technical System on The Basis of Lifecycle Modeling and Big Data Concept

Natalya Kondratyeva, Sagit Valeev
Dept. of Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
knv24@mail.ru, vss2000@mail.ru

***Abstract* — Increasing efficiency of fatigue testing complex technical systems (for example, aircraft engine) is possible with its technical and economic assessment on base of relationship with the economic effect from the system operation. The amount of data is received during the engine life cycle. It should be properly processed to build the lifecycle model. Big Data concept can be suggested as an effective tool for storing and analyzing the life cycle information.**

***Index Terms* — complex technical system, fatigue test, efficiency, aircraft engine, lifecycle, modeling, Big Data.**

I. Introduction

Simulating the life cycle of a complex technical system is a modern trend. For instance, it is urgent for assessing the efficiency criterion of fatigue test in the framework of the system life cycle. Modern gas turbine engines (GTE) are bright example of complex technical systems consisting of several thousands of parts and components. Engine condition during a test and set operation is characterized by various parameters, measured at intervals of several seconds to several minutes [1-4].

In addition, the data on the financial performance indicators at various stages of the engine life cycle (LC) is collected. The amount of data received during the life cycle of GTE (development, production, operation, testing, and repair) can be measured in terabytes and sometimes in petabytes. Thus, the urgent task is to process this information on the basis of Big Data technologies and Cloud Computing in order to build predictive models at the stage of engine tests during development and production [5-8].

One of the main factors limiting the ability to create a simulation model of the life cycle of a complex technical system for selection of fatigue test parameters is the difficulty of assessing the efficiency criterion of fatigue test in the framework of the engine life cycle. The reason for this kind of problem in the first place is the need for preprocessing large volumes of heterogeneous data, collected at different stages of lifecycle of the complex technical system [9-12].

II. Problem Definition

It is expected that the efficiency of fatigue test of gas turbine engines is possible due to using the results of simulation of engine life cycle in the justification of test parameters. As a criterion of the efficiency of fatigue tests, absolute indicator of the efficiency of capital investments, annual operating costs, profits from the engines and operation, minimum unit costs due to wear and tear, cost of the engine operation, engine reliability and others can be used (Fig. 1).

Relation between the economic efficiency of the product and the parameters of fatigue tests is also available through other indicators characterizing the effect of improving the reliability: maintainability, storage ability, durability, and complex indicators, such as availability etc.

In general, the total cost presented in Fig. 1 is the cost associated with the consumption of all kinds of resources, since the start of the research and development of the engine until the completion of the withdrawal of its utilization.

The value of the selected criterion of efficiency is becoming more specified step by step as the engine passes life cycle stages and information on the engine is becoming more detailed [11].

All the arrows in Fig. 1 present numerous information flows. These data are necessary for estimating efficiency criteria and test parameters optimization. Also central to the suggested approach are data warehouses (DW_1, DW_2), collecting information on engine life cycle, including life spending details, life hours left, maintenance and repair, where and when, financial parameters etc.

To choose the parameters of the fatigue test of the aircraft engine in order to provide maximal efficiency of the engine lifecycle, it is necessary to solve the following sequential tasks:

- Justify the choice of efficiency criteria fatigue tests of the product.
- Define the parameters of the engine life cycle that need to be taken into account in the simulation model.

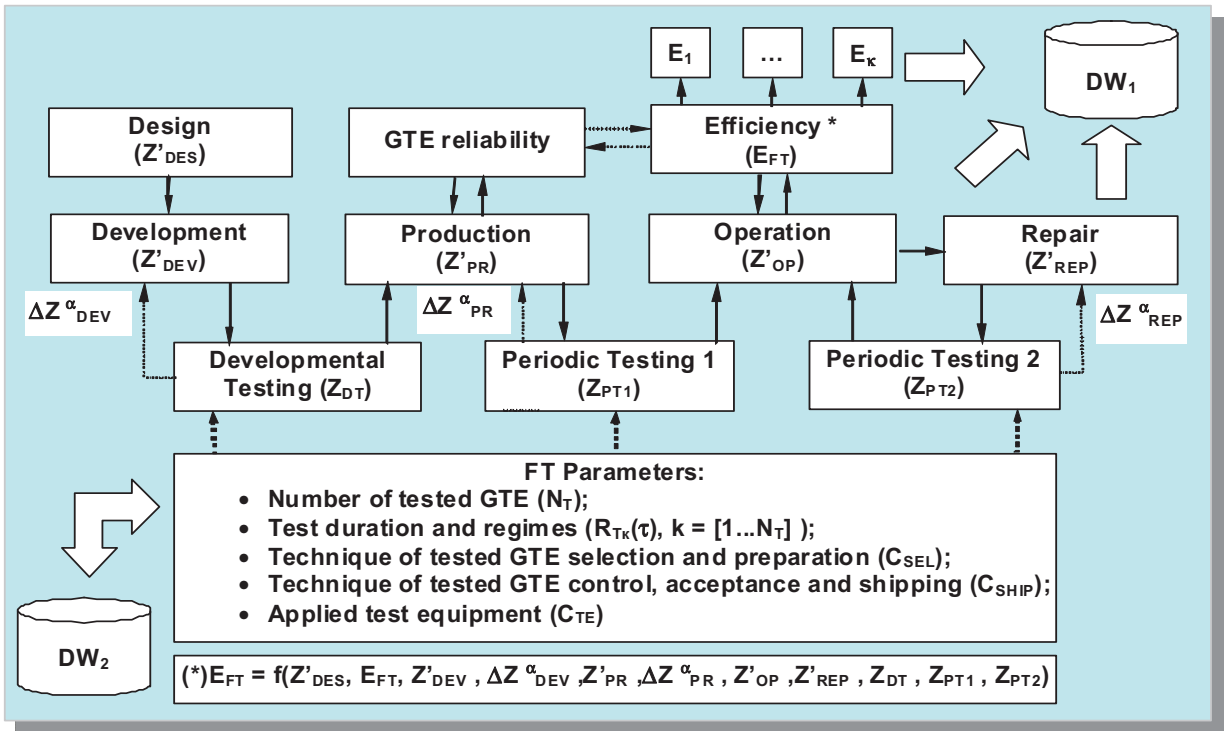


Fig. 1. Relation between economic efficiency of the engine lifecycle stages and the parameters of fatigue test

- Select a modeling tool and develop a simulation model of the life cycle.
- Make a series of computer simulation within planned experiments.
- Determine the optimal values of the fatigue test parameters on the basis of life cycle modeling [3].

III. Fatigue Test Optimization

For complex technical objects such as gas turbine engines, depending on the amount of available background information, various simulation schemes in operation at a fixed resource and operation of on technical condition can be applied.

The objective of maximizing the efficiency E is achieved by optimizing the number of tested engines N_T, test modes $R_T(\tau_T)$ and duration τ_T:

$$\begin{cases} E = \max\ f\,[N_T,\ R_T(\ \tau_T),\ \tau_T;\ C_\Sigma]; \\ R_T(\ \tau_И) \in G_R,\ \ \tau_T \in G_\tau, \end{cases} \qquad (1)$$

where C_Σ – function including lifecycle parameters remaining unchanged in the tests optimization; G_R, G_τ – set of the test regimes and duration.

To solve the task (Eq. 1), it is necessary to make function f more specified. It is possible to provide the following decision in case of the aircraft engine. As it is known, the most complete reproduction in tests operational defectiveness the engine for the minimum time is the primary requirement for accelerated testing of complex technical systems:

$$E \sim \max\ (K_A = \tau_{op} / \tau_T); \qquad (2)$$

$$\Pi_T[P_0, R_T(\ \tau_T),\ \tau_T] = \Pi_{op}[P_0, R_{op}(\ \tau_{op}),\ \tau_{op}]; \qquad (3)$$
$$P_0 = \text{idem}\ ;\quad R_T(\ \tau_T), R_{op}(\ \tau_{op}) \in G_R\ ,$$

where $\Pi_T[P_0,\ R_T(\tau_T),\ \tau_T]$ – engine defectiveness in test depended on initial condition vector P_0, regimes $R_T(\tau_T)$ and test duration τ_T; $\Pi_{op}[P_0,\ R_{op}(\tau_{op}),\ \tau_{op}]$ – engine defectiveness in operation depended on initial condition vector P_0, regimes $R_{op}(\tau_{op})$ and operation duration τ_{op}; K_A – test acceleration factor; G_R – set of loading regimes in which defectiveness of the engine components and parts is remained.

As far as it is hard to comply with condition (3) in real life, test optimization is provided as follows:

$$E \sim \begin{cases} \max K_A\ ; \\ \delta\Pi = \left|\Pi_T[P_0, R_T(\tau_T), \tau_T] - \Pi_{op}[P_0, R_{op}(\tau_{op}), \tau_{op}]\right|; \end{cases} \qquad (4)$$
$$P_0 = \text{idem}\ ;\quad R_T(\tau_T), R_{op}(\tau_{op}) \in G_R\ ,$$

or, combining factors K_A and $\delta\Pi$, the choice of the test parameters can be made under condition:

$$\Phi_\Sigma = \max \sqrt{\alpha_1 \left(\sum_{i=1}^{n} \left(\left(\Pi_{op.i.v_i}[P_0, R_{op}(\tau_{op}), \tau_{op}] \right) / \right.\right.} \qquad (5)$$
$$\left. \left. \Big/ \left(\Pi_{Ti}[P_0, R_T(\tau_T), \tau_T] - \Pi_{op.i.v_i}[P_0, R_{op}(\tau_{op}), \tau_{op}] \right) \right)^2 \right)^2 +$$
$$+ \alpha_2 \left(\frac{\tau_{op}}{\tau_T} \right)^2;\ \overline{\delta\Pi} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\Pi_{Ti} - \Pi_{op.i.v_i}}{\Pi_{op.i.v_i}} \right] \le \overline{\delta\Pi}^*;$$
$$R_T(\tau_T), R_{op}(\tau_{op}) \in G_R;\quad \tau_T, \tau_{op} \in G_\tau;\quad \nu_i \in [1 \ldots N_{op}],$$

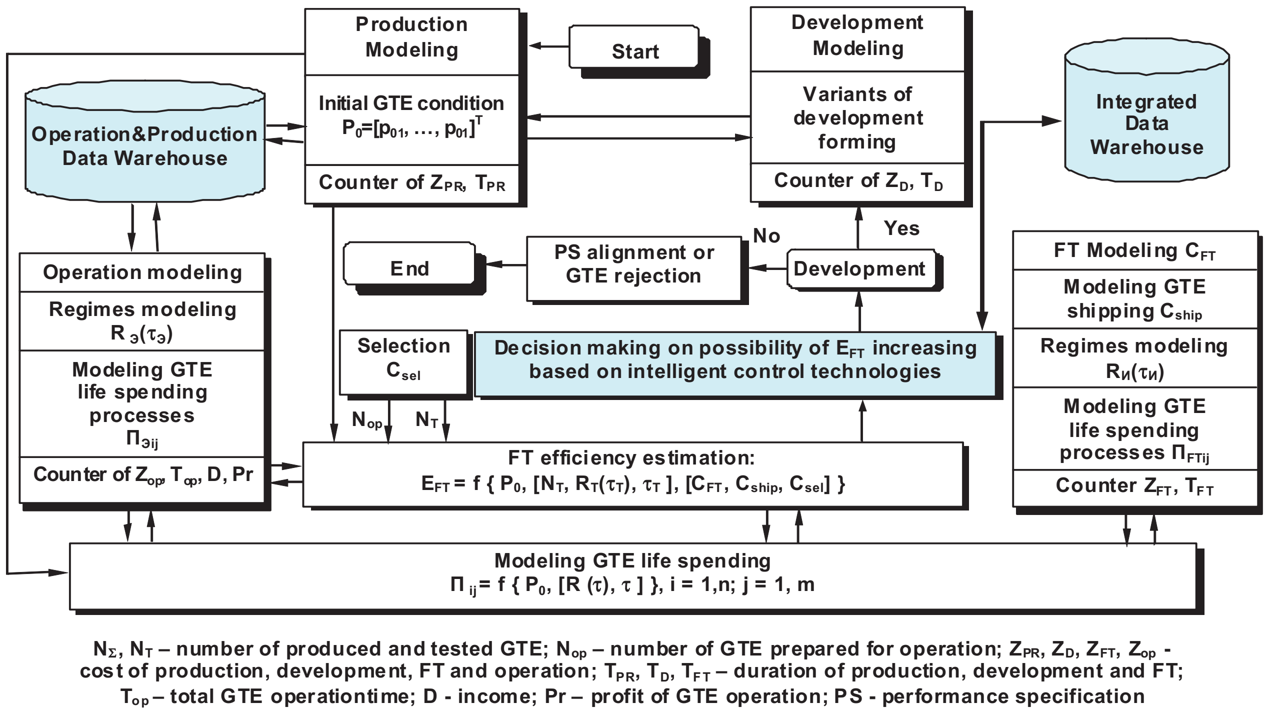


Fig. 2. Generalized scheme of gas-turbine engine (GTE) lifecycle model building for choosing engines test parameters

where Π_{Ti} и $\Pi_{op.i.\nu i}$ – defectiveness of *i*-th part of the engine in test and in operation; τ_T, τ_{op} – loading duration in test and in operation; N_{op} – number of operated engines; α_1, α_2 – weight factors ($\alpha_1 + \alpha_2 = 1$); $\overline{\delta\Pi}^*$ – maximal acceptable average relative difference between operational and testing defectiveness of the engine.

Taking into account profit from the engines and operation Pr , function Φ_Σ can be presented as convolution of factors $\delta\Pi$, K_A и Pr:

$$\Phi_\Sigma = \max\sqrt{\alpha_1\left(\sum_{i=1}^{n}\left(\left(\Pi_{op.i.\nu_i}[P_0, R_{op}(\tau_{op}), \tau_{op}]\right) \Big/ \left(\Pi_{Ti}[P_0, R_T(\tau_T), \tau_T] - \Pi_{op.i.\nu_i}[P_0, R_{op}(\tau_{op}), \tau_{op}]\right)\right)^2\right)^2 + \alpha_2\left(\frac{\tau_{op}}{\tau_T}\right)^2 + \alpha_3 Pr^2}\,; \qquad (6)$$

$$\overline{\delta\Pi} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\Pi_{Ti} - \Pi_{op.i.\nu_i}}{\Pi_{op.i.\nu_i}}\right] \le \overline{\delta\Pi}^*;$$

$$R_T(\tau_T), R_{op}(\tau_{op}) \in G_R; \quad \tau_T, \tau_{op} \in G_\tau; \quad \nu_i \in [1 \ldots N_{op}].$$

While optimizing, it is necessary to consider $\Pi_{op.i.\nu i}$ in Eq. 6 to be of different value (Fig. 2) caused by operation strategy, engine production quality etc. One of the main findings of this approach is applying intelligent control and management technologies for decision making on possibility of efficiency increasing.

The initial data for decision making are lifecycle modeling results, and the range of implemented intelligent technologies is wide as well.

IV. Big Data Concept Implementation

One of the main reasons of limiting the ability to build a simulation model of the engine lifecycle for selection of fatigue test parameters is the difficulty of assessing the efficiency criterion of fatigue test in the framework of the engine life cycle.

The cause for this kind of problem firstly is the need for preprocessing large volumes of heterogeneous data, collected at different stages of lifecycle of the complex technical object. The main data and knowledge flows in modeling lifecycle of the aircraft engine are shown in Fig. 3, where DB_1, DB_2 – integrated databases; KB_2 – knowledge base; D_1 – formal GTE description; D_2, D_3, D_4, D_5 – data necessary for GTE design, production, testing and operation; K_1 – GTE operation knowledge; K_2 – control of test parameters; K_3 – control of operation parameters; K_4 – reclamation control.

Under these circumstances, Big Data technology can serve as an effective tool for storing and analyzing the life cycle information. For example, MapReduce allows performing batch processing queries database that very significantly reduces the processing time. Hadoop effectively manages distributed resources based on streaming data access. These technologies are suitable for preprocessing source model data as well as simulation results.

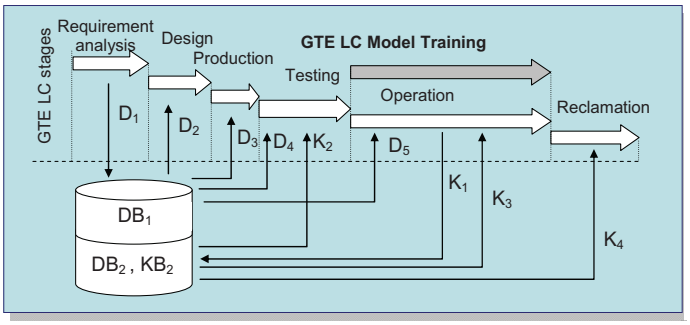


Fig. 3. Data and Knowledge Flows in GTE LC Modeling

V. Conclusion

Aircraft engines can be considered as complex technical systems consisting of several thousands of parts and components. Engine condition during the lifecycle is characterized by numerous parameters. It is offered to increase efficiency of fatigue testing the aircraft engines is possible with its assessment on base of relationship with the economic effect from engines operation. For this purpose, the simulation model of engine life cycle is implemented. To build the lifecycle model and optimize test parameters, Big Data concept is proposed.

References

[1] A. Guichvarov, N. Kondratieva, "Technical and economic assessment of aircraft engines fatigue testing on base of simulation modeling," in Proc. of AIAA/ASME/SAE/ASEE Joint Propulsion Conf. & Exhibit. 37th, Salt Lake City, Utah, 8-11 July, 2001, AIAA-2001-3817.

[2] A. S. Guichvarov, N. V. Kondratyeva, "Method of GTE fatigue tests optimization on the basis of lifecycle modeling," in Aviation Industry, 2003, No. 3, pp. 21-25.

[3] N. V. Kondratyeva, "Determination of endurance test parameters for aircraft based on the life cycle simulation and Big Data technology," in Proc. of the 1st Russian Workshop "Big Data in aviation", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 27 April 2015. Ufa: USATU, 2015, pp. 30-32.

[4] N. V. Kondratyeva, "GTE endurance test justification in framework of the life cycle with implementing Big Data technologies," in Proc. of the 2nd Russian Conference "Free Flight-2015", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 22-24 December 2015. Ufa: USATU, 2016, pp. 93-99.

[5] A. Adamov, "The Truth about Cloud Computing as new Paradigm in IT," in Proc. of the 3rd IEEE International Conference on Application of Information and Communication Technologies (AICT2009), Azerbaijan-Baku, 14-16 October, 2009.

[6] G. Smorodin, O. Kolesnichenko, "Big Data as the big game changer: Big Data-driven world needs Big Data driven ideology," in Proceedings of The 9th IEEE International Conference on Application of Information and Communication Technologies (AICT2015), Russia, Rostov-on-Don, 14-16 October, 2015, pp. 40-43.

[7] Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Eds. EMC Education Services. Indianapolis: Wiley & Sons, Inc., 2015, pp. 55-86.

[8] Information Storage and Management: Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments, 2nd ed., Eds. EMC Education Services. Indianapolis: Wiley & Sons, Inc., 2015, pp. 67-93.

[9] S. S. Valeev, "Big Data technologies in aviation," in Proc. of the 2nd International Conference "Information Technologies for Intelligent Decision Making Support", 15-17 May, 2014, Ufa: USATU, 2014, pp. 150-152.

[10] S. S. Valeev, A. D. Kulakov, "Information technologies Big Data in Aviation," in Proc. of the 1st Russian Workshop "Big Data in aviation", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 27 April 2015. Ufa: USATU, 2015, pp. 21-23.

[11] S. S. Valeev, A. D. Kulakov, N. V. Kondratyeva, "The problems of Big Data processing in aviation," in Proc. of the 2nd Russian Conference "Free Flight-2015", Zhukovskiy-Ufa, M. M. Gromov Flight Research Institute (FRI)-USATU, 22-24 December 2015. Ufa: USATU, 2016, pp. 9-15.

[12] S. S. Valeev, N. V. Kondratyeva, "Reference model for large scale industry management based on Big Data concept," in Proc. of the 4th International Conference "Information technologies for Intelligent Decision Making Support," Vol. 1, Russia, Ufa, May 17-19, 2016. Ufa: USATU, 2016, pp. 15-18.

Batch Clustering Algorithm for Big Data Sets

Rasim Alguliyev [1], Ramiz Aliguliyev [1], Adil Bagirov [2], Rafael Karimov [1]
[1] Institute of Information Technology of Azerbaijan National Academy of Sciences
Baku, Azerbaijan
rasim@science.az, r.aliguliyev@gmail.com, rafael.karimov@gmail.com
[2] Federation University Australia
Ballarat, Australia
a.bagirov@federation.edu.au

***Abstract* — Vast spread of computing technologies has led to abundance of large data sets. Today tech companies like, Google, Facebook, Twitter and Amazon handle big data sets and log terabytes, if not petabytes, of data per day. Thus, there is a need to find similarities and define groupings among the elements of these big data sets. One of the ways to find these similarities is *data clustering*. Currently, there exist several data clustering algorithms which differ by their application area and efficiency. Increase in computational power and algorithmic improvements have reduced the time for clustering of big data sets. But it usually happens that big data sets can't be processed whole due to hardware and computational restrictions. In this paper, the classic *k*-means clustering algorithm is compared to the proposed *batch clustering* (BC) algorithm for the required computation time and objective function. The BC algorithm is designed to cluster large data sets in batches but maintain the efficiency and quality. Several experiments confirm that batch clustering algorithm for big data sets is more efficient in using computational power, data storage and results in better clustering compared to *k*-means algorithm. The experiments are conducted with the data set of 2 (two) million two-dimensional data points.**

***Index Terms* — Big Data, Big Data Clustering, Clustering Algorithms, k-means, Batch Clustering.**

I. Introduction

Clustering of big data has gained popularity in the last decade due to increased demand to process of large data sets. Data clustering algorithms were developed as a powerful tool to analyze and relate large amount of data generated by different systems. Many applications produce large amounts of data that demand efficient data processing algorithms. The main purpose of clustering is to find similarities among the elements of the given data set and categorize them into groups by a certain metrics. That is clustering is a process of organizing objects or elements into groups whose members are similar to each other. So the purpose of clustering is to define groups in a set of unlabeled data. Data clustering is considered as one of the important *unsupervised learning* techniques [1]. Today data clustering has got wide areas of applications in marketing, online sales, biology and medicine, libraries, banking and insurance, city-planning and utilities, earthquake studies, WWW, Internet of Things (IoT), etc. [2]. Currently there are several data clustering (a.k.a. *unsupervised classification*) algorithms that differ by their efficiency and quality [3]. The question then arises: how to define the best algorithm for the clustering problem? It is possible to demonstrate that there is no "best" method which does not depend on the purpose of clustering and the given data set. So it is up to the user who can select criterion in the way that the outcome of clustering will satisfy his/her requirements. Generally, there are several requirements that clustering algorithms should satisfy:

- Clustering algorithm should be scalable, i.e. it can be applicable to a growing volume of data sets;
- The algorithm should handle different attribute types as the scope of application can be different;
- The algorithm should be able to define clusters of any shape;
- There should be minimal knowledge about the domain when defining parameters;
- Ability to filter noise and define outliers;
- The algorithm should be able to cluster data sets of high dimensions;
- The result of clustering should be interpretable and usable.

In this paper, it is generally mentioned about some of the main data clustering algorithms and classic *k*-means is compared to the proposed *batch clustering* (BC) algorithm. The main clustering types (algorithms) can be classified as follow but they are not limited to the below [12]:

- Density-Based Clustering;
- Exclusive or Partition-Based Clustering;
- Overlapping or Fuzzy Clustering;
- Model-Based Clustering;
- Hierarchical (Agglomerative or Divisive) Clustering;
- Grid-Based Clustering;
- Probabilistic Models (Gaussian or Bernoulli Mixture Models) for Clustering.

In *Exclusive or Partition-Based Clustering* data are assigned into groups in an exclusive (excluding) method, i.e. if a certain data element or object assigned to a certain cluster then it cannot be assigned to a different cluster.

But on the contrary, in *Overlapping Clustering* a certain data element can be assigned to several clusters at the same time with a certain coefficient (degree of membership). This membership is defined through fuzzy sets.

In *Hierarchical Clustering* algorithm each nearest two clusters are assigned to each other, i.e. defined as the union between them. In the first condition every datum is treated as a cluster. After some iteration we get the wanted final clusters.

Another kind of clustering which is based on the probabilistic approach is called *Probabilistic Clustering* algorithm.

The below are the highly used algorithms for data clustering:

- K-means Clustering and its variations (Kernel K-means, Weighted K-Means, Genetic K-Means, etc.);
- Fuzzy C-means (FCM);
- Hierarchical (Agglomerative or Divisive) Clustering (Single and Complete Link, etc.);
- Bernoulli Mixture Model;
- Mixture of Gaussians;
- Expectation-Maximization Algorithm.

Any of the mentioned algorithms can be related to certain type of clustering methods indicated above. For example, k-means can be considered as an exclusive (excluding) clustering method [4], [5]. But Fuzzy C-means (FCM) is related to an overlapping or Fuzzy clustering method [7], [8]. According to its name, hierarchical clustering is distinct [6], [9], [10], [11]. Mixture of Gaussian or Bernoulli algorithm belongs to the probabilistic clustering type [13].

For comparison the well-known *k*-means algorithm will be used. Theoretically, a new approach was introduced to classify big data sets in batches, which is effective in reduced computational resources. Empirically, several experiments were conducted to evaluate the efficiency and quality of the proposed method by using big data sets. The BC algorithm is more efficient for clustering of big data sets. Indeed, batch clustering algorithm is based on the *k*-means algorithm and can be considered as a new version of *k*-means, but offers a new approach in processing of large data sets of several dimensions. Numerical experiments over large data sets have shown that BC algorithm is more efficient over classic *k*-means algorithm in using computational power of hardware with restricted resources.

For qualitative estimation of the proposed algorithm we will compare required calculation time for clustering (centroids and assigning elements to centroids) and the value of *objective function* for classic *k*-means algorithm and the BC algorithm. We will use data set with synthetically generated 2 (two) million and 2-dimensional data elements. But similar experiments were conducted over the real data sets taken from UCI Machine Learning Repository.

II. Related Work

What do we mean when we talk about Big Data and related fields? What volume of data we consider as big? Facebook alone processes more than half petabytes (PB) of data daily. The Library of Congress has lately classified more than 170 billion tweets, i.e. all the tweets from the beginning of Twitter. Can we assume it big? If we cannot load and process the data in computational memory then we treat the data as big [14].

Big Data has several major problem areas: volume, variety, velocity, validity, veracity and volatility. Here we will deal with the limited storage and processing volume problems of Big Data and main clustering approaches. As we know a large data set can be problem to store and process in the working memory of the computational system due to storage limitations.

There are two main methods of big data sets clustering: distributed clustering in parallel systems and sample clustering by taking samples according to some rules. Each of these methods has been applied to many of the clustering methods, like *k*-means and its variations, hierarchical, density-based clustering, probabilistic, etc.

Recent works in accelerating calculation of the *minimal spanning tree* (MST) for clustering of big data involve the *DualTreeBoruvka* (DTB) method [16], and the *Filter-Kruskal* (FK) method [15]. The proposed methods in [15] and [16] focus mainly on the rapid MSTs, and because of their accuracy, it is possible to obtain hierarchical *single linkage* (SL) clustering partitions from Minimal Spanning Tree. One of the hierarchical single linkage algorithms for big data clustering that is known about is ***clu****stering with* ***s****calable* ***v****isual* ***a****ssessment of* ***t****endency* (clusVAT) algorithm proposed by Timothy Craig Havens, James C. Bezdek et al. in [17]. There it is named as sVAT-SL. That algorithm creates accurate single linkage partitions in a minor class of data sets, but with broad set of inputs the clusVAT algorithm leaves the single linkage pattern. In the paper [14], J. C. Bezdek et al. have compared clusiVAT to SL clusters constructed with fast MST that is the nature, quantity, velocity and other characteristics of clusiVAT partitions over big data single linkage partitions taken from Filter-Kruskal Minimal Spanning Trees.

In the paper [17] by T. C. Havens, J. C. Bezdek et al., the *scalable Visual Assessment of Tendency* (sVAT) method has been extended to output hierarchical SL partitions in big data clustering. The proposed sVAT method was developed to ensure visible confirmation of cluster counts in large data sets. The improvement described in [17] for sVAT makes possible to output the SL partition with the visible confirmation conveniently. The proposed sVAT method is an extensible approach for imaging the clusters and their counts in a given big data set. This algorithm first marks a representative sample, after that creates an image that indicates the cluster counts as the number of black squares on the diagonal of the contrast picture. The sVAT-SL method calculates an SL partition of the sVAT marked data, after that expands that partition for the whole data set.

III. Proposed Method

As we indicated, clustering of big data sets can be a problem due to memory and other computational restrictions. In this paper, it is proposed to cluster a given large data set in batches by using *k*-means algorithm. That is take some portions of data elements from the given data set and process it. Then take next portion and process it and so on until all the elements of data set are processed. After that the whole given

data set is also clustered by *k*-means algorithm for efficiency and quality comparison. Later qualitative indicators for both of these approaches are measured. For qualitative indicators the below are considered:

- Time T required for calculation of centroids and assigning data elements to centroids (clusters);
- The value of the objective function J (squared error function):

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a given distance (Euclidean) measure between a data element $x_i^{(j)}$ and the cluster center c_j, i.e. J is an indicator of the Euclidean distance of the n data elements from their corresponding cluster centers.

Assuming that we have a data set consisting of n data elements, m is a number of elements in a small subset of data set, k is a number of centroids, the proposed *batch clustering* algorithm can be defined as follows:

1. We take m number of data elements from the given data set of n elements ($m < n$);
2. m number of data elements are processed in RAM of computer by the *k*-means algorithm to find k number of centroids ($k < m$);
3. Then take the k centroids along with $m - k$ number of data elements from the remaining data set;
4. Repeat steps 2 – 3 until all the elements of the initial data set are processed;
5. Assign all elements to the k centroids calculated in the last step 4.

IV. Experiment

In this paper, we will compare the clustering results of classic *k*-means algorithm with the proposed batch clustering algorithm which is also based on *k*-means. The result of experiment will be compared for qualitative indicators such as time required for calculation of centroids and assigning data elements to centroids; and objective function J (squared error function).

The parameter of PC on which the experiment is conducted is as follows:

- CPU: Core i7 2.2 GHz
- RAM: 8 GB
- Video Card: 1 GB
- HDD: 1 TB
- Microsoft Windows 7 Ultimate 64 bit

Initial data for experiment are below:

- Number of elements in data set: 2,000,000
- Dimension of data element: 2-dimensional
- Size of data file: 100 MB
- Source: synthetically generated
- Number of centroids: 10

The result of calculations is below:

Classic k-means

- Time spent calculating centroids: $T = 225$ sec.
- Time spent calculating objective function: $T = 979$ sec.
- Objective function: $J = 33997$.

Batch clustering

- Number of elements in each portion: $m = 100{,}000$ elements.
- Time spent calculating centroids: $T = 223$ sec.
- Time spent calculating objective function: $T = 932$ sec.
- Objective function: $J = 33996$.

As seen from the above calculation results, batch clustering algorithm produces better results over classic *k*-means algorithm. Besides, we have a big gain in using computational power with restricted resources.

V. Conclusion

As mentioned above clustering of Big Data sets is a problem due to computational power restrictions. There are some research papers in this field to handle this challenge. In this paper, we proposed batch clustering algorithm which is based on classic *k*-means algorithm, but assumes clustering of large data sets in a chosen portions, thus efficiently using computational power. A number of computational experiments have been conducted over synthetically generated data sets and real data sets taken from UCI Machine Learning Repository. For each experiment the elements of data sets have been chosen with different dimensions. As seen from the experiment the proposed BC algorithm produces better results compared to classic *k*-means algorithm. The main advantage of the BC algorithm is that it makes possible to process the elements of large data sets in batches, thus enabling of handling any size of big data set.

References

[1] Charu C. Aggarwal, Chandan K. Reddy "Data Clustering. Algorithms and Applications", CRC Press, Taylor & Francis Group, 2014

[2] Osmar R. Zaïane, "Principles of Knowledge Discovery in Databases – Chapter 8: Data Clustering", *University of Alberta*, 1999. [Online]. Available: http://www.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html. [Accessed: 03- May- 2016].

[3] J. B. MacQueen: "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297, 1967.

[4] Andrew Moore, "K-means and Hierarchical Clustering – Tutorial Slides", *Carnegie Mellon University*, 2001. [Online]. Available: http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html. [Accessed: 03- May- 2016].

[5] Brian T. Luke, "K-Means Clustering", 2007. [Online]. Available: http://www.btluke.com/kmeans.html. [Accessed: 03- May- 2016].

[6] Hans-Joachim Mucha and Hizir Sofyan, "Nonhierarchical Clustering", *Humboldt University of Berlin*, 2000. [Online]. Available: http://edoc.hu-berlin.de/series/sfb-373-papers/2000-49/PDF/49.pdf. [Accessed: 03- May- 2016].

[7] J. C. Dunn, “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters”, Journal of Cybernetics 3: 32-57, 1973.

[8] J. C. Bezdek, “Pattern Recognition with Fuzzy Objective Function Algoritms”, Plenum Press, New York, 1981.

[9] S.C. Johnson, “Hierarchical Clustering Schemes” Psychometrika, 2:241-254, 1967.

[10] R. D'andrade, “U-Statistic Hierarchical Clustering” Psychometrika, 4:58-67, 1978.

[11] Stephen P. Borgatti, “How to explain hierarchical clustering”, *Analytic Technologies*, 1994. [Online]. Available: http://www.analytictech.com/networks/hiclus.htm. [Accessed: 03- May- 2016].

[12] Maria Irene Miranda, “Clustering Methods and Algorithms”, *Indian Institute of Technology Bombay*, 1999. [Online]. Available: http://www.cse.iitb.ac.in/infolab/Data/Courses/CS632/1999/clustering/dbms.html. [Accessed: 03- May- 2016].

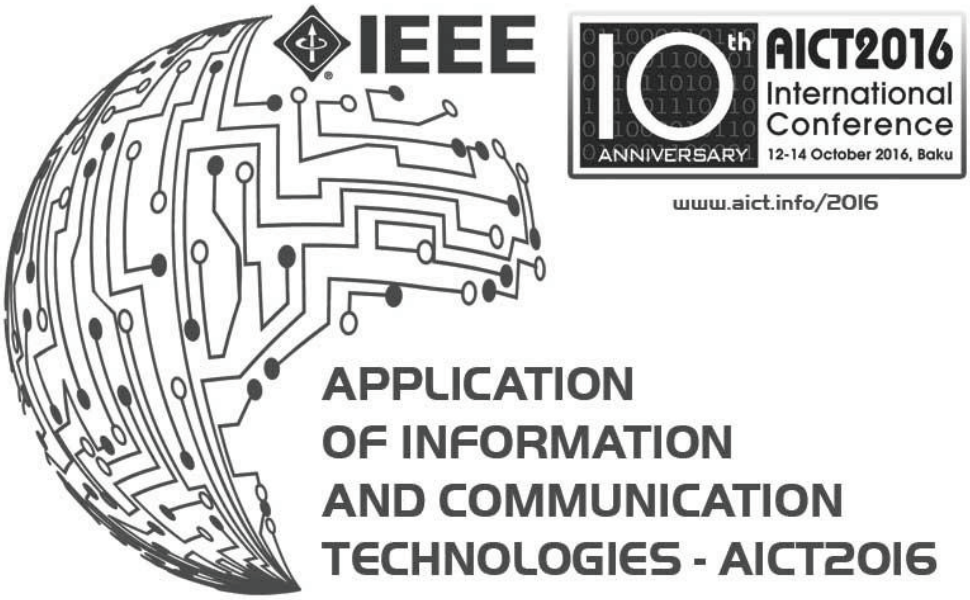
[13] A.P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum Likelihood from Incomplete Data via the EM algorithm”, Journal of the Royal Statistical Society, Series B, vol. 39, 1:1-38, 1977.

[14] D. Kumar, M. Palaniswami, S. Rajasegarar, C. Leckie, J. C. Bezdek, T. C. Havens “clusiVAT: A mixed visual/numerical clustering algorithm for big data”, ResearchGate, Conference Paper, October 2013.

[15] V. Osipov, P. Sanders and J. Singler, "The Filter-Kruskal minimum spanning tree algorithm", in Proc. ALENEX, Jan. 2009, New York, SIAM, 52-61, 2009.

[16] W. B. March, P. Ram and A. G. Gray, “Fast Euclidean Minimum Spanning Tree: Algorithm, Analysis and Applications”, in Proc. 16th ACM SIGKDD, ACM, NY, 603-612, 2010.

[17] T. C. Havens, J. C. Bezdek and M. Palaniswami, "Scalable single linkage clusters for big data," in Proc. 2013 IEEE ISSNIP, 396-401, 2013.



CONFERENCE PROCEEDINGS

SESSION 2.

DATA MINING AND DATA ENGINEERING

12-14 October 2016, Baku, Azerbaijan

Evaluation of Machine Translation Output in Context of Inflectional Languages

Daša Munková
Department of Translation Studies
Constantine the Philosopher University in Nitra
Nitra, Slovakia
dmunkova@ukf.sk

Jozef Kapusta
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
jkapusta@ukf.sk

Michal Munk
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
mmunk@ukf.sk

Jaroslav Reichel
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
jaroslav.reichel@ukf.sk

Abstract — **Objective of the paper is to evaluate metrics of automatic evaluation of machine translation output using manual metrics – fluency and adequacy. We tried to answer the question to which extent the manual evaluation correlates with the automatic evaluation of MT output from/to Slovak to/from English. We focused on metrics based on the similarity and statistical principles (WER, PER, CDER and BLEU-n). We found out, that the manual evaluation, namely fluency and adequacy metrics correlates with automatic metrics of MT evaluation for less spoken language and low resource language such as Slovak. The contribution also consists of system proposal for both, manual (based on POS tagging) and automatic (based on reference) evaluation of MT output.**

Index Terms — **Manual evaluation, Automatic evaluation, Inflectional language, Natural language processing**

I. Introduction

If we want to improve the performance of machine translation (MT) system, it is inevitable its evaluation through its output. Evaluation of MT system it as old as machine translation itself [1-6]. It has been much written about the machine translation yet. Carbonell and Wilks [7] remarked that MT evaluation is better understood that MT itself, even that it is its inseparable part.

There are two main ways how to assess the MT output. Manually (human) or automatically, both with original methodologies with a wide range of proposed methods. Though the manual evaluation, using the intrinsic and extrinsic methods, we assess accuracy and usability of the MT output.

Intrinsic methods (evaluation scales, ranking, error analysis, etc.) assess subjectively the quality of MT output based on their comparison with reference (a "golden standard"). In most cases, the evaluators assess subjectively the attributes of the translation quality such as adequacy or fluency of MT output. In other words, intrinsic methods focus on the language quality. Evaluation scales comprise two attributes of translation quality – fidelity and intelligibility. They were very frequently used during the MT evaluation campaigns, e.g. DARPA [3], IWSLT (International Workshop on Spoken Language Translation, [8]), WMT (Workshop on Statistical Machine Translation, [9]) a NIST open workshop of MT evaluation (NIST Machine Translation Evaluation Workshop) organized by LDC (Linguistic Data Consortium) (LDC, 2002, 2005; [10]). Evaluation scales and ranking assess the translation quality in terms of "correctness" by contrast to error analysis, which measures "error rate" of MT output. In a way, the error analysis demonstrates objectively and explicitly the lack of accuracy (fidelity) or clarity of the MT output. Based on Llitjós et al. work [12], Vilar et al. [11] introduced a framework for error analysis and error classification, where five categories of errors were classified: missing words (missing content words/lexical errors, or missing filler words/ functional semantic categories), word order (word order on word level (a local shift within the phrases) or on phrase level (relocation of the phrase within one sentence), incorrect words (inappropriate equivalent), unknown words and punctuation. This classification has become widely used in recent years [13]. Extrinsic methods, so called task-oriented methods such as post-editing or reading with understanding are focused on the effectiveness of MT output with the respect to specific task. The aim of the last mentioned is an extraction of the crucial information from the MT output. Laoudi et al. [14] proposed a multiple level of information extraction of MT output, so called wh-questions. Lo and Wu [15] formed a metric called MEANT. The principle of this metric is the usability of MT output, i.e. to what extent the given MT output can help a reader to understand successfully "who did what to whom, when, where, why and how".

In this paper, we examine MT quality in relation to Slovak language as an inflectional type of language and with morphological complex system often causing inadequate translation, especially in the case of machine translation (MT

output). Slovak language is characterized by the robust usage of morphology (word modification and word formation using inflections- endings or vowel changes) and complex agreement system with a loose word order (verb agrees with subject, direct object and indirect object). Compare to English, Slovak has more grammatical endings and more flexible word order than English with a few inflections and a fixed word order.

The objective of this paper is to answer the question to which extent the manual evaluation (metrics of manual MT evaluation) correlates with the metrics of automatic evaluation of MT output in special case from/to inflectional language to/from analytical language (from/to Slovak to/from English). For the needs of our work, we will only focus on metrics based on the similarity and statistical principles that are widely used in the automatic evaluation of machine translation outputs (WER, PER, BLEU-n, etc.).

Many experiments showed that automatic metrics for MT evaluation correlate with human judgment [17-20]; whereby human judgment is in the form of "adequacy and fluency" quantitative scores [21].

A number of researches [17-20] focusing on the automatic evaluation of MT output has been realized, especially for dominant languages such as English, German or Spanish. In these researches were proved, that metrics of intrinsic automatic MT evaluation correlate well with human intrinsic metrics- fluency and adequacy (using the scores of these metrics) [21]. However, there are no researches or a few focusing on MT evaluation of Slovak language. For the reason, our partial goal presented in this paper is to provide an insight into MT evaluation and machine translation itself in context of European, minority and low resource language - Slovak.

II. System for Evaluation of MT Output

A system for manual evaluation of machine translation output was created. In our case, manual evaluation (human evaluation) comprises evaluation scales (fluency and adequacy), post-editing and error classification. The system offers a simple interface for effective human evaluation, i.e. the user has a space for post-editing; an option to mark fluency and adequacy of MT output on the scales from 1 to 5 (where 5 means all meaning/flawless language, 4 means most meaning/good language, 3 means much meaning/non-native language, 2 means little meaning/disfluent language and 1 means none/incomprehensible) and also an option to determine an occurred error, (whether the error belongs to language, accuracy, terminology or style). The system enables to determine several errors (more then one) for the given sentence. Within the error, evaluator has to decide if the occurred error is minor, major, critical or neutral.

For a better description of the relationships between the system and its user, we defined the following functional requirements for the application:

- to upload and process a file with the source text,
- to upload and process a file with MT output, which is already aligned with the source text (texts were aligned using the algorithm and software hunalign [22]),
- to select a sentence for post-editing and to post-edit the given sentence,
- to mark the fluency and adequacy of MT output on the evaluation scales (from 1 to 5),
- to assign basic error category of errors (language, accuracy, terminology and style) and the extent of error rate (from neutral to critical).

System was created as a web application due to simplicity and accessibility of the solution, necessity of a central entries database and the need of accesses of several users.

By the system design for manual evaluation of machine translation output, various diagrams were created, which help to explain its implementation. The ERD scheme of the application with database design were the basis of the system (Fig. 1).

III. Method

A. Data

Data were collected from the origin Slovak textbook dealing with the issue of machine translation. We chose this text because of the terminology, many expressions and abbreviations were written in Slovak as well as in English and the second important thing was that we had its human translation (English version, translation by 2 human translators and checked by a native speaker specialist in machine translation). Slovak text was translated into English by Google Translate Service (a free online machine translation application based on Moses). Later, MT output was evaluated by 14 human translators. First, they assessed fluency and adequacy of each sentence. They used manual intrinsic method- 5-point evaluation scale for adequacy as well as for fluency. Later, they used another manual intrinsic method, post-editing, to post-edit the MT output. Both evaluation, manual and automatic, were carry out within our created system evaluation MT output.

B. Metrics

We chose 7 automatic metrics, 3 represent error rates (WER, PER and CDER) and 4 represent accuracy (BLEU-1 to BLEU-4). By the metrics of error rate the higher score/value, the lower the translation quality. On the other, by the metrics of accuracy, the higher score, the better the translation quality. Word error rate (WER), Position-independent error rate (PER) and Cover disjoint error rate (CDER) are based on edit distance, i.e. on Levenshtein distance measuring the similarities between two strings/sequences (one source-MT output and one target- reference sequences and a distance represents a number of insertions, deletions and substitutions needed to transform a source sequence to target. Bilingual evaluation understudy (BLEU-n) is based on n-grams. Intrinsic metrics of error rate and intrinsic metrics of accuracy for automatic MT evaluation are described in more detail in [28] and [22].

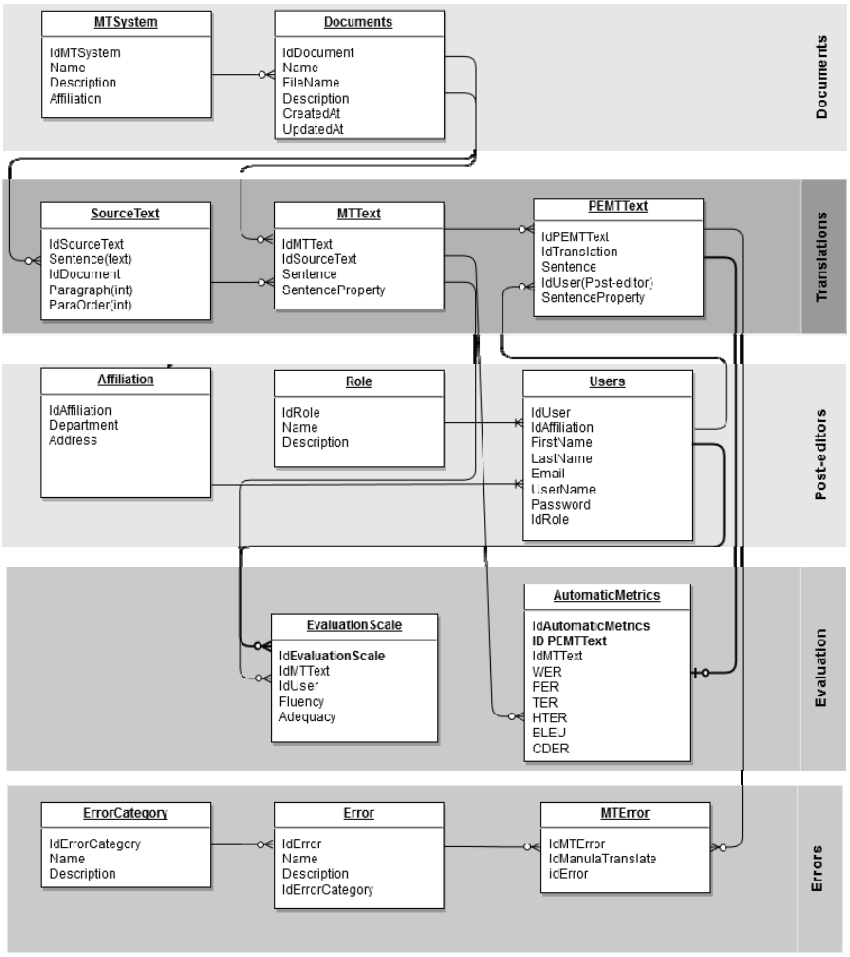


Fig. 1. Application of the ERD scheme.

C. Analysis of Validity of the Intrinsic Automatic MT Metrics of the Evaluation

The aim of the research was to assess the translation quality of MT output. We used the analysis of validity to evaluate intrinsic automatic metrics of MT output using the manual intrinsic metrics for MT evaluation, the fluency (F) and adequacy (A).

In other words, to answers the question to which extent the manual evaluation correlates with the automatic evaluation of MT output from Slovak to English.

We used our own created tool (system) for evaluation, in which algorithms of automatic metrics were implemented. It computes the scores based on the similarities between the hypotheses (MT outputs) and references (human translations-gold standard).

D. Results

14 human translators assessed the translation quality of MT output using the metrics fluency and adequacy.

Scores of intrinsic human metrics, fluency (F) and adequacy (A), were the criterions for validity of the intrinsic automatic metrics for MT evaluation. Scores of fluency were measured on the 5-poit scale, where 1 means incomprehensible, 2 disfluent, 3 not-native, 4 fluent and flawless language.

Scores of adequacy were also measured on the 5-poit scale where 1 means none, 2 a little, 3 much, 4 most and 5 means all meaning.

TABLE I. CORRELATIONS AMONG THE SCORES OF AUTOMATIC METRICS OF ERROR RATES AND FLUENCY AND ADEQUACY.

	MT_PER	MT_WER	MT_CDER
Fluency (F)	-0.490	-0.508	-0.553
Adequacy (A)	-0.433	-0.439	-0.481

The results showed (Tab. 1), there is a medium to large degree of linear dependency between (PER, WER and CDER) and (F and A), i.e. between error rates and fluency, as well as between error rates and adequacy. The values are changing together in the reverse direction (Fig. 2).

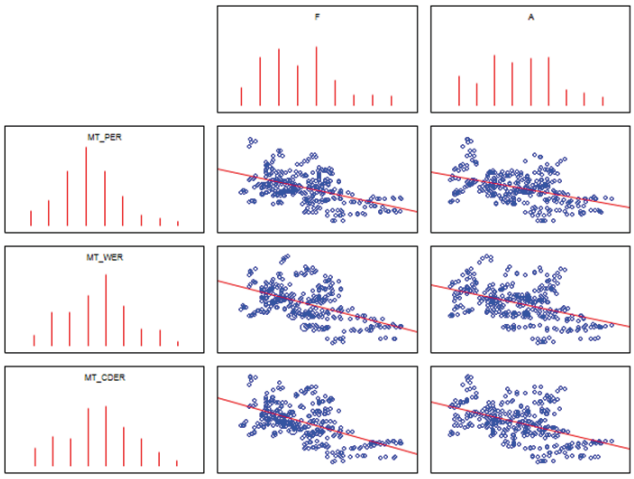


Fig. 2. Visualisation of correlations of metrics (PER, WER and CDER) and metrics (adequacy and fluency) of MT evaluation.

The correlation coefficients are statistically significant (Tab. 2).

TABLE II. VALUES OF SIGNIFICANCE OF CORRELATIONS OF THE AUTOMATIC METRICS OF ERROR RATES AND FLUENCY AND ADEQUACY.

	MT_PER	MT_WER	MT_CDER
Fluency (F)	0.0000	0.0000	0.0000
Adequacy (A)	0.0001	0.0001	0.0000

The same was proved between the automatic metrics of accuracy and fluency and also between them and adequacy, i.e. the correlation coefficients are statistically significant. The values of metrics of accuracy and values of fluency and adequacy are changing together in the same direction (Fig. 3).

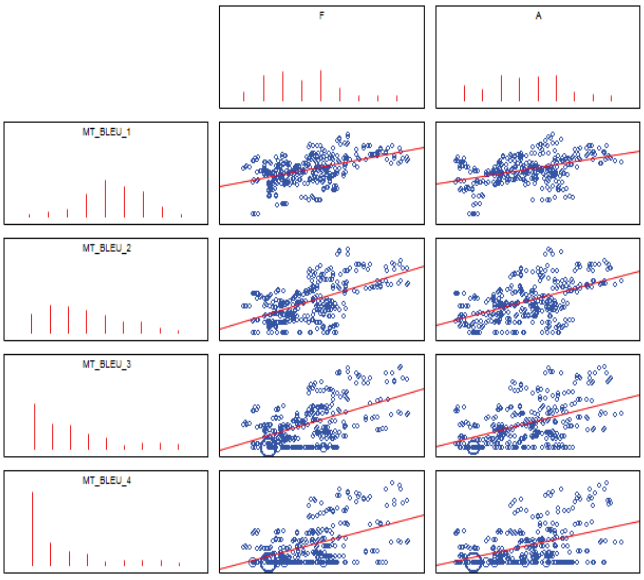


Fig. 3. Visualisation of correlations of metrics (BLEU-n) and metrics (adequacy and fluency) of MT evaluation.

The values of multiply correlation coefficients achieved the value almost 0.6 in case of fluency ($R_{F,(PER,WER,CDER)} = 0.559$, $R_{F,(BLEU-1,BLEU-2,BLEU-3,BLEU-4)} = 0.573$) and also in the case of adequacy ($R_{A,(PER,WER,CDER)} = 0.483$, $R_{A,(BLEU-1,BLEU-2,BLEU-3,BLEU-4)} = 0.532$).

Validity is the degree to which the assessment measures what it is intended to measure. It is closed tied with reliability, i.e. the lower reliability, the lower concurrent validity (the lower the correlation among automatic and manual MT metrics), which results in calculation of a corrected coefficient of correlation. In both cases (adequacy and fluency) coefficients of corrected coefficient of correlation achieved values higher than 0.6, which means, metrics of error rate for automatic MT evaluation can be regarded as valid.

IV. CONCLUSION AND FUTURE DIRECTION

Williams [23] claims, if the translation quality assessment, done manually or automatic, it should be as objective as possible, its methodology with metrics and have to be reliable and valid. However, the manual evaluation, using the intrinsic and extrinsic methods, is regarded as the most reliable, bit is highly time and labor consuming; expensive and requiring trained bilingual human reviewers) [24].

Results of the presented research showed, that the manual evaluation correlates with automatic evaluation, namely intrinsic method (evaluation scales-fluency/adequacy) with metrics of error rate or correctness of automatic MT evaluation in context of less spoken language such as Slovak. They showed that the automatic metrics are also valid for Slovak language. We verified the concurrent validity of automatic metrics for MT evaluation by the manual MT metrics using intrinsic method (evaluation scales – fluency and adequacy) and a corrected multiple correlation coefficient between the manual and automatic MT metrics of evaluation.

We also presented a system for the evaluation of MT system, where the text representation was realized using the transaction-sequence model which is detail described in [25-27]. The system consists of two sub-systems, one comprising intrinsic methods (evaluation scales and error analysis) and extrinsic method (post-editing) of manual evaluation of MT; the second is focused on automatic evaluation (metrics of error rate/correctness) and automatic error analysis (error classification), based on POS tagging.

For the future direction, we want to check whether it is valid vice versa, i.e. if the intrinsic automatic MT metrics correlate with error analysis (intrinsic method) or with post-editing (extrinsic method). Further, we follow up with metrics of error rate and classification of MT errors (which error rate is related to error category) and automate this procedure based on the morphological annotation (e.g. POS tagging).

We believe it will be useful not only for translation community (human or machine), but also for the community of natural language processing or philology.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-14-0336 and Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contracts No. VEGA-1/0559/14.

REFERENCES

[1] J, Hutchins, "Current commercial machine translation systems and computer-based translation tools," in International Journal of Translation, vol. 17, no. 1-2, 2005.

[2] D, Arnold, "Why translation is difficult for computers," in H. L. Somers (Ed.), Computers and Translation: A Translator's Guide, 2003, pp. 119-142.

[3] J. S. White, "How to evaluate machine translation," in H. Somers (Ed.), Computers and Translation: A translator's guide, 2003, pp. 211-244.

[4] M. King, A. Popescu-Belis, and E. Hovy, "FEMTI: Creating and using a framework for MT evaluation," in Proceedings of Machine Translation Summit IX., 2003, pp. 224-231.

[5] C. Callison-Burch, M. Osborne, and P. Koehn, "Re-evaluating the role of BLEU in machine translation research," in EACL-06, 2006, pp. 249-256.

[6] C. Callison-Burch, C. Fordyce, P. Koehn, Ch. Monz, and J. Schroeder, "(Meta-) evaluation of machine translation," in StatMT '07, 2007, pp. 136-158.

[7] J. Carbonell and Y. Wilks, "Machine Translation: An In-Depth Tutorial," in The 29th Annual Meeting of the ACL, 1991.

[8] M. Paul, "Overview of the IWSLT 2009 evaluation campaign," in IWSLT-09, 2009, pp. 1-18.

[9] P. Koehn, B. Haddow, P. Williams, P., and H. Hoang, "More Linguistic Annotation for Statistical Machine Translation," in Proceedings of the Fifth Workshop on Statistical Machine Translation and Metrics MATR, 2010.

[10] G. Doddington, "Automatic evaluation of machine translation quality using n-gram cooccurrence statistics," in HLT-02, 2002, pp. 138-145.

[11] D. Vilar, G. Leusch, H. Ney, and R.E. Banchs, "Human evaluation of machine translation through binary system comparisons," in StatMT '07, 2007, pp. 96-103.

[12] A. F. Llitjós, J.G. Carbonell, and A. Lavie, "A framework for interactive and automatic refinement of transfer-based machine translation," in Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT 05), Budapest, 2005, pp. 87–96.

[13] M. Popovic and H. Ney, "Word error rates: Decomposition over POS classes and applications for error analysis," in ACL '07-WSMT, 2007.

[14] J. Laoudi, C. R. Tate, and C. R. Voss, "Task-based MT evaluation: From who/when/where extraction to event understanding," in LREC-06, 2006, pp. 2048-2053.

[15] C. Lo and D. Wu, "MEANT: An inexpensive, high-accuracy, semiautomatic metric for evaluating translation utility via semantic frames," in ACL-11, 2011, pp. 220-229.

[16] C. Callison-Burch, M. Osborne, and P. Koehn, "Re-evaluating the role of BLEU in machine translation research," in EACL-06, 2006, pp. 249-256.

[17] S. Banerjee ande A. Lavie, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," in ACL'05, 2005, pp. 65-72.

[18] G. Doddington, "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics," in HLT-02, 2002, pp. 138-145.

[19] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, "A study of translation edit rate with targeted human annotation," in ASMTA, 2006.

[20] P. Koehn, Statistical Machine Translation. Cambridge: University Press, 2010.

[21] A. Lavie and A. Agarwal, "Meteor: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments," in ACL '07- WSMT, 2007, pp. 228-231.

[22] D. Varga, L. Németh, P. Halácsy, A. Kornai, V. Trón, and V. Nagy, "Parallel corpora for medium density languages," in Proc. RANLP 2005, 2005, pp. 590–596.

[23] M. Williams, "Translation Quality Assessment In Mutatis Mutandis. vol 2, no 1., 2009, pp. 3 – 23.

[24] D. Munková and M. Munk, "An automatic evaluation of machine translation and Slavic languages," in Proceedings of the 8th International Conference on Application of Information and Communication Technologies (AICT-14), Astana, 2014, pp. 447–451.

[25] D. Munková, M. Munk, and Ľ. Adamová, "Modelling of Language Processing Dependence on Morphological Features," in Advances in Intelligent Systems and Computing, 2013, pp. 77-86.

[26] D. Munková, M. Munk, and M. Vozár, "Data Pre-processing Evaluation for Text Mining: Transaction/Sequence model," in ICCS 2013, 2013, pp. 1198-1207.

[27] D. Munková, M. Munk, and M. Vozár, "Influence of Stop-Words Removal on Sequence Patterns Identification within Comparable Corpora," in Advances in Intelligent Systems and Computing, 2013, pp. 67-76.

[28] D. Munková and M. Munk, "Automatic Metrics for Machine Translation Evaluation and Minority Languages," in Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015, Volume 381 of the series Lecture Notes in Electrical Engineering pp 631-636

2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)

Math modeling of passenger traffic in the monorail transport system

Antonova V.M., Kuznetsov N.A.
The Kotel'nikov Institute of Radio-Engineering and Electronics of Russian Academy of Sciences
Moscow, The Russian Federation

Volkov D.O.
"GMK Innovations" LLC
Moscow, The Russian Federation

Starostenko A. M.
Control Systems Laboratory
Moscow Institute of Physics and Technologies
Moscow, The Russian Federation

***Abstract*— The paper is devoted to simulation of passenger transportation of monorails. Probability distribution model of passenger traffic is considered, on the basis of which the input and output passenger traffic in the stations is generated and used for simulation of passenger transportation. As a result, we obtain the estimates of parameters of transport system for determination of the necessary transportation capacity.**

***Index Terms*— probabilistic model, Gaussian distribution, simulation of passenger transportation, public transport**

I. Introduction

Currently, there is a serious imbalance of supply and demand in the transport sector in the most developed countries of the world: the volume of traffic, particularly of large cities, approaches the limits of capacity of existing highways and at rush hour even exceeds them. There are two ways to meet the traffic demand: extensive and intensive. For a long time the extensive approach aimed at expanding the existing infrastructure has been the primary way to solve the problem. However, currently this option is considered to be unviable [1], and high-tech approaches involving construction of new types of transport systems are becoming more and more popular. One of such types of transport is overhead monorail, which does not require the allocation of "scarce" areas and fits perfectly into the existing infrastructure.

To provide comfortable services for passengers, the necessary capacity and capability of transport system is to be known before the design stage. That is why we need to simulate and evaluate the distribution of passenger traffic and parameters of passenger transportation. However, although the simulation of passenger transportation in cities is carried out around the world for more than fifty years, and great theoretical and practical material was obtained in this area, the creation of timetables is a challenge and it should be designed for each case individually.

Thus, there is a very time-consuming task of obtaining statistical data on passenger traffic changes on the entire length of the monorail line. The solution to this problem by continuous examination is practically impossible because of considerable complexity and hence high cost of research. It is therefore necessary to build a simulation model that adequately describes the processes and allows to obtain the necessary statistical material [2] and create an optimal schedule before the start of the system with a minimum labor intensive.

II. Problem statement

The paper is devoted to simulation of passenger traffic and passenger transportation management of monorail transport system. The solution of these problems is necessary to create a supervisory control system.

The goal of management is to minimize the waiting time of passengers at the station. We assume that the transport system is controlled by changing the system parameters, such as the number of wagons in the train and the time interval of motion.

III. The simulation of passenger transportation

A. The simulation of incoming and outgoing passenger traffic

We consider the incoming traffic flow distribution as a two-humped distribution by analogy with passenger transportation of public transport systems, especially metro. According to the study of the Moscow subway, there are two peaks of passenger flow: from 8 to 9 a.m. and from 5:30 to 6:30 p.m., as shown in Figure 1 [3].

Firstly, during simulation we generate the incoming passenger flow as a mathematical model that uses two-humped distribution obtained by combining three normal distributions with peaks in the morning, afternoon and evening [4].

The afternoon distribution is added in order to avoid failure to 0 during the period from 11:00 a.m. to 2:00 p.m. The morning and evening peaks are high and apparent, as the intensity of the traffic flow is very high in these hours. The afternoon peak is smaller and wider, it is due to the decrease in passenger traffic intensity.

At the first part, we determine the input parameters for our model:

1. The number of stations s = [1,2…10]
2. The total number of passengers from 5 to 1 a.m. – Q(s)
3. Share of passengers transported in the morning peak-$a_1(s)$
4. Share of passengers transported in the afternoon peak–$a_2(s)$
5. Share of passengers transported in the evening peak–$a_3(s)$
6. The dispersion of the morning peak – $d_1(s)$
7. The dispersion of the afternoon peak – $d_2(s)$
8. The dispersion of the evening peak – $d_3(s)$

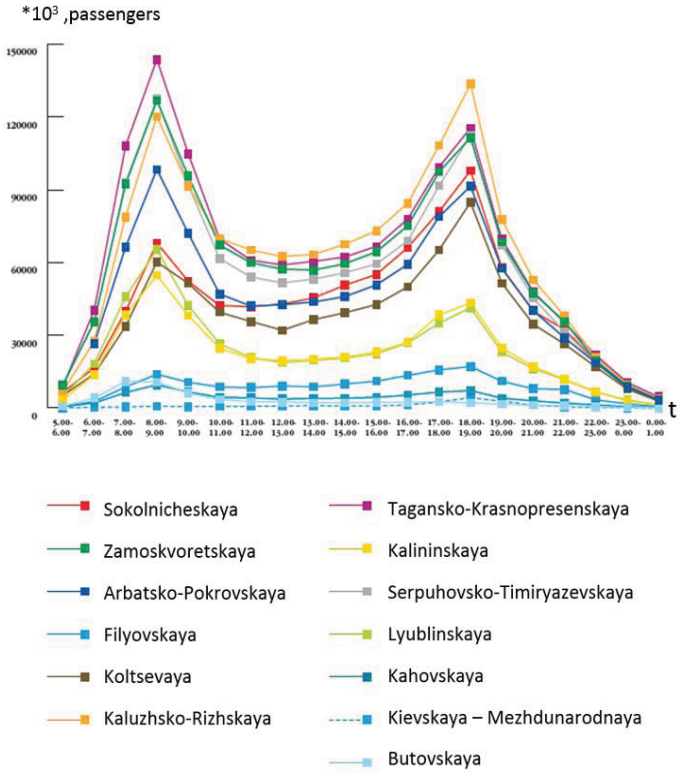


Fig. 1. *The hourly distribution of passengers on a working day in 2010.*

9. Table of stations importance - coefficients that describes how many people go out on special station.

For example, for passengers at the "Station 1", the table of coefficients can be as follows (the total number of stations is three):

	Station 2	Station 3
Morning	1/3	2/3
Afternoon	2/3	1/3
Evening	1/3	2/3

10. Train timetable

11. Travel time between stations

At the second part, main algorithm is implemented. It includes generating of input passengers, passenger transportation and outgoing passengers. The algorithm consist of two cycles – the first cycle is by stations, we iterate it from first to last in the direction of the train movement. Second, which is included in the first - by people entering the station. In the second cycle, we generate incoming and outgoing passenger traffic, simulate transportation of passengers.

Let us consider all steps of the algorithm.

Start cycle 1 - For every station s.

Start cycle 2 - For every incoming passenger.

Step 1. All Incoming passengers are generated in series by one person. Entry time is calculated according to the algorithm below on step 3. Now at this step we calculate the exit station for passenger, applying the uniform distribution, which is given by the formula distribution density, as in Eq 1. The length distribution [0, 1] is divided into parts, which lengths are determined according to the table of stations importance, and depending on where the random variable falls, we make a decision about exit the station.

$$f_X(x) = \begin{cases} 1, x \in [0,1] \\ 0, x \notin [0,1] \end{cases} \quad (1)$$

Step 2. As on step 1, we apply Equation 1 to select one of three normal distributions, which part of passengers enter the first, second or third peak. The length distribution [0, 1] is split into parts, in accordance with a_1 (s), a_2 (s), a_3 (s). Depending on part, where the random variable falls, we make a decision about which normal distribution for the passenger we should use.

Step 3. We use the normal distribution to computing the passenger's arrival time at the station - $t_{instation}$. Equation 2 shows density distribution of probability.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

$$\sigma, \mu - set\ the\ initial\ conditions$$

Step 4. According to the train timetable, we determine in which train passengers sit. If the train is packed, we will send him to the next train. So we calculate the boarding time - $t_{intrain}$.

Step 5. We count the waiting time of passengers with Equation 3.

$$t_{wait} = t_{intrain} - t_{instation} \quad (3)$$

We compare it with waiting time of the previous passenger, and get the maximum waiting time - t_{wait}^{max}.

Step 6. We add a time of train movement to the necessary station, which is given by the initial conditions, and obtain the exit time - t_{out}.

Go to the next passenger.

End cycle 1.

Go to the next station.

End cycle 2.

As a result, we get for each passenger set of characteristics – the arrival time at the station, the waiting time of the train, exit time. As well as the maximum waiting time of passengers.

B. Modelling of transportation of passengers transport system

Monorail transport system has the ability to change the transport capacity by two parameters - the number of wagons (N) in the train and the interval of movement (T). We introduce the following limitations: the number of wagons may vary from 1 to 4, and the minimum interval of movement T_{min} = 1 minute. One wagon accommodates 65 people. Defined by the total time-distance trains between stations by the time, it starts and goes to a full stop. These parameters are as close as possible to the transport system, being developed in Russia [5].

The program of modeling these parameters is set in the form of timetables:

Hour (t)	Interval of movement (T)	N wagons (N)
5:00 a.m.	7 min.	1
6:00 a.m.	5 min.	2
7:00 a.m.	3 min.	3

...

The criterion of success of the functioning of the transport system is the average maximum waiting time of passengers. According to the requirements, its value must not exceed 15 minutes, x <15 minutes. Figure 2 shows the Block schema of the algorithm of supervisory level of the transport system management.

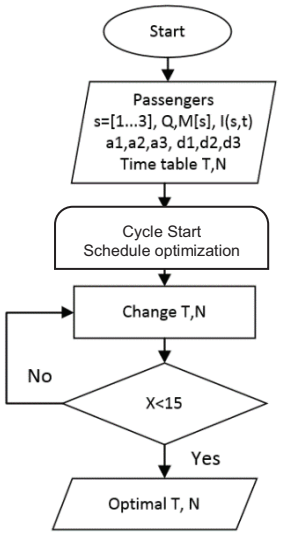


Fig. 2. *Block schema of the algorithm of supervisory level of the transport system management*

Every iteration changes the movement interval T and number of wagons N in the train so that the mean maximum waiting time of passengers at the stations does not exceed 15 minutes.

We carry out simulations with the following input data:

1. There are three stations, s = 3;
2. Peak in the morning is at 8:30 am, Evening - 6:00 p.m., Afternoon peak at 1:00 p.m.;
3. 15000 passengers enter in the Station 1, in the Station 2 – 10000, at the Station 3 – 15000. Q (1) = 15000, Q (2) = 10000, Q (3) = 15000;
4. Morning share of the passengers $a_1(1) = a_1(2) = a_1(3) = 0{,}32$, afternoon $a_2(1) = a_2(2) = a_2(3) = 0{,}26$, evening $a_3(1) = a_3(2) = a_3(3) = 0{,}42$;
5. Morning dispersion $d_1(1) = d_1(2) = d_1(3) = 1{,}5$, afternoon dispersion $d_2(1) = d_2(2) = d_2(3) = 5$, evening dispersion $d_3(1) = d_3(2) = d_3(3) = 2{,}5$;
6. M(1) =

	Station 2	Station 3
Morning	1/3	2/3
Afternoon	2/3	1/3
Evening	1/2	1/2

M(2) =

	Station 1	Station 3
Morning	1/2	1/2
Afternoon	1/2	1/2
Evening	1/2	1/2

M(3) =

	Station 1	Station 2
Morning	1/2	1/2
Afternoon	1/3	2/3
Evening	2/3	1/3

Input passenger traffic at the station 1 is shown on Figure 3, I(1, t).

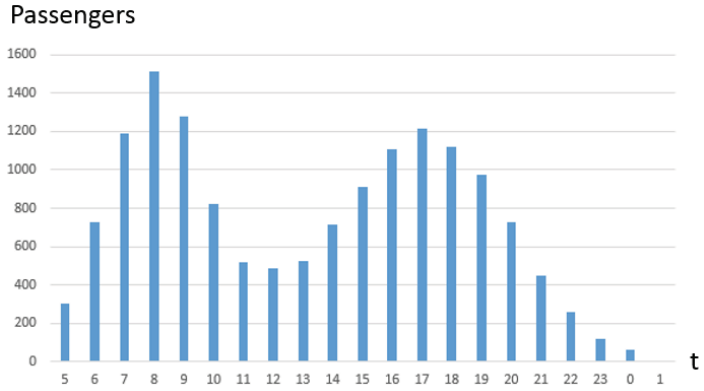


Fig. 3. *Input passenger traffic at the station 1, passengers/ hour*

Let us make the simulation with the schedule specified in the table, as shown in Figure 4:

Hour	Interval of movement	Number of wagons
5:00-6:00 a.m.	7,5 min.	1
6:00-7:00 a.m.	7,5 min.	2
7:00-8:00 a.m.	7,5 min.	2
8:00-9:15 a.m.	7,5 min.	2
9:15-10:15 a.m.	7,5 min.	2
10:15 a.m. - 3:30 p.m.	7,5 min.	2
3:30-7:45 p.m.	7,5 min.	2
7:45-9:30 p.m.	7,5 min.	2
9:30 p.m. - 1:00 a.m.	7,5 min.	1

Fig. 4. *Initial timetable*

The average maximum waiting time of passengers is shown in Figure 5.As seen from the graph, the standby time reaches nearly 5 hours, which is unacceptable based on the system requirements. It is necessary to reduce this time to 15 minutes or less.

After a few iterations, we get that for a given passenger traffic optimal timetable will be as shown in Figure 6.

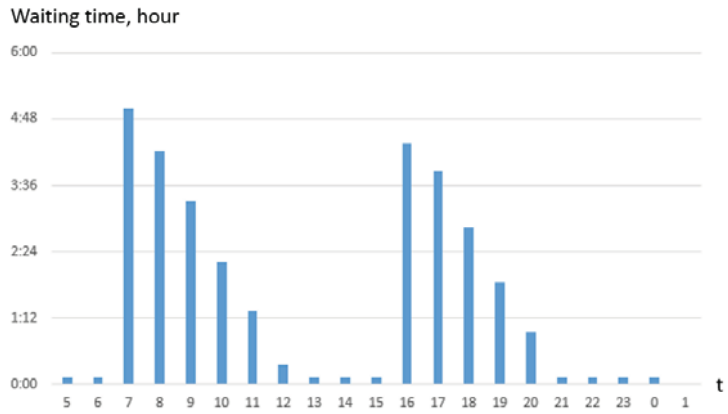


Fig. 5. *Waiting time of passengers, Hour/ daytime*

Hour	Interval of movement	Number of wagons
5:00-6:00 a.m.	7,5 min.	1
6:00-7:00 a.m.	4,5 min.	2
7:00-8:00 a.m.	3 min.	3
8:00-9:15 a.m.	3 min.	3
9:15-10:15 a.m.	3 min.	2
10:15 a.m. - 3:30 p.m.	7,5 min.	2
3:30-7:45 p.m.	3 min.	2
7:45-9:30 p.m.	3 min.	2
9:30 p.m.-1:00 a.m.	7,5 min.	1

Fig. 6. *Optimal timetable*

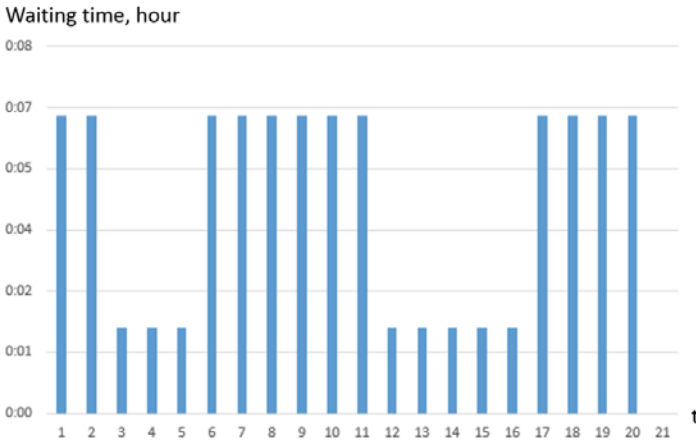


Fig. 7. *Waiting time of passengers, Hour/ daytime*

As Figure 7 shows, the average maximum waiting time is less than 7 minutes.

IV. Conclusion

In this work, we carried out a successful modeling of the passenger traffic in the transport system with two adjustable parameters – the interval of movement, and the number of wagons in the lineup. The developed model and the software is very flexible and allows you to make changes quickly in the studied transport system and to conduct simulations with any specified parameters: the number of stations, the number of incoming passengers, the number of wagons in the train headway etc. With the help of search engine optimization algorithms an optimal (for a given criterion) timetable for the transport system can be selected.

Acknowledgment

Research was supported by The Ministry of Education and Science of Russia, unique identifier - RFMEFI58214X0003.

References

[1] Inose H., Hamada, T. Control of traffic. // Moscow: Transport, 1983. — 248.

[2] L. Kleynrok. Queueing Systems. // M.: «Mechanical engineering», 1979. – 431.

[3] Analytical report Moscow metro 2010, URL:http://www.mosmetro.ru.

[4] Albert N. Shiryaev, «Probability», Vol 1. // M: MCNMO, 2007. – 928.

[5] Project «Design and testing of an experimantal elevated road and automatic traffic control system of above ground transport by H-Bahn technology», URL:http://igmk.ru/proekty/proekt-h-bahn/h-bahn-project

Behavior Analysis and Event Detection with Statistics of News Release on Insurgent Website

Min Tang, Duoyong Sun, Li Bo, Zihan Lin
College of Information System and Management
National University of Defense Technology
Changsha, China
tangmin3756@126.com

***Abstract*—Behavior analysis and events detection of insurgent organization are important issues in intelligence information research and public security management. Due to the covertness of insurgent activities, the relationships between organization and events are hard to detect. Using open source data provides an effective way to obtain information of organizational activities. One important issue in this field is mining hidden frequency pattern with noisy or incomplete information. In this paper, we study the activities of insurgent organization and related event with news data mined from insurgent website. A covert behavior mining method is proposed to process the open source data and analyze news release behaviors. The method can mine frequency pattern of news releasing in the case of losing part of news information. Three methods are employed to detect the events based on the processed data. The performance of the proposed method is evaluated through real data. With the results, we show how this method can be used to improve the events detection and surveillance of insurgent organizational activities.**

***Keywords*—web mining, behavior analysis, event detection, abnormal recognition.**

I. INTRODUCTION

Nowadays, with the development of Internet and information technology, the insurgent groups can easily acquire sensitive information and make a wide range of panic or public opinion [1]. Some researchers showed that insurgents organization use the Internet (website, social network, instant messaging software) as a platform to broadcast related information or inflammatory speech [2][3]. This is an ideal way, because it can spread widely with small probability of being detected. Another reason for the news website usage is that the news can be deleted momentarily, which increases the difficulty of behaviors detection. The Internet is becoming a convenient tool for the insurgent groups, and also many researchers have focused on the web mining for public security.

The researches of security have become critically important issue in recent years. It has been proved that the information technology played a significant role in public security management [4]. To identify and analyze attributes of news web page, different insurgent/extremist website with genuine sites are compared, the prepare metrics is further used for identification [5]. By using the information (mainly text) in website and forums of extremist and insurgent groups, the Dark Web Forum Portal [6] is constructed, which consists of four types of functions, forum browsing, statistics analysis, multilingual translation and social network visualization [7]. Web content mining [8] need to gather accurate infromation from different sites in order to discover knowledge in huge data. It makes use of all kinds of data mining techniques to organize and extract information automatically from webpages. The automated coding [9] technology makes the web mining more intelligent, and the Global Terrorism Database [10] can be constructed. The link prediction method is also used for identifying hidden groups of insurgents or criminals in the security and counter-terrorism domains [11].

Statistical analysis on the data of news release can get a lot of useful conclusions. However, it may lose effectiveness when the news data is incomplete. The news data on insurgent website are always deleted after the events, which protect the website from being discovered by the public security department. It also increases the difficulty of identifying whether the website is related with insurgent events. Most of current researches are based on the text information in Internet, and few studies related to missed information. As the text has been deleted from the website, we can only analyze the frequency of activities, which is also useful in event detection and insurgent website identification. The most common methods used for abnormal recognition are Sliding Window Detection (SWD) [12] and Local Outliers Mining Algorithm (LOMA) [13][14]. Both are methods of statistical analysis of temporal sequence data, and the accuracy rely on the integrity of the data.

In this paper, we study the news release behavior on insurgency website for behavior analysis and event detection of insurgent's organization. In section 2, we describe the methods of statistical analysis and covert behavior mining used for news release data in this paper. The statistical analysis gives general understanding of the releasing behaviors on website, and the method of covert behavior mining provides the way to get the frequency of deleted news. In section 3, three methods, threshold detection, SWD and LOMA are used to detect the events based on the processed data. The results of statistical analysis and event detection are presented in section 4, as well as the comparative results with the real events. Conclusion and future work is put forward in the last section.

II. DATA COLLECTION AND PROCESSING

A. Statistical Analysis of the News Release Website

It can get a lot of information from a news records. For each news record, we extract four meta-data, which are the news title, the release time, article ID, and the news briefing. These four meta-data contains general attributes used for different analysis. The release time is the key meta-data used in this method, which provides the time information of the time series data about the events and organization behavior. The news title and the news briefing are used for event content mining, which is not the focus of this paper. The article ID has no practical consequence for the news itself, but it provides a way to obtain covert post behavior of the news on website. For each record $Record_i$, we use two attributes, $Rtime_i$ (release time) and $RartID_i$ (article ID).

The statistical results of the behavior are measurements of the organization evolution [15]. For behavior analysis on news website, the most straightforward feature is the number of the released news. There is an obvious difference between the insurgent related website and the general news website. The release frequency of insurgent website relates to the insurgent events. The implicit relationship is what we want to discover in this method. Records can be assigned to different sets S_j based on the time difference. We define the time interval of S_j is $[S_j(t_1), S_j(t_2)]$, then,

$$S_j = \{Record_i \mid Rtime_i \in [S_j(t_1), S_j(t_2)]\} \quad (1)$$

The number of records in time interval $[S_j(t_1), S_j(t_2)]$ is $|S_j|$, which is time series varable of the release behaviors. The time interval $[S_j(t_1), S_j(t_2)]$ can be defined as required.

B. Covert Behavior Mining

For covertness, some of the news on the website are deleted after the event, and this partly hides the release behavior of the organization. Although the deleted news text is impossible to obtain, but the release behavior frequency can be captured. In this section, we propose a method to mine the covert behavior frequency. As described previously, the frequency of the news every day is a time series variable. We regard the data as a stream data, and assume the time is the sampling point with the step of one day. Thus the time series data can be handled as a function of $f(t)$, which can be defined as:

$$f(t) = |S_j| = |\{Record_i \mid Rtime_i = S_j(t_1) = S_j(t_2) = t\}| \quad (2)$$

where the time t is the independent variable, and the frequency $f(t)$ of time t is the dependent variable. Assuming there are two adjacent news records, $RartID_1$ and $RartID_2$. The time of the records are $Rtime_1$ and $Rtime_2$ $(Rtime_1 = t_1, Rtime_2 = t_2)$, and the frequencies at the time points are respectively $f(t_1)$ and $f(t_2)$. If the article ID is continuous, then there is no covert release behavior between the two records, and if it is not, there are two possible situations.

Time continuous: In this situation, the time t_1 and t_2 are continuous, which means that the covert behavior occurs at time t_1 or t_2. Then the covert release behavior should be assigned to the sets at the two time points. Since the records are contiguous increasing, the covert frequency is calculated as:

$$CovertFre = RartID_2 - RartID_1 \quad (3)$$

The frequency assignment is carried out as the following equations:

$$CovertFre(t_1) = CovertFre \times \frac{f(t_1)}{f(t_1) + f(t_2)} \quad (4)$$

$$CovertFre(t_2) = CovertFre \times \frac{f(t_2)}{f(t_1) + f(t_2)} \quad (5)$$

And the frequency can be recalculated as:

$$f(t_1) = f(t_1) + CovertFre(t_1) \quad (6)$$

$$f(t_2) = f(t_2) + CovertFre(t_2) \quad (7)$$

The frequency assignment is actually recalculated based on the frequency of the two contiguous times.

Time discontinuous: Time discontinuous indicates that there are covert behaviors between time points t_1 and t_2. For assignment, there is a time set $T = \{t_i \mid t_1 < t_i < t_2\}$, the purpose of the covert behavior mining is to sign the $CovertFre$ to the time set. Here we use the linear interpolation method for processing. Assuming the function in interval $[t_1, t_2]$ is linear, and the function expression is as:

$$f(t) = k \times t + b \quad (8)$$

The variable k and b can be gotten with the two cells $(t_1, f(t_1))$ and $(t_2, f(t_2))$, then we compute the estimation of the convert frequency as:

$$E(CovertFre) = \sum_{t_i \in T} f(t_i) \quad (9)$$

If $E(CovertFre) \geq CovertFre$, the assignment results can bet calculated with the equations as:

$$\begin{cases} f(t_i) = k \times t_i + b' \\ \sum_{t_i \in T} f(t_i) = CovertFre \end{cases} \quad (10)$$

If $E(CovertFre) < CovertFre$, the assignment results can bet calculated following the equations as:

$$\begin{cases} f(t_i) = k \times t_i + b'' \\ \sum_{t_i (t_1 \leq t_i \leq t_2)} f(t_i) = f(t_1) + f(t_2) + CovertFre \end{cases} \quad (11)$$

Where $t_1 \leq t_i \leq t_2$, and Equations (10) and (11) means that $f(t_1)$ and $f(t_2)$ will be recalculated. The frequency assignment methods in (10) and (11) make the covert behavior frequency linear. It is useful for the trend analysis within a long time interval $[t_1, t_2]$, and also meaningful to find the covert behavior in short interval.

III. EVENTS DETECTION METHODS

Sliding window detection is a frequently used method of stream data processing, which uses a sliding window model to analyze the stream data instead of random sampling. The algorithm makes decisions only based on the latest data, rather than calculating with all the data observed so far or an single sample. The sliding window method is generally used for filtering, and in this paper is used for events detection.

Formally, at each time t, there arrives a new data, which is the frequency of the news release behavior $f(t)$. The data will be out of date at time $t+w$, and the w is the window size or called length. Sliding the window along the time series data, there will be $n-w+1$ subsequences (assuming n times), which is denoted as $G=\{g_1, g_2, \cdots, g_{n\text{-}w+1}\}$, and $g_i=(t_i, \cdots, t_{i+w-1})$. The most common method of using sliding window for events detection is sliding average, which can handle the influence of noise. The sliding average calculates the average of subsequence g_i as the result of time t_i, which reduces the influence of noise in data. Finally, we define the frequency threshold σ_1. And there is a threshold σ_2 for the results of sliding window processing, and there is also a clustering process for the continuous time. It means that if the time points detected are continuous, they should be treated as one event, which lasts for a time period.

IV. RESULTS ANALYSIS

A. Data Discription

In order to analyze the news release behavior of insurgent organization, we use a web crawler to harvest the entire news webpages of an sample website and extract the meta-data of the news to a structural form. The data contains 2020 news webpages of the website from 2009 to 2014. These news text still exist on the website. After processing with the covert behavior mining method, we get 7258 records, which is used for the following analysis.

B. Statistical Analysis

Based on the methods mentioned previously, we get the statistical results of the news release behaviors on the website by year. As shown in Figure 1, both the news mining results and the covert mining results are given. The results give an overall understanding of releasing behavior distribution, which profiles the overall trend of the organization activities.

Figure 2 shows the frequency distribution of news release behavior and the statistical result of the frequency by day. From Figure 2(a) we can find that the number of news released is small at most time. There are some unusual time points when the value of release behavior is much higher. The result shows the frequency of activities in a detailed way, and it is obvious that there are some outliers in the data stream. We get the statistical result of frequency as shown in Figure 2(b). The results show that the frequency of the release behavior follows approximately power-law distribution, which means that minority events hold the frequent activities of the organization.

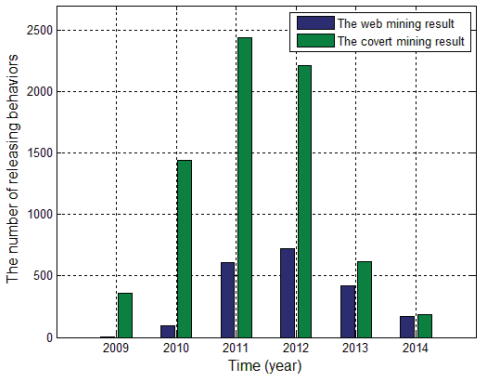


Fig. 1. The Statistical results of the released news

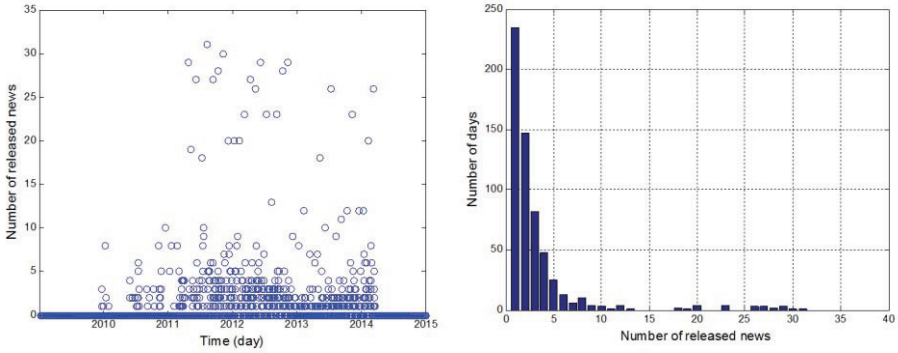
The results of the covert behavior mining are shown in Figure 3. From the disperse points, we can find that the covert behavior has similar distribution with the web mining, but the release frequency is much higher than Figure 2. The number difference can be found in Figure 1. The numerical difference indicats that there are many released news have been deleted. It is not difficult to understand that the website deleted the news from escaping the detection.

C. Events Detection

In this section, we give the results of events detection by the three methods presented in section 3. Table 1 shows the accuracy of different methods ($\sigma_1=\sigma_2=20$, $w=3$). The covert behavior mining method is effective for the SWD and the LOMA, which is because it provides additional informtion for behavior analysis. However, it reduces the accuracy of the threshold detection method, this is because the additional informaiton may contains noise which is sensitve in threshold detection method. Another interesting result is that the covert behavior mining method increases the number of events detected by the three methods, this indicates that the method is effective in mining the losing behavior information even if the news text in the websit has been deleted.

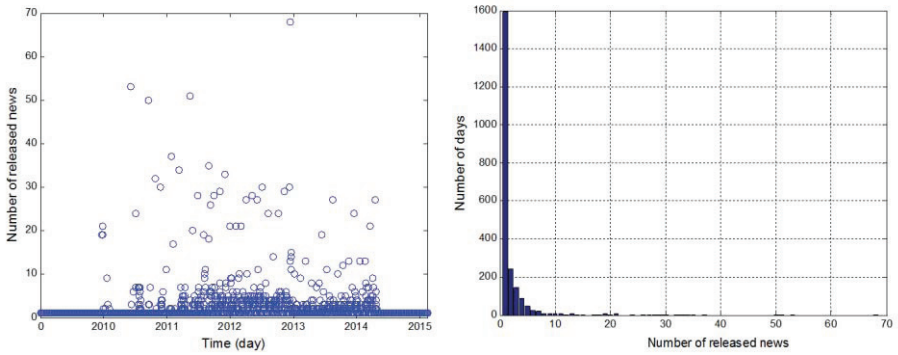
To further analyze the factors influencing the accuracy, we calculated the results of events detection with different thresholds. The results are shown in Table 2 (there is no threshold in LOMA, so we just focus on the threshold detection and SWD). The values in the table are the numbers of events detected by the methods, and the numbers of insurgent events detected are shown in the parentheses. When the threshold is small, more events can be detected, but the accuracy is relatively low. Conversely, the accuracy increases with the value of threshold, but the number of events detected getting fewer. Figure 4 shows the curves of accuracies with different thresholds. The curves shows volatility of accuracy with thresholds increasing. The accuracy of threshold detection method (web data) continue increasing with the value of threshold, which is different from the other three curves as they have peak points. The accuracy of SWD (web data) with low thresholds is higher than the other curves, which is because it can deal with the influence of noise. The SWD (covert data) is better than the threshold detection

(covert data) over the entire threshold range. Despite the high detection accuracy of threshold method (web data) with high threshold value, it can only detect few events. The results shown in Table 2 and Figure 4 indicate that the detection methods are effective in different situation, and the threshold should be set according to the methods and the characteristics of data.



(a)The frequency of news releasing with web mining (b)The frequency statistics of web mining

Fig. 2. *Statistical result of news releasing with web mining*



(a)The frequency of news releasing with covert behavior mining (b) The frequency statistics of covert behavior mining

Fig. 3. *Statistical result of news releasing with covert behavior mining*

TABLE I. EVENTS DETECTION ACCURACY OF DIFFERENT METHODS

Methods	Number of events	Insurgent events	No-isurgent events	Accuracy
Threshold (web data)	17	11	5	64.71%
SWD (web data)	15	10	5	66.67%
LOMA (web data)	16	12	4	75.00%
Threshold (covert data)	32	14	18	43.75%
SWD (covert data)	24	19	5	79.17%
LOMA (covert data)	33	27	6	81.82%

TABLE II. EVENTS DETECTION RESULTS WITH DIFFERENT THRESHOLDS

Threshold value	Web data mining		Covert behavior mining	
	Threshold Detction	SWD	Threshold Detction	SWD
5	91(31)	63(29)	166(41)	98(39)
10	33(15)	28(20)	75(28)	45(19)
15	24(14)	17(14)	42(17)	38(17)
20	17(11)	15(10)	32(14)	24(19)
25	13(10)	11(6)	22(17)	18(16)
30	2(2)	0(0)	12(9)	11(9)

Table 3 shows part of the events detected. The dates in the tables are the event time detected by the methods, which are consistent with the reality. Such as the date of "2012-2-29" is the day after the violate attack at kashgar region of Xinjiang province in China. The events are sorted by the frequency values in threshold method and SWD, and distance in LOMA, which can be seen as the severity of events.

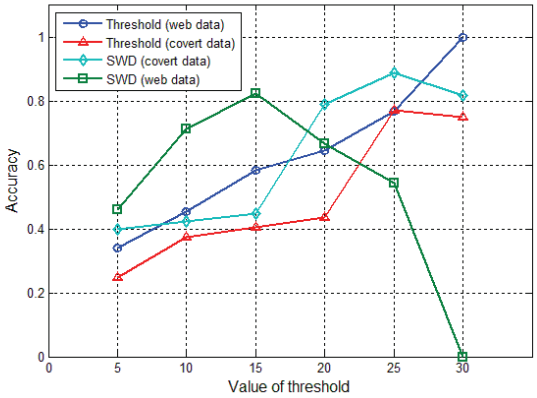


Fig. 4. *Accuracy of events detection with different thresholds*

TABLE III. EVENTS TIME DETECTED BY DIFFERENT METHODS

Events detection with web mining data			
No.	***Threshold***	***Sliding window***	***LOMA***
1	2011-8-1	2011-7-5	2011-7-5
2	2011-11-1	2011-7-20	2011-8-1
3	2012-2-29	2011-8-1	2011-11-1
4	2012-4-9	2011-11-1	2012-2-29
5	2012-6-26	2012-4-8	2012-7-5
Events detection with covert behavior mining data			
No.	***Threshold***	***Sliding window***	***LOMA***
1	2010-8-20	2010-1-20	2010-1-21
2	2010-11-2	2010-8-19	2010-12-20
3	2010-12-20	2011-3-14	2011-1-21
4	2011-1-20	2011-7-5	2011-3-16
5	2011-3-15	2011-11-2	2011-7-5

There are 15 events in Table 3 (after removing duplicate events), and 12 events are the insurgent attacks or the group incidents which are organized by the insurgent organization. The others are the broadcast activities of news or information ralate to the events. This shows that the method can provide additional information for insurgent website identification.

V. CONCLUSIONS

This paper has studied the news release behavior on website for insurgent events detection and organization behavior analysis. With the statistical results, we have found that there is obvious characteristic of frequency distribution between the organization and the events. Three methods, the threshold detection, the SWD, and the LOMA are used for events detection. The result shows that combining information mined from the website can increase accuracy for events detection and obtain more information about activities of insurgent organization. The future work includes the further analysis of the events with content mining method. Based on the results of content analysis, the events classification and early warning can be performed.

ACKNOWLEDGMENT

This work is supported by the National Nature Science Foundation of China under Grant No. 71973138, and the Specialized Research Fund for the Doctoral Program of Higher Education of China under the Grant No.20134307110020.

REFERENCES

[1] Bunker R.J. (2014). Networks, terrorism and global insurgency, Routledge, New York.

[2] Conway M. (2006). Terrorism and the Internet: New Media—New Threat. Parliamentary Affairs, 59(2), pp. 283-298.

[3] Seib P. and Janbek D.M. (2011). Global Terrorism and New Media, Routledge, New York.

[4] Bohannon J. (2009). Counterterrorism's New Tool: 'Metanetwork' Analysis, Science, 325(5939), pp. 409-411.

[5] Patil G.A., Manwade K.B., and Landge P.S. (2012). A Novel Approach for Social Network Analysis & Web Mining for Counter Terrorism, International Journal on Computer Science & Engineering, 4(11), pp. 1816-1825.

[6] Chen H. (2006). Intelligence and security informatics for international security: Information sharing and data mining, Springer, Berlin.

[7] Chen H. (2010). AI and security informatics, Intelligent Systems, 25(5), pp. 82-90.

[8] Ajoudanian S. and Jazi M.D. (2009). Deep Web Content Mining, Proceedings of World Academy of Science: Engineering & Technology, 49(8), pp. 1-10.

[9] Schrodt P.A. and Van B.D. (2013). Automated coding of political event data, Handbook of Computational Approaches to Counterterrorism. Springer, New York, pp. 23-49.

[10] LaFree G. and Dugan L. (2007). Introducing the global terrorism database, Terrorism and Political Violence, 19(2), pp. 181-204.

[11] Fire M., Puzis R., and Elovici Y. (2013). Link prediction in highly fractional data sets, Handbook of Computational Approaches to Counterterrorism. Springer, New York, pp. 283-300.

[12] Fu T. (2011). A review on time series data mining. Engineering Applications of Artificial Intelligence, 24(1), pp. 164-181.

[13] Breuning M.M., Kriegel H.P., Ng R.T., et al., (2000). LOF: Identifying density-based local outliers, Proceedings of the ACM SIGMOD International Conference on Management of Data, Dallas, pp. 93-104.

[14] He Z., Xu X., Deng S. (2003). Discovering cluster-based local outliers, Pattern Recognition Letters, 24(9), pp. 1641-1650.

[15] Clauset A. and Gleditsch K.S. (2012). The developmental dynamics of insurgent organizations, PloS one, 7(11), e48633.

Road Condition Analysis Using 3-axis Accelerometer And GPS Sensors

B. Y. Amirgaliyev, K. K. Kuatov, Z. Y. Baibatyr
Department of Information Systems
International Information Technologies University
Almaty, Kazakhstan
b.amirgaliyev@iitu.kz

***Abstract*—This work proposes road quality data analysis based on accelerometer and GPS sensors built into smartphone. Given paper particularly addresses the problem of road conditions. In order to study on road patterns we collected sufficient amount of data during the driving on different types of roads. In this paper, we provide visual representation of differences of consecutive accelerometer values. All examined cases consider speed of the moving vehicle.**

***Keyowrds*—road quality, GPS, accelerometer, road data analysis.**

I. INTRODUCTION

Nowadays various piece of data can be used to acquire necessary information. Particularly we refer to processing of road data collected from 3-axis acceleration sensor and GPS (Global Positioning System). This technical data provides important clues which serve to understand different patterns occurred during the driving. As a goal we set to study distinct behaviors of accelerometer records with respect to road conditions and some particular driving events. Given research facilitates road quality identification. As a matter of fact, the problem of quality of roads still remains open. The maintenance of roadways calls for regular monitoring of road degrees, which can be classified as smooth and non-smooth. Smooth roads are divided into several types such as perfectly flat surface, smooth but rough and smooth road with some anomalies. These anomalies are damaged parts like pits and potholes. Some physical damages can be caused because of the weather conditions. Non-smooth is also defective type of the road which comprises uneven surface, poorly paved and non-paved roads. The latter one we also name as off-roads. Most of these road issues have their impacts on state of the vehicles. Especially they deteriorate the chassis (vehicle frame) of an automobile. It turns out that manholes and expansion joints belong to road anomalies as well. They can also result in defects of the car. To prevent something like this, drivers need to be notified about bad road or abnormality so they can bypass them. In order to reconstruct damaged part of the road in a short period of time maintenance services must be informed of them instantly. Such awareness can be achieved through the system which will automatically analyze road condition. To offer the solution one need to understand specificity of different patterns on the road. Precisely we only need to explore 3-axis accelerometer values and GPS coordinates provided by sensors built into a device which is attached inside the vehicle. So the idea is that acceleration changes can give distinguishable data in various situations. Using GPS coordinates we store traveling position and calculate average speed at a certain time. Regardless of the practical information we still face some challenging patterns and noisy data.

II. RELATED WORK / PREVIOUS STUDIES

There are some works that introduce own methods of measuring road quality. Some approaches use smartphone sensors for collecting data. In [1] Kalra et al uses smartphone sensors to analyze driving events and road anomalies. They used accelerometer of smartphone placed onto dashboard to collect road and driving data. These data were observed and analyzed by determining thresholds for different patterns. Eriksson et al [2] proposed a system called Pothole Patrol which uses tree-axis accelerometer and GPS sensors for road data gathering. This system had been deployed on 7 taxis driving in Boston. The system is able to detect road anomalies and distinguish them from other unnecessary driving events. Pothole Patrol system with false positive reduction approach was able to identify potholes in need of repair with 90% precision. This system can also determine the location of found potholes on the map. It is done by analyzing GPS coordinates of the detected road anomalies. They also propose several filtering steps which discard events such as turning, braking, sensor orientation change, door slams and slow speed motion of a vehicle. In [3] Mednis et al propose a mobile sensing system for road anomaly detection using smartphones based on Android OS. They provide detailed comparison of four methods Z-THRESH, Z-DIFF, STDDEV(Z) and G-ZERO on gathered accelerometer data. The whole system gives a true positive rate of 90% on their test data. They also claim that road quality annotation has many subtleties [4]. Bhoraskar et al [5] propose an approach using accelerometer, GPS and magnetometer sensors of smartphones to detect road conditions. Their approach allows to use smartphones in any pose with help of two step reorientation method using accelerometer and magnetometer data. Support Vector Machine was used to train classifiers on road quality data for event detection. Their approach is able to identify driving events like brakes and bumps. For bump detection it demonstrates a zero false positive rate, and a 10% false negative rate. In previous paper [6], a design of system which uses a

device with gyroscope and GPS sensors for road quality analysis was presented.

III. Data acquisition

All gathered data were recorded while driving on Almaty roads in Kazakhstan. Collection of road quality data was the first step of conducted studies. This step is the most important part of the analysis because the next steps of this work rely on correctness and accuracy of the data. In this study, data acquisition was carried out by using a smartphone. Almost every smartphone has multiple types of sensors including accelerometer and gyroscope. These sensors provide information about orientation and motion of a smartphone. After consideration of the related works and investigation of those sensors accelerometer was chosen for acquisition of road quality data. Accelerometers provide data about acceleration of the device in the three-dimensional space. The data from accelerometer is given as X, Y and Z accelerations. X axes shows direction parallel to the width side of a device while Y axes shows direction parallel to its height side and Z axes shows direction perpendicular to its display. A smartphone with BOSCH accelerometer sensor was used to collect acceleration data. In this work, the necessity of speed data along with acceleration is considered. Speed data can be obtained using different sensors. The most suitable approach to extract velocity data of a vehicle is to use GPS sensor. GPS gives coordinates of the device on the earth which can be used to find passed distance and speed.

A mobile application was developed in order to gather road quality data. This application receives sensor indications from the device and saves them in a text file. Average frequency of data writings is 100 lines per second. Each line of this text file contains information about longitude and latitude coordinates from GPS, X, Y and Z accelerations from accelerometer, current time, road quality indicator and velocity. These data are used in the next preprocessing step to prepare them for the further analysis. Preprocessing step is required to correct or improve data annotation. Data annotation is firstly handled by a passenger sitting next to driver. This passenger labels all the road events on the smartphone using interface of the developed application. The interface of the application plots the graph of acceleration change and also displays current velocity and several buttons for annotating road conditions. Road conditions include smooth road, off-road, speed bump, manhole and pothole classes. In order to perceive roughness of the road from all sides of a vehicle, smartphone was set on the center of the lower part of dashboard. It is also obligatory that button press event does not affect accelerometer magnitudes.

One of the important aspects of the data acquisition step is full control of annotation process. Since most roads have intervals with different conditions it is hard to make accurate annotation. In some cases, it is even difficult to distinguish between two types of roughness. Therefore, road intervals with particular types of roughness were chosen in advance. This solution ensures the correctness and accuracy of preliminary annotation by a passenger. Each annotated road interval was covered several times with different vehicle speeds. This approach gives us a better understanding of vehicle behavior and accelerometer sensitivity to the roughness of the road at different speeds. At the same time dependency between vehicle speed and accelerometer values can be inferred. Another advantage of this approach is that the same event is repeated multiple times. It leads to more comprehensive analysis of all types of events that took place during annotation process.

IV. Preprocessing

After collecting road quality data using sensors equipped in vehicle it must be reviewed in order to adjust annotated labels. Event annotation cannot be accurate when it is done during the driving. Start and end of time interval of the road events drift frequently. These drifts are corrected manually by second annotation phase. During this annotation phase all the unlogged or ambiguous events are discarded from the collected data.

In this work, road quality data analysis takes into consideration velocity of the vehicle. Speed can be calculated using coordinates from GPS and time intervals passed between those points. It's not difficult to find approximate elapsed distance using GPS coordinates and Earth radius but GPS coordinates usually are not accurate enough and positioning can work slowly, particularly at low speeds and bad conditions for GPS. This causes small errors in road state analysis. To decrease impact of these errors data intervals with vehicle speed lower than 10 km/h were discarded. During the data acquisition step real values of vehicle speed were recorded in a log. Comparison of calculated and real velocity data showed that calculated speed values are almost always lower than real values but it is acceptable to use the former one because errors follow the same fashion.

V. Road Data Analysis

In our study, we utilize multi-axis model of accelerometer along with GPS sensor to obtain and illustrate some signal patterns. More particularly our objective is to demonstrate how X, Y, Z acceleration forces respond to some road conditions during the driving. As we mentioned earlier speed of the moving vehicle has an effect on acceleration data. Here speed serves as a normalization factor. So it was critical to get the velocity information out. For each acceleration values we calculated speed using latitude and longitude coordinates.

First of all, let's list deduced road types (Table 1) and then depict corresponding acceleration values on graphs.

TABLE I. ROAD TYPES

Smooth	Non-smooth
perfectly flat surface; smooth but rough; smooth road with some anomalies (pits, potholes, speed bumps, expansion joints and etc.)	uneven surface; off-road levels (poorly paved and non-paved roads).

In order to derive patterns, we decided to use not the actual acceleration magnitudes but the difference of consecutive

values for each direction. Let's denote these values as deltaA = { deltaAX, deltaAY, deltaAZ }. Thus we eliminate G-force to handle situations when car changes its orientation in space. In this case, mean value of each deltaA tends to zero.

The graphs of deltaA magnitudes for 10 seconds window when car is not moving is illustrated below (Fig. 1). deltaAZ varies between -0.8 and 0.5. deltaAZ values has grater deviation than deltaAX and deltaAY. Here we observe how car engine effects acceleration changes. For the most part it only influence deltaAZ.

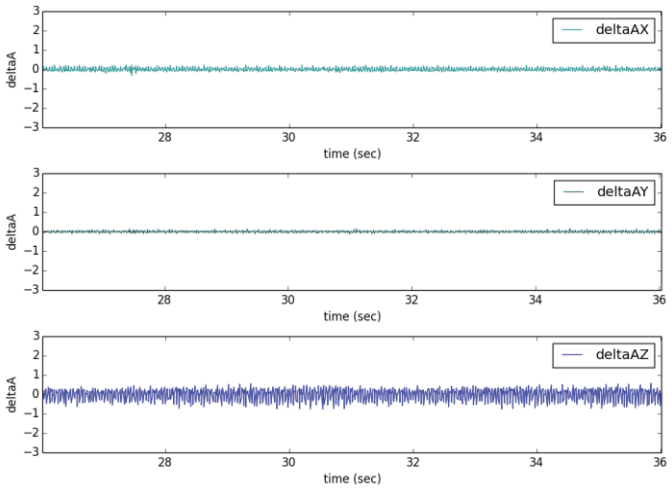


Figure 1. deltaA values when vehicle doesn't move

Graphs below illustrate data for 40 km/h and 60 km/h speed within 10 seconds (Fig. 2-3) for the same road interval. We observed that deltaAX and deltaAZ oscillates between -2.3 and 2.2 on perfectly flat road while deltaAY on average deviates by 0.5 from 0.

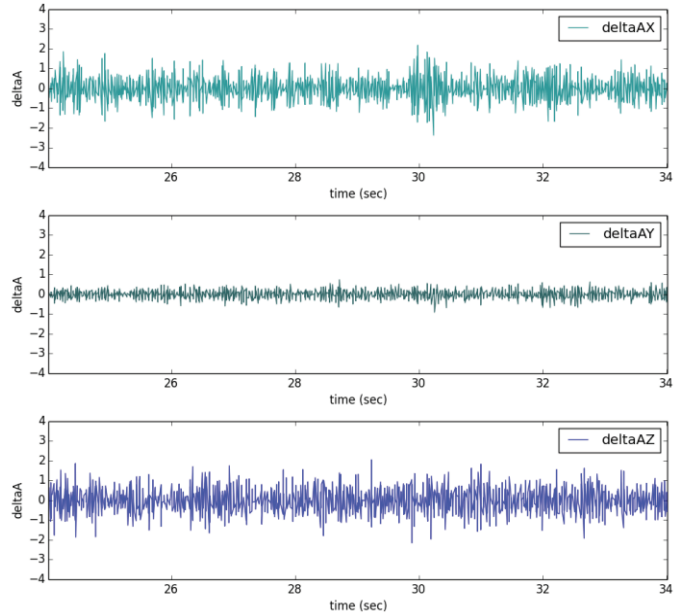


Figure 2. deltaA values when vehicle moves with 40 km/h speed on perfectly flat surface

Graph in Fig. 3 presented for 60 km/h velocity is similar in its range to the graph for deltaA values with 40 km/h. But deviation of amplitude slightly increased. We concluded that only road anomalies can display markedly deviating values with higher speed on perfectly flat road.

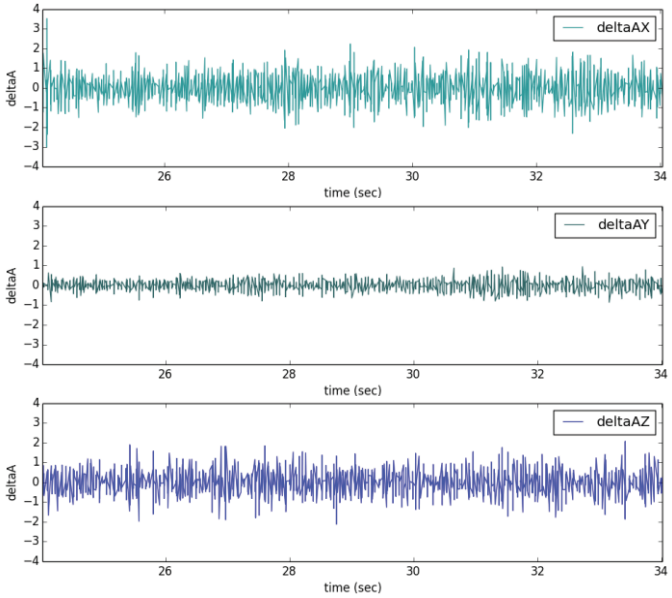


Figure 3. deltaA values when vehicle moves with 60 km/h speed on perfectly flat surface

Fig. 4 illustrates deltaA magnitudes for 80 km/h velocity. As you can observe graph for deltaAX and deltaAY didn't change that much in its deviation range from the previous graph. But deltaAZ increased its amplitude up to 3.2 for 80 km/h on perfectly flat road. We inferred that for a fixed velocity on perfectly flat surface amplitudes of deltaA values must be below some threshold. This border value always depends on the moving speed.

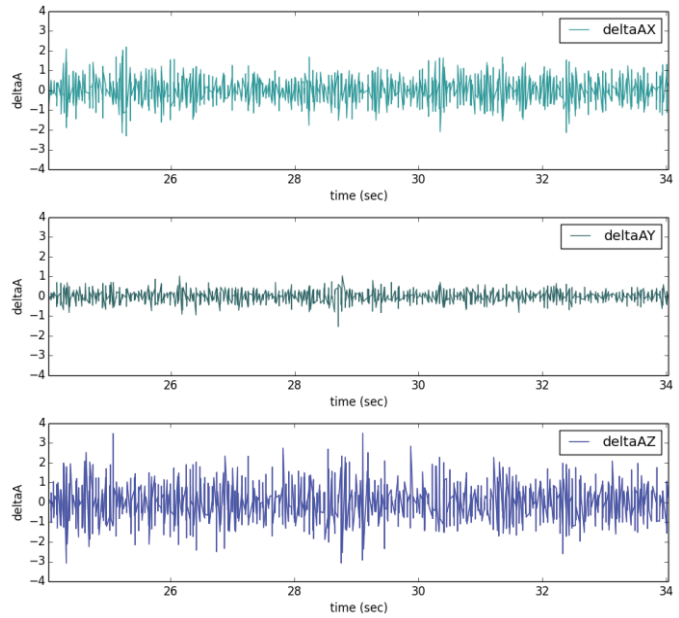


Figure 4. deltaA values when vehicle moves with 80 km/h speed on perfectly flat surface

To understand road anomalies, we purposely hit manholes, pits and potholes. These anomalies are well distinguishable graph patterns on smooth road. In Fig. 5, we demonstrate deltaA records for 40 km/h with two noticeable manholes. We assume that deltaA magnitudes for road anomaly is usually above standard deviation of the smooth road.

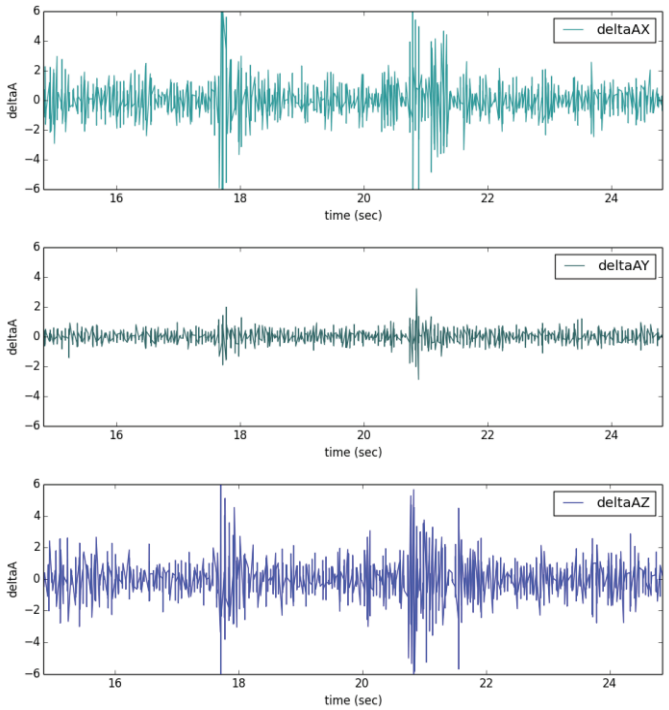


Figure 5. deltaA values when vehicle moves with 40 km/h speed on smooth but rough road with two manholes

In addition, illustrations of expansion joint and two potholes anomalies through deltaAZ data are depicted in Fig. 6-7. When vehicle moves at the high speed we see that deltaAZ responds with high value for the road anomaly. Speed information is necessary to normalize deltaA values and identify road anomalies.

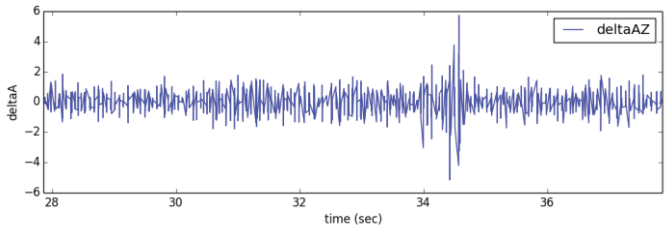


Figure 6. deltaAZ records when vehicle moves with 80 km/h speed on perfectly flat surface with expansion joint

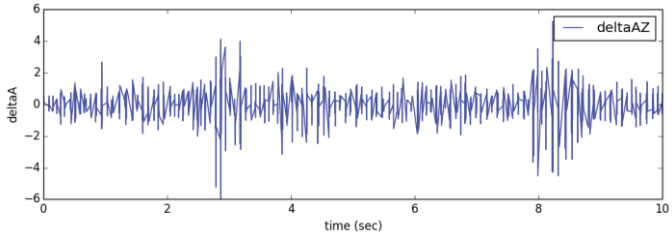


Figure 7. deltaAZ records when vehicle moves with 40 km/h speed on smooth but rough road with two potholes

When car drives on non-smooth, bad road deltaAX magnitudes start varying in a big range even if the speed is low enough. It indicates that off-road can be identified through the observation of deltaAX amplitude and speed. It is anomalous behaviour when speed is low. So we have another pattern for off-road in Fig. 8.

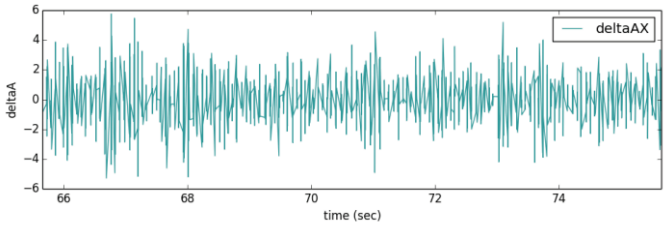


Figure 8. deltaAX records when vehicle moves with 10-20 km/h speed on non-paved road

Having analyzed all the collected road data, two types of anomaly detection features were extracted and compared. To find these features we used 1 second sliding interval over all data. The first feature is found by calculating the ratio of maximum deltaAZ to corresponding mean speed for each interval. In Fig. 9 smooth road and non-smooth road features are illustrated as green and red points accordingly. These features are displayed in two rows in order to avoid confusion. One can see that most values of the first feature values for smooth road doesn't exceed threshold value 0.2. On the other hand, features of non-smooth road are above this threshold but some ratios are below. This is because of some annotation imperfections at the borders of non-smooth intervals.

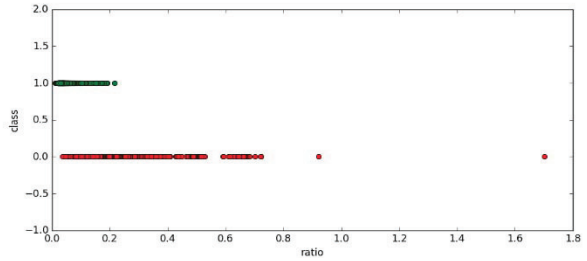


Figure 9. Visualization of the first features for smooth and non-smooth road data

The second examined feature is the ratio of the sum of maximum deltaAX and maximum deltaAZ to corresponding mean speed for each interval. The second feature values are depicted in Fig. 10. It can be seen that there is a little difference between using two feature types.

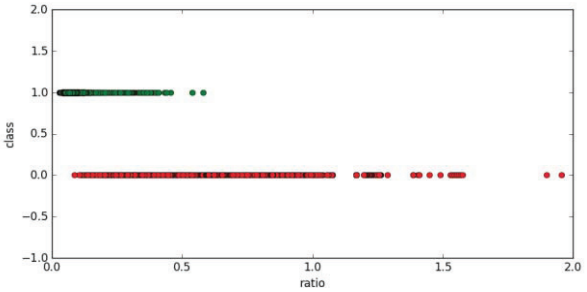


Figure 10. Visualization of the second features for smooth and non-smooth road data

Therefore, we claim that using deltaAX is appropriate for road anomaly detection as well. Particularly off-road intervals can be easily detected by using deltaAX and deltaAZ values.

VI. CONCLUSION

Given paper provides road data analysis based on accelerometer and GPS records. For this work road quality data was gathered using smartphone placed inside the car. Collected data was analyzed using manual annotation and graphical visualization. Comparison of smooth and non-smooth roads was performed using two different road quality features. We proposed a method for distinguishing smooth type from non-smooth type of roads. It was inferred that vehicle speed has impact on accelerometer records. Proposed road quality features take into account vehicle speed factor. This work can be improved by adding machine learning algorithms and advanced signal processing methods.

ACKNOWLEDGMENT

The work was done under the funding of the Ministry of Education and Science of the Republic of Kazakhstan 2015-2017 Scientific Program (No. 4836/GF4).

REFERENCES

[1] N. Kalra, G. Chugh and D. Bansal, "Analyzing Driving and Road Events via Smartphone", International Journal of Computer Applications (0975 – 8887) Volume 98– No.12, July 2014.

[2] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring", Proceedings of the 6th international conference on Mobile systems, applications, and services, Breckenridge, Colorado, USA, June 17–20, 2008. pp. 29-39. ACM.

[3] A. Mednis, G. Strazdin, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time photole detection using android smartphone with accelerometers", In: International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)., (Barcelona), pp. 1–6, June 2011.

[4] G. Strazdins, A. Mednis, R. Zviedris, G. Kanonirs, and L. Selavo, "Virtual Ground Truth in Vehicular Sensing Experiments: How to Mark it Accurately," in The 5th International Conference on Sensor Technologies and Applications (SENSORCOMM 2011), (Nice, France), pp. 295–300, August 2011.

[5] R. Bhoraskar, N. Vankadhara, B. Raman, P. Kulkarni, "Wolverine: Traffic and Road Condition estimation using Smartphone Sensors", In: Fourth International Conference on Communication Systems and Networks (COMSNETS), (Bangalore), pp. 1–6, January 2012.

[6] M. Hasan, B. Amirgaliyev, D. Rakhatov, A. Tuyakbayev, C. Kenshimov, "Real time road mapping for the driving safety purposes", in 2015 International Conference on Connected Vehicles and Expo (ICCVE), (Shenzhen, China), pp. 60–61, October 2015.

Link Grammar Parser for Turkic Languages and algorithms for estimation the relevance of documents

Batura Tatiana Victorovna
A.P. Ershov Institute of Informatics Systems, SB RAS
Novosibirsk State University
Novosibirsk, Russia
tatiana.v.batura@gmail.com@mail.ru

Murzin Feodor Alexandrovich
A.P. Ershov Institute of Informatics Systems, SB RAS
Novosibirsk State University
Novosibirsk, Russia
murzin@iis.nsk.su@mail.ru

Sagnayeva Saule Kairolliyevna
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
sagnaeva_tar@mail.ru

Tazhibayeva Saule Zhaksylykbayevna
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
tazibaeva_szh@enu.kz

Yerimbetova Aigerim Sembekovna
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
aigerian@mail.ru

Bakiyeva Aigerim Muratovna
Novosibirsk State University
Novosibirsk, Russia
m_aigerim0707@mail.ru

***Abstract*—The main problem considered in this paper is creating algorithms for estimation the relevance of documents to the search query on the basis of sentences structure analysis. To decide this problem, we use the relations between words constructed by the program system Link Grammar Parser, based on the so-called link grammar. There were suggested the natural system of links for Turkic languages, created prototypes of Link Grammar Parser for Kazakh and Turkish. In addition, we considered some generalization of the summarization algorithm of Niraj Kumar. Results of research are planned to use in different information retrieval systems.**

***Index Terms*—Natural language processing, syntactic analysis, Link Grammar Parser, relevance, Turkic languages.**

I. Introduction

Due to the fact of the increasing volumes of information networks, the problem of improving the quality of the automatic information extraction becomes more and more topical. Many researchers [1, 2] introduce a deep semantic analysis of texts for making the semantic images of texts that can be the basis for document ranking. However, it requires a careful and long-term work on the creation of suitable tools for natural language processing [3]. Therefore, a search for partial solutions, one of which is presented in this paper, is also useful.

Our main purpose is creating algorithms for estimation the relevance of documents to the search query on the basis of sentences structure analysis. For this, we use the relations between words constructed by Link Grammar Parser [4, 5]. Simple algorithms for calculating the degree of similarity of sentences in a natural language were described by us in [6–8].

The Link Grammar Parser is a syntactic parser, based on link grammar [9]. In the current time, there exist variants for English, Russian, Persian, Arabic, German, Lithuanian, Vietnamese, Indonesian languages.

One of the unsolved problems is the development of a parser like Link Grammar Parser for Turkic languages, which are most frequently represented in Internet, such as Kazakh and Turkish.

We suggested the natural system of links for Turkic languages, created prototypes of Link Grammar Parser for Kazakh and Turkish. In addition, we considered some generalization of the summarization algorithm of Niraj Kumar.

II. Parsers for Turkic languages

Many morphologic and syntactic parsers are developed nowadays. In particular, some approaches applied to agglutinative languages are described in the works [10-13].

The machine translation system from Kazakh into English and vice versa, using the link grammar and statistical approach, is considered in the paper by U. A. Tukeyev et al. [10]. Link Grammar plays an important role in the algorithm they are proposed. The statistical approach is used for translation of polysemantic words. The developed models and algorithms have been implemented in the program of machine translation. According to the linguistic classification, there are six different types of languages: SVO — Subject Verb Object; SOV — Subject Object Verb; VSO — Verb Subject Object,

etc. These schemes reflect the typical structure of sentences. Turkic languages belong to the type SOV. A list of 13 links that naturally reflect the most important syntactic links between words in the sentences in the Kazakh language is described in [10]. It is important that the same links can be used in the development of parsers for other Turkic languages, due to the high degree of similarity not only of their syntax but also the morphology and vocabulary.

In [11], the "statistical parser" of dependencies of the Turkish language is described, which is based on the statistical models of learning based on the sentences in the Turkish language from the Turkish Dependency Treebank. As a result, the parser produces the dependency relationships between inflective groups — lexical units within the subsets of words in a sentence.

The research [12] shows that the morphological and lexical information can improve parsing accuracy substantially. The proposed IG-based (inflectional group) models consistently outperform word-based models. This result has been obtained both for the probabilistic and the classifier-based parser, although the probabilistic parser requires careful manual selection of relevant features to counter the effect of data sparseness. A similar result was obtained in respect of lexicalized authors, in this case, although the improvement was only demonstrated in the classifier, which is probably due to its greater resistance was the scarcity of data based on the parsing. By combining a deterministic classifier based parsing approach with an adequate use of the model IG (inflectional group) on the basis of representations of morphological information and lexicalization, the authors concluded that they managed to achieve the highest accuracy for parsing the Turkish Treebank.

That is, in contrast to the system of Link Grammar Parser which uses a dictionary containing the specifications that describe the relationship, in this case the link grammar is derived from the statistics.

The Turkish link parser considered in [13] is "not a lexical analyzer" in fact. At the first stage, a morphological analyzer is applied and some morphological descriptions are compared to the initial words. These descriptions are based on the analysis of the suffixes of words, which is natural for agglutinative languages. There are lexical items of only certain functionally important words. Then the links are established between morphological descriptions, not between the initial words. Apparently, it is possible to return to the initial sentence and carry the derived links to the words, but it is not considered in the work. This approach is used to describe the Turkish grammar in the terms of Link, but it is clear that it is applicable to other Turkic languages. It should be noted that this kind of research was carried out by other authors [14–16].

The development of dependency parser based on the Kazakh language treebank is described in the paper [14, 15]. The most difficult step in this work is the creation of a free open-source dependency treebank. The authors note that their work is in the initial stage, so it is too early to talk about results.

Creating a treebank for the Kazakh language is a very laborious and time-consuming process that requires the work of a large number of linguists. Currently, we did not aim to create the Kazakh treebank. However, we do not exclude that this work will be done in the future.

Çağrı Çöltekin [16] initiates a new experience for morphologic segmentation, stemming, lemmatization, unknown words differentiation conversion of grapheme to phoneme, hyphenation and a morphological disambiguation on the data of the Turkish Language. The tools, which are promoted in the research, give new possibilities to build in a free open-source morphologic analyzer for Turkish natural language processing.

Of course, there is no doubt of the significance of Çağrı Çöltekin's new results on morphologic segmentation. However, there is a question of scholarly interest, whether it is possible to apply a set of open-source tools instruments for other Turkic languages of Oghuz, Karluk, Kypchak groups? In this respect it is interesting to compare the effectiveness of our research on a broad range of Turkic languages.

III. Measures of the similarity between sentences and texts

The simple methods allowing us to define measures of the similarity between sentences are described in [6, 7]. Sentences can be considered as vectors, i.e. $\bar{x} =< x_1,\ldots,x_n >$, $\bar{y} =< y_1,\ldots,y_m >$, where x_i, y_i are words. Further, the sentences are analyzed by Link Grammar Parser. Then all pairs $< i_1, i_2 >, < j_1, j_2 >$ are considered such that the words x_{i_1}, x_{i_2} and y_{j_1}, y_{j_2} are connected by links of the same type. Thereby the words x_{i_1}, y_{j_1} and x_{i_2}, y_{j_2} must be similar according to some criterion, for example, they are synonyms, etc. The function words (for example, articles, conjunctions, etc) may be omitted. Let I be a set of all pairs mentioned above, and its cardinality is $|I| = n$. Measures of similarity of two sentences can be defined as follows:

$$\mu_0(\bar{x},\bar{y}) = n/\max(n_1,n_2) \text{ or } \mu_1(\bar{x},\bar{y}) = 2n/(n_1+n_2) \quad (1)$$

Where n_1, n_2 are the numbers of links obtained by Link Grammar Parser after the analysis of $\bar{x},\bar{y}$, respectively.

The paper [17] describes the generalization of the summarization algorithm of Niraj Kumar [18]. In the algorithm of Niraj Kumar, the theme weight is defined as a sum of all words concerning to this theme. The weights of the words (it is possible to consider them, as key words) are calculated by means of algorithm of reference ranging (PageRank algorithm) and proceeding from the assumption, that the text, as a whole, can be presented in the form of the oriented graph. It is the most interesting and important part of the algorithm. Let us note, that the algorithm works not with some selected sentences, but with the text as a whole.

Niraj Kumar considers the order of words. We consider more, namely (in addition), syntactic relations. Analogously our method uses Link Grammar Parser. Our investigations are oriented onto processing news articles, reviews from social networks, etc. We consider the possibility of applying this algorithm to estimate the relevance of posts published on the Internet to the selected articles published before.

We assign a graph $G_i(V_i, E_i)$ constructed by Link Grammar Parser on a base of a sentence S_i. In this graph, V_i is the set of words and E_i is a set of triplets $< v_1, v_2, t >$ where $v_1, v_2 \in V_i$ are vertices and t is the type of link. Thus we obtain $G_1,...,G_k$ are the graphs of sentences. On the next stage, we build a graph $G(V,E)$ by combining the sentence graphs $G_1,...,G_k$. Here $V = \bigcup_{1 \le i \le k} V_i$ is the set of all words from the sentences; E is a set of quadruples $< v_1, v_2, t, n >$ where, as previously, $v_1, v_2 \in V_i$ are vertices, t is the link type, and the additional parameter $n = \left| \{ i : < v_1, v_2, t > \in E_i \} \right|$ is the occurrence number of the $< v_1, v_2, t >$ triplet. Notions introduced by Niraj Kumar are generalized for this case.

IV. Links indicating syntactic features of words

Turkic languages belong to the group of agglutinative languages. A small limited set of links can be used, such as proposed in [10].

We have identified the following basic connections in the Kazakh and Turkish languages: AS is an attribute of a subject; AO is an attribute of an object; E is an adverbial modifier; J connects a postposition and a noun; OV is a direct object; OJV is an indirect object; S connects a subject and a predicate.

If we consider syntactic features of words in a sentence, then each part of speech can be associated with a formula of possible connectors: a noun may act as a subject connected to an attribute; a verb has to be at the final position and end a sentence, etc.

Here is an example of a sentence structure in the Turkish language: <N_S>: {AS–} & {OV+} & S+. Besides, a noun may act as an object, on the left of which is an attribute, on the right is a postposition and predicate. Such structure is generally described by the formula: <N_O>: {AO–} & {OV+} & {OJV+}.

V. Model of semantic markup of sentences

In order that to make a transition from the morphological and syntactic to semantic links, it is more convenient to carry out further considerations in terms of predicates. We have double predicates because we use link grammar. Thus, syntactic links discussed in the previous section, in some cases, can be saved in the form of predicates: AS (adjective, noun); AO (adjective, noun); E (adverb, verb); OJV (Nd (noun) | Na (noun) | Ni (noun) | Nl (noun) | Nb (noun), verb); S (Nn | Pn), verb), etc. Note that under this approach the predicates OV (x, y) and OJV (x, y) contain information about the verbal coordination, that is, they depend on the use of a specific case before the certain verb. In the future, we plan to carry out an additional study of verbal coordination in the Kazakh and Turkish languages. Now it is possible to consider the semantic predicate of possession: OF (Possessor, Possessed) = OF (Ng (noun) | Pg (pronoun), Np3 (noun)). The predicate OF (x, y) describes, for example, the phrase: *kadının elbisesi* ("women's dress", i.e. dress which belongs to the woman), where *kadın* is a stem of a word ("female"); *ın* is a genitive suffix; *elbise* is a stem of a word ("dress"); *si* is a possessive suffix. Consider the sentence: *Ben kardeşin kitabını okuyorum. (I am reading the brother's book.)*. Let us write this sentence with the help of the predicates: READ (ben, OF (kardeşin, kitabını)). The predicate OF enables emphasize the possessive pronouns. Figure 1 shows a parsing example, *Менің қарным ашқан жоқ. (I am not hungry.)*, containing the first person possessive pronoun (the link OF1 is responsible) and the negative form of the verb (the link VN is responsible).

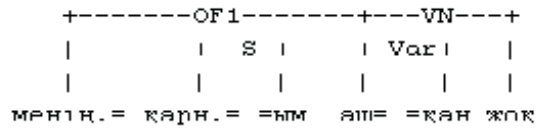


Fig. 1. Possessive pronouns in Kazakh language

A sentence parsing example with the possessive pronoun *Senin ne istedigini bilmiyorum. (I don't know what you want.)* is shown below.

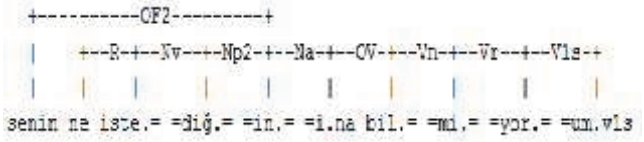


Fig. 2. Possessive pronouns in Turkish language

Semantic predicates of place LOC (verb, adverb) and time of action TIME (verb, adverb) are interesting from the perspective of further research. The predicate FOR (Ng (noun) | Pg (pronoun), postposition) describes a combination of a postposition "için" with a noun or pronoun in the genitive case.

VI. Conclusion

The study of the Turkic languages is stipulated by the need to analyze information in social networks, such as socio-economic, political, and radical Islamism. Investigations of this kind allow us to use Internet and social networks as a tool for influencing public sentiment and identifying social risks.

In this paper, we considered algorithms for the comparison of texts (as a sequence of sentences) and estimation of their

similarity. This method is applicable only to the sentences that can be quite correctly parsed by the Link Grammar Parser. The proposed measure takes into account lexical, syntactic and semantic relations between words.

We carried out experiments to assess the relevance of texts in English, Russian, Turkish and Kazakh languages. The volume of the Kazakh dictionary for the Link at the moment is about 500 words and 100 affixes. The size of texts in both languages for our experiment was 11-27 Kb.

There were used three methods of a comparison of sentences: measures μ_0, μ_1 and the generalization of the summarization algorithm of Niraj Kumar.

For example, the following results of the comparison of the Kazakh texts on the base of the generalized Niraj Kumar algorithm were obtained:
$(text\_1, text\_2) = 0.4727$; $(text\_1, text\_3) = 0.4364$;
$(text\_1, text\_4) = 0.4$; $(text\_2, text\_3) = 0.766$;
$(text\_2, text\_4) = 0.2215$; $(text\_3, text\_4) = 0.2123$.

The analogous results of the comparison of the Turkish texts are as follows:
$(text\_1, text\_2) = 0.6041$; $(text\_1, text\_3) = 0.5833$;
$(text\_1, text\_4) = 0.75$; $(text\_2, text\_3) = 0.1305$;
$(text\_2, text\_4) = 0.1188$; $(text\_3, text\_4) = 0.2055$.

Of course, here there is a very important question of a choice of a threshold. Namely, in which case we can say that a given text is similar another one, i.e. in particular, it is relevant to a query.

The accuracy of the Link Grammar Parser is mostly dependent on the completeness of dictionaries. In the preparation of dictionaries it is necessary to take into account some specific morphological characteristics with respect the agglutinative structure of the Turkic languages.

During the research, we also faced with problems of linguistic ambiguity: how to describe the homogeneity of the sentence part, using the link grammar, how to cope with homonymy of stems and affixes, etc.

REFERENCES

[1] G. Salton, "Automatic Information Organization and Retrieval," 1968, p.514.

[2] G. V. Lezin, V. A. Tuzov, "The semantic analysis of the text in Russian: semantico-syntactical model of the sentence," Economic-mathematical researches: mathematical models and information technologies. Is. 3. Saint Petersburg: Nauka, 2003, pp. 282–303. (in Russian)

[3] T. V. Batura, F. A. Murzin, "The machine-oriented logic methods of representation of semantics of the text in natural language," Novosibirsk: A.P. Ershov Institute of Informatics Systems SB RAS. Publishing Company of NGTU, 2008, p.248. (in Russian)

[4] D. Temperley, D. Sleator, J. Lafferty, Link Grammar Documentation, http://www.link.cs.cmu.edu/link/dict/index.html

[5] D. Sleator, D. Temperley, "Parsing English with a Link Grammar," Pittsburgh: School of Computer Science Carnegie Mellon University, 1991, p. 93.

[6] F. Murzin, A. Perfliev, T. Shmanina, "Methods of syntactic analysis and comparison of constructions of a natural language oriented to use in search systems," Bull. Nov. Comp. Center, Comp. Science. Iss. 31. 2010, pp. 91-109.

[7] F. Murzin, A. Perfliev, T. Shmanina, "Methods of syntactic analysis and comparison of constructions of a natural language oriented to use in search systems," Vestnik of Novosibirsk State Univ. Ser.:Information Technologies.. vol. 9, Iss. 4. 2012, pp. 13-28. Novosibirsk. (in Russian)

[8] T. V. Batura, F. A. Murzin, A. A. Perfliev, T. V. Shmanina, "Methods of the increase of the efficiency of information search on the basis of syntactic analysis," Novosibirsk: A.P. Ershov Institute of Informatics Systems SB RAS, Publishing Company of SB RAS, 2014, p. 76. (in Russian)

[9] D. Temperley, "An Introduction to the Link Grammar Parser," 2014, http://www.abisource.com/projects/link-grammar/dict/introduction.html#1

[10] U. A. Tukeyev, A. K. Melby, Zh. M. Zhumanov, "Models and algorithms of translation of the Kazakh language sentences into English language with use of link grammar and the statistical approach," Proceedings of IV Congress of the Turkic World Math. Baku, Society, 2011, p.474.

[11] Gülşen Eryiğit, Kemal Oflazer, "Statistical Dependency Parsing of Turkish," Proceedings of EACL 2006, 11th Conf. of the European Chapter of the Association for Comp. Linguistics. Trento, Italy, 2006, pp. 89-96.

[12] Gülşen Eryiğit, Joakim Nivre, Kemal Oflazer. "Dependency Parsing of Turkish, Computational Linguistics," vol. 34, N. 3. 2008, pp. 357–389.

[13] Ozlem Istek, Ilyas Cicekli, "A Link Grammar for an Agglutinative Language," Proceedings of Recent Advances in Natural Language Processing (RANLP 2007), Borovets, Bulgaria, 2007, pp. 285-290.

[14] F. M. Tyers, J. Washington, "Towards a free/open-source universal-dependency treebank for Kazakh," Proceedings of the 3rd Conference on Turkic Languages (TurkLang 2015), 2015, pp. 276-289.

[15] J. N. Washington, I. Salimzyanov, F. M. Tyers, "Finite-state morphological transducers for three Kypchak languages," Proceedings of the 9th Conference on Language Resources and Evaluation, LREC2014, 2014, pp. 3378-3385.

[16] Çağrı Çöltekin, "A Set of Open Source Tools for Turkish Natural Language Processing," Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) Ed. by N. Calzolari, 2014, pp. 1079-1086

[17] T.A. Batura, F.A. Murzin, D.F. Semich, A.M. Bakiyeva, A.S. Yerimbetova, "On some graphs connected with texts in a natural language, link grammar and the summarization process," Bull. Nov. Comp. Center, Comp. Science. Iss. 38. 2015, pp. 37-49.

[18] N. Kumar, K. Srinathan, V. Varma, "Using graph based mapping of co-occurring words and closeness centrality score for summarization evaluation," Proceedings of the 13th Intern. Conf. on Computational Linguistics and Intelligent Text Processing (CICLing'12), Ed. by A.Gelbukh, 2012, pp. 353-365.

Speech Recognition using Support Vector Machines

Kamil Aida-zade
Institute of Control Systems of Azerbaijan National Academy of Sciences, Baku, Azerbaijan
kamil_aydazade@rambler.ru

Anar Xocayev
Computer Engineering
Qafqaz University
Azerbaijan, Baku
axocayev@qu.edu.az

Samir Rustamov
ADA University,
Institute of Control Systems of Azerbaijan National Academy of Sciences, Baku, Azerbaijan
samir.rustamov@gmail.com

***Abstract*—In this article we applied Support Vector Machines to acoustic model of Speech Recognition System based on MFCC and LPC features for Azerbaijani DataSet. This DataSet has been used for speech recognition by Multilayer Artificial Neural Network and achieved some results. The main goal of this work is applying SVM techniques to the Azerbaijan Speech Recognition System. The variety of results of SVM with different Kernel functions is analyzed in the training process. It is shown that SVM with radial basis and polynomial kernels give better recognition results that Multilayer Artificial Neural Network.**

***Index Terms*—Speech Recognition, Feature Extraction, Support Vector Machines, Linear Predictive Coding, Mel Frequency Cepstral Coefficients.**

I. Introduction

Speech recognition is a main core of spoken language systems. Speech recognition is a complex classification task and classified by different mathematical approaches: acoustic-phonetic approach, pattern recognition approach, artificial intelligence approach (knowledge based approach), dynamic dime warping, connectionist approaches (artificial neural networks), support vector machine. Most of state-of-the-art systems use hidden Markov models (HMM), multilayer artificial neural networks (MANN) and convolution neural networks.

A Hidden Markov Models are robust to the uncertain or incomplete information which result from confusable sounds, variations in word sequence, homophones words, speaker rate and accent, and contextual effects in speech recognition. However, there are some disadvantages of HMM: density functions of the model are not adequately reflects the structure of the data; increasing the representative power of the model is not the best criterion for achieving optimum speech recognition performance; difficult to analyze the errors of an HMM systems in an attempt to improve its performance [4,11].

Another powerful classification algorithm is Multilayer Artificial Neural Networks which have been widely used in speech recognition. An artificial neural networks consist of neurons which affect each other by network weights. Some important properties (trainability, generalization, nonlinearity, robustness, uniformity, parallelism) of MANN makes them attractive for hardware implementation. The main disadvantages of MANN's are having local minumuma and requirement plenty of iterations over the large amount of labeled training data in the training process.

Support Vector Machines are one of powerful machine learning algorithm that have been used for data classification. SVMs maximize the margin between boundary points of the classes and the separating hyperplane. For the training of SVMs has been used quadratic programming problem and don't have convergence problem. In spite of that, SVMs face two essential problems in speech recognition: the variable length data sequences and multi-class decision[8,11].

Recently, Artificial Neural Networks [1,2,7] and Hidden Markov Model [3] have been applied for Azerbaijan speech recognition systems. In the work we focus on application Support Vector Machines techniques to the speech recognition based on MFCC and LPC features for Azerbaijani DataSet. This DataSet has been used for speech recognition by MANN and got some results. The main goal of this work is applying SVM techniques to the Azerbaijan Speech Recognition System. The variety of results of SVM with different Kernel functions in training process is analyzed. It is shown that, SVMs with radial basis and polynomial kernels give better recognition result that MANN.

II. Support Vector Machines

SVM is a discriminative classifier that have been used in many real world applications. SVMs define the "best" separating hyperplane by maximizing the margin between boundary points of the classes and the separating hyperplane. (Fig.1)[6]. These boundary points are referred as support vectors. SVMs use linear and nonlinear separating hyperplanes for data classification. Let's explain how SVM find an optimal hyperplane which categorizes new examples.

Let's denote feature vector by $x_i \in R^n, i = 1 \dots M$, where M is a number of training samples, n-is number of features of speech signal. Our goal is to classify speech unit into the two classes $y_i = -1$ (out of vocabulary unit), or $y_i = +1$ (in vocabulary unit) by hyperplane. This hyperplane is described as following form

$$w \cdot x + b = 0,$$

where w is a normal to the plane, b is a bias.

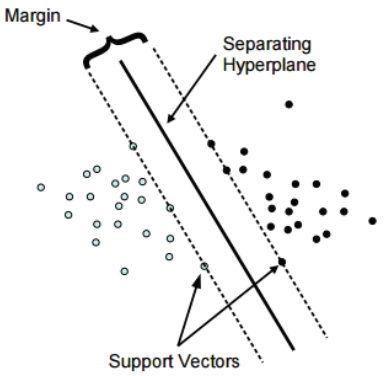


Fig. 1. Classification of two linear separable classes

SVMs construct this hyperplane (w and b) to be far from the boundary of both classes.

$$x_i \cdot w + b \geq +1 \ \ for \ \ y_i = +1 \quad (1)$$

$$x_i \cdot w + b \leq -1 \ \ for \ \ y_i = -1 \quad (2)$$

In order to find best hyperplane we need to maximize the margin with constraints in (1)-(2). In this case the margin is equal to $\frac{1}{\|w\|}$, therefore maximizing margin is equivalent to minimizing $\frac{1}{2}\|w\|^2$. This optimization problem solved by the method of Lagrange multipliers [9].

$$L_P \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{M} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{M} \alpha_i \quad (3)$$

We differentiate L_P with respect to w and b

$$\frac{\partial L_P}{\partial w} = 0 \ \Rightarrow \ w = \sum_{i=1}^{M} \alpha_i y_i x_i \quad (4)$$

$$\frac{\partial L_P}{\partial b} = 0 \ \Rightarrow \ \sum_{i=1}^{M} \alpha_i y_i = 0 \quad (5)$$

Substituting (4) and (5) into (3) gives Dual form L_D of the Primary L_P:

$$L_D \equiv \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i H_{ij} \alpha_j$$

where $H_{ij} = y_i y_j x_i \cdot x_j$

Minimizing L_P is equivalent to maximizing L_D:

$$\max_{\alpha} \left[\sum_{i=1}^{M} \alpha_i - \frac{1}{2}\alpha^T H \alpha \right]$$

$$\sum_{i=1}^{M} \alpha_i y_i = 0, \quad \alpha_i \geq 0, i = 1, \ldots, M$$

It is a convex optimization problem and can be solved by Quadratic Programming method which returns "best" α and it is allow us to calculate w by the formula (4).

In the next step we need to define the set of Support Vectors S which indices hold $\alpha_i > 0$ condition and calculate b:

$$b = \frac{1}{N_s}\sum_{s \in S} \left(y_s - \sum_{m \in S} \alpha_m y_m x_m \cdot x_s \right)$$

Each new sample x^* is classified by $y^* = sgn(w \cdot x^* + b)$ in the test process.

If it is impossible to find a linear separator, SVMs transform data (feature vectors) into an high dimensional space, where the training data become linearly separable, i.e. SVMs transform n-dimensional feature vector x in to an N-dimensional feature vectors[10]:

$$\Phi: R^n \rightarrow R^N$$

This transformation is realized via Kernel functions. There are defined different types of Kernel functions in SVM method.

Linear Kernel: $K(x_i, x_j) = x_i^T x_j$

Radial Bases Kernel: $K(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)}$

Polinomial Kernel: $K(x_i, x_j) = (x_i \cdot x_j + a)^b$

Sigmoidal Kernel: $K(x_i, x_j) = tanh(ax_i \cdot x_j - b)$

where a and b are Kernel's parameters.

Each new sample x^* is classified by the following formula for the non-linear Kernel based SVM

$$y^* = sgn\left(\sum_{s \in S} y_s \alpha_s K(x^*, x_s) + b\right)$$

SVM classification is a binary classification machine learning algorithm. That's why it is necessary to modify decision making part of the algorithm for the multiclass tasks in real world situations. Two most popular methods are used for multiclass classification: "one against one" (1A1) and "one against all" (1AA) techniques. The 1AA SVM multiclass approach involves the division of an N class dataset into N two-class cases, while 1A1 approach involves training a separate classifier for each pair of classes. This leads to N(N−1)/2 classifiers. This method is less sensitive to the imbalanced datasets but is much computationally expensive. In the work we use "one-against-all" SVM multiclass technique.

III. ALGORITHM OF CALCULATION OF SPEECH FEATURES

We applied SVM to the connected digit recognition task. The first step in speech recognition is to calculate feature vectors from the speech signal. We use two feature extraction algorithms for speech recognition: Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC). According to acoustics, the amplitude spectrum of a speech signal is dominant under $4\,kHz$. Therefore the speech signals is passed through a first-order finite impulse response filter.

$$s(n) = s_{in}(n) - \alpha \cdot s_{in}(n-1)$$

where $\alpha -$ is the filter coefficient $(\alpha \in (0{,}95;1))$, $s_{in}(n) -$ is the input signal.

We applied VAD algorithm to detect end point of an utterance in a speech signal. Speech is quazi-stationery signal in a short time interval. Therefore the input signal is divided into overlapping frames of N samples.

$$s_f(n) = s(n) \cdot w(n),$$

$$w(n) = \begin{cases} 1, & K \cdot r < n \leq K \cdot r + N, \ \ r = 0,1,2,\ldots,M-1, \\ 0, & otherwise, \end{cases}$$

where M is the number of frames, f_s is the sampling frequency, t_f is the frame length measured in time, and K is the frame step.

$$N = f_s \cdot t_f .$$

We use $f_s = 16kHs$, N=400 and K=160 in our system

The Hamming window is used to minimize signal discontinuities after framing:

$$s_w(n) = \left\{0{,}54 - 0{,}46\cos\left(\frac{2\pi(n-1)}{N-1}\right)\right\} s_{frame}(n), \quad n = 1,...,N .$$

MFCC feature extraction algorithms:

1. Fast Fourier transformation

$$bin_k = \left|\sum_{n=1}^{N} s_w(n) e^{-i(n-1)k\frac{2\pi}{N}}\right|, \quad k = 0,...,N-1 .$$

2. Mel filtering.

$$Mel(x) = 2595\lg\left(1+\frac{x}{700}\right), \qquad x = 700\cdot\left(10^{\frac{mel}{2595}} - 1\right),$$

$$f_{c_i} = Mel^{-1}\left\{Mel\{f_{start}\} + \frac{Mel\{f_s/2\} - Mel\{f_{start}\}}{NF} i\right\}, \quad ,$$

$$i = 1,...,NF-1$$

$$cbin_i = round\left\{\frac{f_{c_i}}{f_s} N\right\}, \quad NF = 24$$

$$fbank_k = \sum_{i=cbin_{k-1}}^{cbin_k} \frac{i - cbin_{k-1} + 1}{cbin_k - cbin_{k-1} + 1} bin_i +$$

$$+ \sum_{i=cbin_k+1}^{cbin_{k+1}} \left(1 - \frac{i - cbin_k}{cbin_{k+1} - cbin_k + 1}\right) bin_i , \; k = 1,2,..., NF-1.$$

where

$$cbin_0 = round\left\{\frac{f_{start}}{f_s} N\right\};$$

$$cbin_{24} = round\left\{\frac{f_s/2}{f_s} N\right\} = \frac{N}{2} .$$

3. Non-linear transformation.

$$f_i = \ln(fbank_i), \qquad i = 1,..,NF-1 .$$

4. Cepstral coefficients.

$$C_i = \sum_{j=1}^{NF-1} f_j \cdot \cos\left(\frac{\pi \cdot i}{NF-1}(j-0.5)\right), \; i = 1,..,12 .$$

5. Cepstral Mean Subtraction.

$$mc_j(q) = C_j(q) - \frac{1}{M}\sum_{i=1}^{M} C_i(q), \quad q = 1,2,,...,12$$

The LPC coefficients are calculated by the following formula:

$$c(k) = -a_p(k) - \sum_{i=1}^{k-1}\left(1 - \frac{i}{k}\right) a_p(i) c(k-i) , \qquad k = 1,...12 .$$

where $a_p(k)$ are found by applying Levinson-Durbin algorithm.

IV. EXPERIMENTS RESULTS

We applied SVM to the connected digit recognition task. Therefore we have 10 classes and our multiclass SVM system classify every utterance into 10 classes or reject recognition. In the work we use "one-against-all" method for multiclass classification.

In the training process, we use digits that pronounced by 70 different persons. Then speech features are extracted from the every speech signal via LPC and MFCC algorithms. Speech signals have different length in every utterance, however SVM applied to equal length feature vectors. We use the Lagrange interpolation method to lead utterances to the same scale.

The trainig and test process are carried out by the LibSVM toolkits[5,9]. As shown in table 2, SVM with Radial Bases kernel on LPC features gives better result, whereas SVM with polynomial kernel on MFCC features. According to the experiments' results the overall performance of SVM based speech recognition sligtly better than ANN [7] on the same DataSet.

TABLE I. THE RESULTS OF SPEECH RECOGNITION USING DIFFERENT KERNEL SVM ALGORITHMS.

	Feature type	
Kernal type	**MFCC**	**LPC**
Linear	95.01%	92.57%
Polynomial	**96.23%**	96.27%
Radial basis function	19.99%	**97.82%**

REFERENCES

[1] K.R.Ayda-zade, S.S.Rustamov. "Research of Cepstral Coefficients for Azerbaijan speech recognition system". Transactions of Azerbaijan National Academy of sciences. “Informatics and control problems”. Volume XXV, №3. Baku, 2005, p.89-94.

[2] K.R. Aida-Zade, S.S.Rustamov, E.E.Mustafayev. "Principles of Construction of Speech Recognition System by the Example of Azerbaijan Language". International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey 2009, p. 378-382.

[3] A. Abbasov, R. Fatullayev, A.Fatullayev. "HMM-based large vocablary continuous speech recognition system for Azerbaijani". Proceedings of PCI-2010, Baku 2010, Vol. 1, pp. 23-26

[4] M.A.Anusuya, S.K.Katti, “Speech Recognition by Machine: A Review”. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009. p.186

[5] Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: A Library for Support Vector Machines". March 4, 2013

[6] D. Meyer, “Support Vector Machines”. FH Technikum Wien, Austria, August 5, 2015.

[7] K.R.Aida-Zade, C.Ardil, S.S.Rustamov. "Investigation of combined use of MFCC and LPC Features in Speech Recognition Systems". International Journal of Signal Processing. Open Access Refereed Research Journal. Volume 3:2006, pp. 74-80.

[8] J. Padrell-Sendra, D.Martin-Iglesias, F.Diaz-de-Maria. "Support Vector Machines for Continuous Speech Recognition".

EUSIPCO 2006. 14th European Signal Processing Conference, Florence, Italy, September 4-8, 2006.

[9] T. Fletcher. "Support Vector Machines Explained". UCL.2009.

[10] A.Ganapathiraju, J.Hamaker, J.Picone. "Support Vector Machines for Speech Recognition". 2002.

[11] N.Smith, M. Gales. "Speech Recognition using SVMs". Advances in Neural Information Processing Systems 14. MIT Press, 2002.

[12] Aida-zade K.R., Rustamov S.S., Mustafayev E.E., Aliyeva N.T. "Human-Computer Dialogue Understanding Hybrid System". International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey 2012.

[13] Rustamov S.S. "An Application of Neuro-Fuzzy Model for Text and Speech Understanding Systems". PCI'2012. The IV International Conference "Problems of Cybernetics and Informatics", Baku, Azerbaijan, volume I, pp. 213-217, 2012.

[14] Rustamov S.S. "On an Understanding System that Supports Human-Computer Dialogue". PCI'2012. The IV International Conference "Problems of Cybernetics and Informatics", Baku, Azerbaijan, Volume I, pp. 217-221, 2012.

On a Mathematical Model of the Generalized Sequential Partially Covering Problem

Fidan Nuriyeva[1,2]
Department of Computer Science,
Dokuz Eylul University[1]
Izmir, Turkey
Institute of Control Systems of ANAS[2]
Baku, Azerbaijan
nuriyevafidan@gmail.com

***Abstract*—In this paper, Generalized Sequential Partially Covering Problem is presented. A Binary Integer Linear Programming model of the problem is proposed and it is explained with an example. The model has been implemented in GAMS and solved using the CPLEX solver.**

***Index Terms*—Bandpass problem, band collocation problem, mathematical modelling, binary integer linear programming problem, sequential partially covering problem, telecommunication applications.**

I. Introduction

The Bandpass Problem (BP) whose first mathematical model was presented in 2009 is a combinatorial optimization problem which may be used in telecommunication systems [1]. Due to the development in the technology, some problems become invalid or useless and they need to be updated. In this sense, Nuriyev et. al. announced the Band Collocation Problem (BCP) by extending the BP due to incompatibility with real life implementations at the present time [2].

The BP is related to transmitting data over fiber optic networks using the Dense Wavelength Division Multiplexing (DWDM) technology [1]. The data is transmitted from a source to other stations on different wavelengths in a single fiber optic cable. Stations add/drop data onto/from the cable via an optical device called Add/Drop Multiplexers (ADM). Special cards in ADMs control each wavelength. They can add/drop (extract) data at some wavelengths to/from a network path [3].

A Sequential Partially Covering Problem (SPCP) is a sub problem of the Band Collocation Problem (BCP) [4]. The aim of the BCP is to minimize hardware costs by organizing network traffic using wavelength division multiplexing (WDM) system [5]. Sequential Partially Covering Problem was firstly introduced at the 5th International Scientific Conference of Students and Young Scientists Theoretical and Applied Aspects of Cybernetics, (TAAC 2015) which was held in Kiev, Ukrain on 23-27 November and a method based on dynamic programming is also proposed in [6].

In this paper, a mathematical model of the Generalized Sequential Partially Covering Problem (GSPCP) is proposed.

II. A Sequential Partially Covering Problem (SPCP)

A definition of SPCP is as follows [6]:

Let A[m] be a sequence with m elements such that $A(i) \in \{0, 1\}$, $i = 1, 2, \ldots, m$.

Let S_l be a cover with 2^l cells, where $l=0, 1, \ldots, t$ and $t = \lfloor \log_2 m \rfloor$.

Let $d(S_l) = 2^l$ and $p(S_l) = p_l$ be the size and value of S_l with $l = 0, 1, 2, \ldots, t$ respectively.

Here,

$$p(S_0) < p(S_1) < p(S_2) < \ldots < p(S_t) \text{ and}$$
$$2.p(S_l) > p(S_{l+1}) \ (0, 1, \ldots, t - 1).$$

The aim is to cover all elements which is equal to "1" in $A[m]$ with S_l $(l = 0, 1, \ldots, t)$ with a minimum cost.

As it is seen, the size of S_l covers changes as a power of 2 to l. Determining size of S_l covers as q_l a definition of the Generalized Sequential Covering Problem is given below.

A mathematical model of the SPCP is given in [7].

III. Definition of the Generalized Sequential Partially Covering Problem (GSPCP)

In this section, a generalized version of the sequential partially covering problem will be discussed.

Let $M[m]$ be a sequence with m elements such that $M(i) \in \{0, 1\}$, $i = 1, 2, \ldots, m$.

Let S_l be a cover with q_l cells, where $l = 1, \ldots, n$, $n < m$.

Let $d(S_l)$ be a size of the cover of S_l. Here $d(S_l) = 1$ and other elements is $d(S_l) = q_l$, $(l = 2, \ldots, n)$, $q_l \in N$, $q_i \neq q_j$, $i \neq j$, $i, j = 1, 2, \ldots, n$.

Covers is sorted by ascending order depending on their size. $d(S_1) < d(S_2) < d(S_3) < \ldots < d(S_n)$.

Let $p(S_l)$ be a cost for a cover S_l.

$$p(S_1) = \begin{cases} 0, & \text{if the cell } M(r) \text{ covered by } S_1 \text{ is equal to 0} \\ c_1, & \text{if the cell } M(r) \text{ covered by } S_1 \text{ is equal to 1,} \end{cases}$$
$$(r = 1, 2, \ldots, m)$$

Here, c_i is a value defined for a one dimensional cover.

Here, we assume that

$p(S_1) < p(S_2) < p(S_3) < \ldots < p(S_n)$ and

$q_{l+1} \cdot c_l > q_l \cdot c_{l+1}$ $(l = 1, \ldots, n - 1)$.

As it seen from the equation given above, there is not any cost because the cells with "0" are not covered.

The aim is to cover all elements which is equal to "1" in M[m] with a minimum cost.

The definition of the GSPCP is as follows: Let A be a binary array of a q_k - *covers*. However, several q_k - *covers* cannot have a common element. A q_k - *cover* may include zero elements. Let c_k be the cost of forming a q_k - *cover*. The GSPCP is to minimize the total cost of q_k - *covers* so that all 1s are covered.

IV. An Example

Let an elements of the sequence of *A[n]* be as follows:

101110011011110

Let first take an overview to SPCP. In this problem a length of the sequences is

$d(S_l) = 2^l$, $d(S_0) = 1$, $d(S_1) = 2$, $d(S_2) = 4$, $d(S_3) = 8$.

$t = \lfloor \log_2(15) \rfloor = 3$.

Let $p(S_0) = 1000$, $p(S_1) = 1600$, $p(S_2) = 2560$, $p(S_3) = 4100$. Then an optimal cover for SPCP will be as follows: S_0, S_2, S_3, $p = 1000 + 2560 + 4100 = 7660$.

S_0		S_2		S_3
1	0	1110	0	11011110

Now let take an overview of this sequence for GSPCP. Assume that, we have covers which are given below:

$d(S_1) = 1$, $d(S_2) = 3$, $d(S_3) = 4$, $d(S_4) = 5$ and $d(S_5) = 7$.

$p(S_1) = 1000$, $p(S_2) = 2100$, $p(S_3) = 2560$, $p(S_4) = 2900$, $p(S_5) = 3500$. Then optimal cover (S_4, S_5) for GSPCP problem is as following: $P = 2900 + 3500 = 6400$.

S_4		S_5	
10111	00	1101111	0

V. A Binary Integer Linear Programming Model of GSPCP

A boolean programming model of the GSPCP is as follows: For a given binary array A and the costs c_k of q_k - *covers* we define two decision variables as follows:

$$y_i^k = \begin{cases} 1, & \text{if the entry } i \text{ is the first element of a } q_k \text{ - cover} \\ 0, & \text{otherwise, } (i,\ k = 1, \ldots, m) \end{cases}$$

$$x_i^k = \begin{cases} 1, & \text{if } a_i \text{ is an element of a } q_k \text{ - cover} \\ 0, & \text{otherwise, } (i,\ k = 1, \ldots, m) \end{cases}$$

We can formulate the boolean programming model of the GSPCP as follows:

$$Minimize \sum_{k=1}^{m} \sum_{i=1}^{m-q_k+1} c_k y_i^k \quad (1)$$

subject to

$$q_k y_l^k = \sum_{i=l}^{l+q_k-1} x_i^k; \quad k = 1,\ldots,m; \quad l = 1,\ldots,m-q_k+1 \quad (2)$$

$$\sum_{i=l}^{l+q_k-1} y_i^k \le 1; \quad k = 1,\ldots,m; \quad l = 1,\ldots,m-q_k+1 \quad (3)$$

$$\sum_{k=1}^{m} x_i^k \ge a_i; \quad i = 1,\ldots,m \quad (4)$$

$$\sum_{k=1}^{m} x_i^k \le 1; \quad i = 1,\ldots,m \quad (5)$$

$$\sum_{k=1}^{m} \sum_{i=1}^{m-q_k+1} q_k y_i^k \ge \sum_{i=1}^{m} a_i \quad (6)$$

$$\sum_{i=1}^{m-q_k+1} q_k y_i^k = \sum_{i=1}^{m} x_i^k; \quad k = 1,\ldots, m \quad (7)$$

$$x_i^k \in \{0,1\}; \quad y_i^k \in \{0,1\}; \quad i,k = 1,\ldots, m \quad (8)$$

The objective function (1) represents the total cost of forming q_k - *covers*. Constraint (2) guarantee to find the coordinates of q_k - *covers*. Constraint (3) guarantee that no two bands may have a common element. Constraint (4) say that each non-zero entry of the array has to be an element of a q_k – *cover*. Constraint (5) guarantee that any entry of the array belongs to at most one q_k - *cover*. Constraint (6) ensures that the total length of all q_k – *covers* cannot be less than the number of 1s. In other words, all 1s must be covered. Finally, constraint (7) express that the number of entries which are covered is equal to total length of the q_k – *covers*. In this model all decision variables are binary as in (8). The binary integer linear model (1)-(8) finds the q_k - *covers* with their coordinates that cover all 1s with the minimum cost.

VI. Conclusion

In this study, we focused on a generalization of the SPCP, which is a sub problem of the BCP that is proposed in [2]. A model which is proposed for SPCP in [7] is generalized for GSPCP.

We introduced a binary integer programming formulation for the GSPCP. In order to test the accuracy of the model, it has been coded in GAMS (the General Algebraic Modeling System) and solved using CPLEX solver [8].

In future, we are planning to generalize an algorithm for GSPCP based on dynamic programming that is proposed for SPCP in [6].

A length of the band collacations changes as multiples of 2 in SPCP, whereas it is an arbitrary number in GSPCP. Because of this feature, this problem could have many applications such as job sharing in different workplaces, promotion campaigns of companies and etc.

REFERENCES

[1] D. A. Babayev, G. I. Bell and U. G. Nuriyev, "The Bandpass Problem, Combinatorial Optimization and Library of Problems", J. Comb. Optim. Vol.18, 2009, pp. 151-172.

[2] U. Nuriyev, M. Kurt, H. Kutucu and A. Gursoy, "The Band Collocation Problem and Its Combinatorial Model", The International Conference Mathematical and Computational Modelling in Science and Technology, August 2-7, 2015, pp.140-141.

[3] I.P. Kaminov et al., A Wideband All Optical WDM Network, IEEE Journal on Selected Areas in Communications, Vol 14(5), pp.780–799, 1996.

[4] U. Nuriyev, H. Kutucu, M. Kurt and A. Gursoy, "The band collocation problem in telecommunication networks", Book of Abstract of the 5th International Conference on Control and Optimization with Industrial Applications (COIA-2015), August 27-29, pp. 362-365, 2015.

[5] A. Gürsoy, H. Kutucu, M. Kurt and U. Nuriyev, "A Binary Integer Linear Programming Model for the Band Collocation Problem", Proceedings of the 5th International Scientific Conference of Students And Young Scientists, "Theoretical and Applied Aspects of Cybernetics", (TAAC'2015), Kyiv, "Bukrek", Ukraine, November 23-27, 2015, pp. 131-134.

[6] F. Nuriyeva, "On a Solution of a Sequential Partially Covering Problem", Proceedings of the 5th International Scientific Conference of Students and Young Scientists "Theoretical and Applied Aspects of Cybernetics", (TAAC 2015), Kiev, Ukrain, November 23-27, 2015, pp. 32-36.

[7] F. Nuriyeva, H. Kutucu, A. Gürsoy, M. Kurt and U. Nuriyev, "On the Mathematikal Model of the Sequential Partially Covering Problem", Abstracts Book of the "*International Conference on Mathematics and Mathematics Education, (ICMME-2016)"*, Firat University, Elazığ, Turkey, May 12-14, 2016, pp. 395-397.

[8] GAMS development Corporation (GDC), General Algebraic Modeling Systems (GAMS), Washington, DC. www.gams.com, 2008.

[9] The Band Collocation Problem Library (BCPLib), http://www.izmir.edu.tr/bps (Lad: 10.10.2015)

Applying of ontological engineering to represent knowledge and training sessions

Bulat Kubekov
Department of Computer and program engineering
«Turan» university
Almaty, Kazakhstan
b.kubekov@mail.ru

Anar Utegenova
Department of Computer and program engineering
Kazakh National Technical University after K.I.Satpaev
utegenova77@mail.ru

Vitaliy Naumenko
Department of Computer and program engineering
«Turan» university
Almaty, Kazakhstan
naumenko.v5@mail.ru

***Abstract*— There are we consider the training technology, based upon the applying of ontological engineering for analysis and processing an educational information and design ontological models and "Mind Map" workspace capabilities. It demonstrate the scene of holding a lecture and using the "Mind Map" capabilities. Applying of the technology let to realize purposeful process of organization of students' educational activities upon the knowledge and skills at the expense of significant improving cognition capability.**

***Keywords-component:* ontology model, ontological engineering, support meaning, concretizing meaning, represent of knowledge.**

I. INTRODUCTION

The education quality depends on innovational constructing methods and applying such methods and means, which could let an education more effective.

There are the educational technologies are the forms of creating and transferring knowledge based upon the technical base as, in these researches, we use multi-paradigm concepts and mechanisms, object-oriented and generating programming in West Europa and America higher schools. There are the results of researches of these paradigms of information technologies and ontological engineering for represent knowledge and training lectures [1].

To solve the organization task and represent knowledge in our researches we use ontological engineering. Ontology as the way to represent semantical data, is defined conceptualization, underlying knowledge representation formalisms domain [2]. Ontology model is the hierarchical way of represent the set of concepts and theirs relations, the way which let to specify terminology on education content of the subject and relations between them. Ontological engineering as the heart of the "knowledge control" conception, directed to creation, propagation, execution and using of knowledge.

Methodical basis of organization and representation of knowledge in ontological engineering is the analysis of commonality and variability, because it is the only way to create abstractions by the human brain. The analysis of commonality and variability creates possibility to distinguish the support and concretizing concepts in the educational material and to endow every concept as mandatory, optional or alternative.

The main methodic to find and fixate commonality and variability is the characteristic design, which let to construct recycled educational components and to accept them to generate new educational programs, subjects and courses.

There was created the language of specification of knowledge for storage, computer processing and using models of ontologies in the educational process, and by dint of which everyone could describe ontological models, execute them and store in the conceptual database.

For processing the expressions of knowledge, construction and visualization of ontologies models there was developed a software interpreter, by dint of which you can enter expressions of knowledge, check the syntax, choose the model of representation, record the expression of knowledge in conceptual database and visualize the expression of knowledge graph. Below, on the Figure 1, you can see the interface software interpreter.

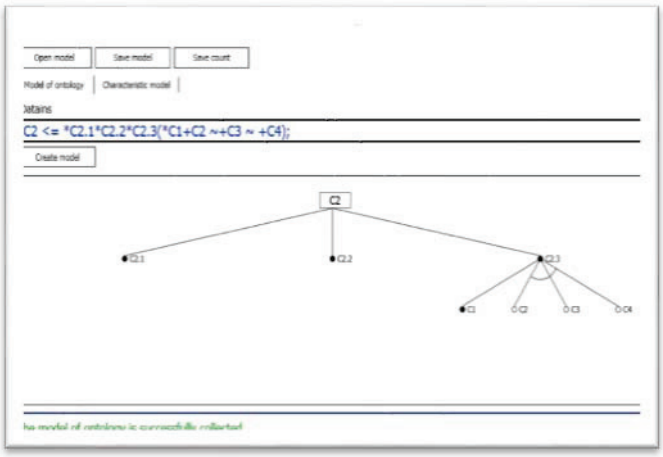


Fig. 1. Interface software interpreter

II. Thematic Structure Of The Discipline

During the lectures or self-study subjects, by means of the medium “Mind Map” the thematic structure of the discipline are demonstrating to the students. The thematic structure shows the place, the role and meaning of the subject and its constituent themes in the curriculum and qualification training of graduates, and besides that, emphasizes competences, which the discipline develops for the future professional activity. Further, the student chooses the topic of discipline and study it in the traditional presentation format. In subsequent phases, educational content, conclusions, are provided to the student, but in the format of ontology models and expressions of knowledge, as well as the recommended literature and sources of information.

Thus, the student has the ability to visualize and understand the role and importance of discipline for his professional development. He have the opportunity to learn the topic in the ordinary, traditional and knowledge views, which obviously helps, firstly, to increase the cognitive ability to learn and as a consequence, effective memorization, reproduction and application of knowledge to solve practical tasks and exercises, provided by this discipline.

The main method for these purposes of technology training is a method of extracting the concepts of the discipline, constructing hierarchical structures from the support and specifying concepts - models of ontology, modeling of the commonality and variation properties, by dint of aggregation, composition, alternative and presentation of models ontology expressions of knowledge.

Next, the script will show the visual form of lectures. Figure 2 shows the structure of the discipline "Software Engineering" and the structure of educational modules on each of the topics.

Fig. 2. The structure of the topics and topic educational modules

The first educational module – the content of educational material of the discipline topic in traditional exposition preceded by information on the role of the studied discipline in professional activity, as we can see on Figure 3.

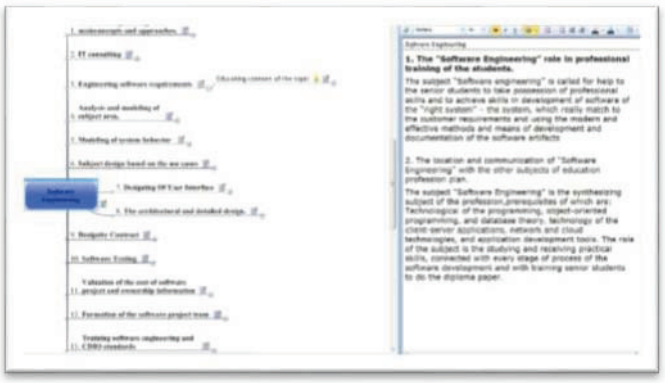


Fig. 3. Educational module "Information about the role of educating discipline in professional activity"

Next, the student chooses a topic, for example “Engineering software requirements”, read the information about the professional competence develops the topic and begin to learn it in common format, as we can see on Figure 4.

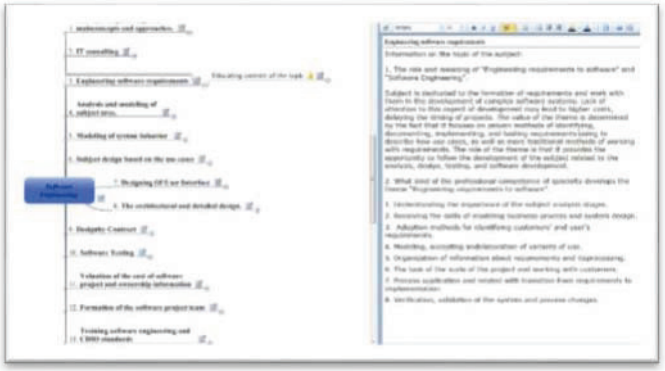


Fig. 4. Educational module "The usual format of academic information"

The following educational modules: "Models, concepts and expressions of knowledge", "Summary and Conclusions", "Control questions and exercises on the topic" develop the topic, but knowledge-representation, with using the support and specifying concepts of the topic and their representations, ontology models and expressions of knowledge, as shown at the Figure 5.

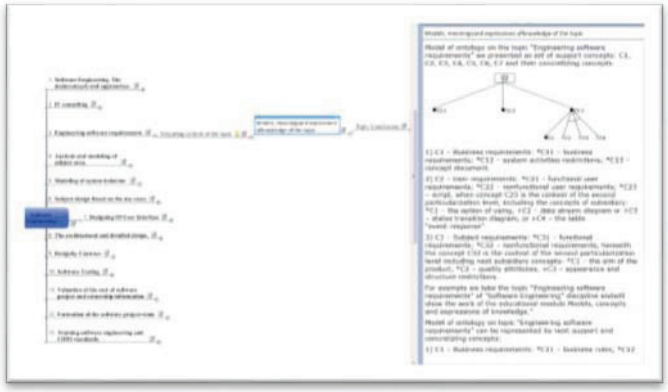


Fig. 5. Educational module “Models, concepts and knowledge expressions”

After the educational module “Models, concepts and knowledge expressions” follow the modules in which with

dedicated concepts and expressions of knowledge, formulate conclusions, as well as recommended educational literature and information sources on the topic, as shown at the Figure 6.

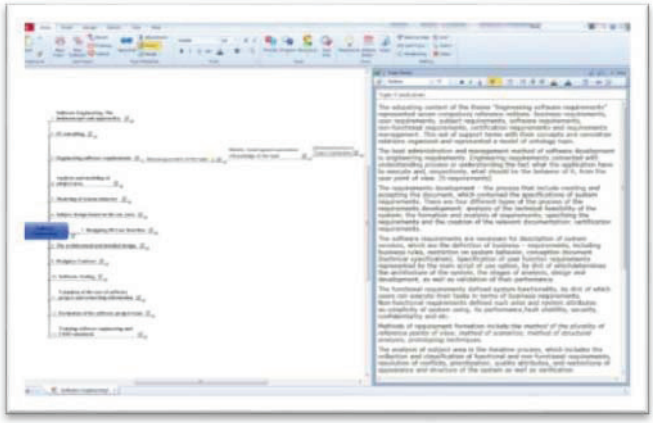


Fig. 6. Educational module "Summary and conclusions on the theme.

III. Models Of Ontology On Topic "Engineering Software Requirements"

For example we take the topic "Engineering software requirements" of "Software Engineering" discipline and will show the work of the educational module Models, concepts and expressions of knowledge".

Model of ontology on topic "Engineering software requirements" can be represented by next support and concretizing concepts:

*1) C1 – Business requirements: *C11 – business rules; *C12 – system activities restrictions; *C13 – concept document.*

*2) C2 – User requirements: *C21 – functional user requirements; *C22 – nonfunctional user requirements; *C23 – script, when concept C23 is the context of the second particularization level, including the concepts of subsidiary: *C1 – the option of using, +C2 – data stream diagram or +C3 – states transition diagram, or +C4 – the table "event-response"*

*3) C3 – Subject requirements: *C31 – functional requirements; *C32 – nonfunctional requirements, herewith the concept C32 is the context of the second particularization level including next subsidiary concepts: *C1 – the aim of the product, *C2 – quality attributes, +C3 – appearance and structure restrictions.*

*4) C4 – system requirements: *C41 – organization of system requirements; *C42 – system requirements quality metrics; +C43 – return requirements.*

*5) C5 – Nonfunctional requirements: *C51 – requirements to the product, herewith C51 is the context of the second particularization level including next subsidiary concepts: *C1 – exploitation requirements, *C2 – efficiency requirements; *C3 – reliability requirements; +C4 – portability requirements, *C5 – failures processing, *C6 – interface requirements. C52 – organization requirements, herewith the concept C52 is the context of the second particularization level including next subsidiary concepts: *C1 – exiting requirements, +C2 – realization requirements, *C3 – standarts requirements. *C53 – external requirements, herewith the concept C53 is the context of the second particularization level including next subsidiary concepts: *C1 – interaction requirements, +C2 – ethic requirements, +C3 – legal requirements.*

*6) C6 – Certification requirements: C61 – reliability; + C62 – Consistency; *C63 – fullness; *C64 – feasibility*

*7) C7 - Requirements management: *C71- management planning; * C72 - modification management.*

Next, for some support concepts of the topic, we create knowledge expressions and visually show each of them by graph. For example, the knowledge expression of support concept C2 – User requirements, will have the following form:

$$C2 <= *C2.1*C2.2*C2.3(*C1+C2 \sim +C3 \sim +C4); \quad (1)$$

and the graph of the model will be:

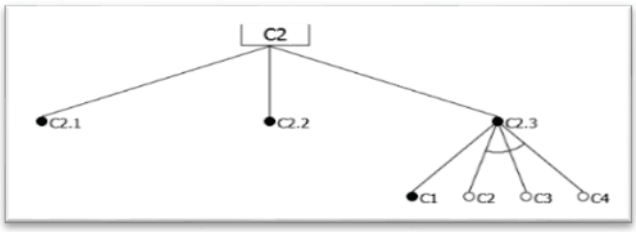


Fig. 7. Graph of the model

The knowledge expression of support concept C5 – Nonfunctional requirements will have following form:

$$C_5<=C_{5.1}(*C_1*C_2*C_3+C_4*C_5*C_6)*C_{5.2}(*C_1+C_2*C_3)*C_{5.3}(*C_1+C_2+C_3); \quad (2)$$

and the graph of the model will be:

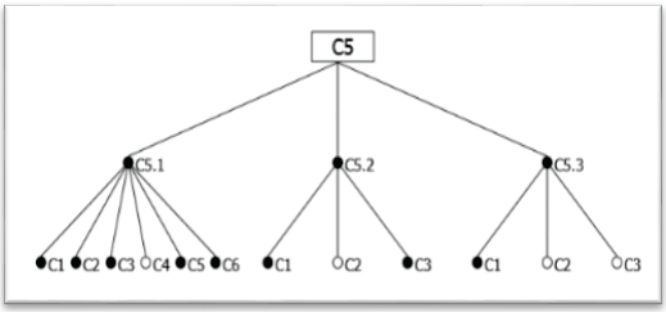


Fig. 8. Graph of the model

Conclusion

Thus, the possibility of applying the concepts and tools of ontological engineering are quite effective methods of presentation and organization the educational material for the training and monitoring of students' knowledge. The innovativeness of the proposed approach is that models of ontologies and expression of knowledge allows to structure

and to model the conceptual content of academic disciplines and effectively apply them in training, and the possibility of storage and computer processing expressions allow knowledge to solve problems of design disciplines and educational programs.

Visibility, clarity and a clear structuring of learning content subjects using ontology models and expressions of knowledge contributes to the qualitative improvement of cognitive abilities of students and, consequently, better memorization, reproduction and use of knowledge. This technology contributes to the acquisition of teaching experience of formation and application of knowledge, not only in the educational process, but also in everyday life, and motivates the acquisition of knowledge in their future professional activities.

REFERENCES

[1] B.S. Kubekov2, A.K. Muslimova1, M. Kh. Ishimova1, I.T. Utepbergenov1. The use of multiparadigm approach to knowledge modeling. 7th International Conference on Education and New Learning Technologies. Barselona (SPAIN) 6th-8th of July, 2015. pp. 8366-8368.

[1] [2] Bulat Kubekov, Janna Kuandykova, Irbulat Utepbergenov, Anar Utegenova. Applying of the conceptual model of knowledge for formalization of concepts of educational content. 9th International Conference "Applying of information and communication technologies-AICT2015", Rostov-on-Don, Russia, 14-16 October 2015. pp.588-594..

Visualization of Parallel Modeling of Dynamical Processes on stochastic Kronecker graphs

Anastasia A. Lantseva, Claudia K. Bochenina, Sergey V. Ivanov
ITMO University
Saint-Petersburg, Russia
stasja16@inbox.ru, k.bochenina@gmail.com, sergei.v.ivanov@gmail.com

***Abstract*— This article focuses on the visualization of parallel modeling of dynamic processes on complex networks. The proposed method based on the parallel visualization of each part of parallel algorithm of modeling. This approach is applied to visualization of the stochastic Kronecker graph. The experiments have shown that our approach facilitates the visual analysis and helps to investigate the dynamic processes into network.**

***Index Terms*— Visualization, Complex Network, Parallel algorithm, Dynamic processes, Kronecker graph.**

I. Introduction

Recently, the interest has increased to complex systems and networks. The visualization of these objects is the required part of scientific knowledge. The ability to observe the data, their interrelations and movements helps more detail to understand of the nature of these data, also to see the dynamic. The complex network is the system that is including the real objects and the relationships between them [3], [4], and [5]. Usually complex network is simulated by a graph with a dynamic evolving structure over time, which has a specific structure and characteristic features. Often many studies require the dynamic graph visualization [6], [7], [8]. If in the case of the consistent system the software visualization is the auxiliary part, in the parallel computing it is the necessary part. Really, the visualization is used as way of debugging of parallel program. Frequently the tasks of such visualization are an ability to display such notions, as compute nodes, sending of messages, multithreading, parallel processing. Now there are a many works and studies that have described the certain software visualization [9], [10], [11], [12]. Sometimes the programs of visualization of parallel processing base on a parallel card and a space-time diagram. The parallel card [13], [14] displays a possible parallelism of tasks. It is the data structure for regrouping and the visual method to display parallel processes. The card shows the history of processing of processes as the event's flow on the timeline. The space-time diagram [15], [16] is being represented as the event's flow in the two-dimensional space, where the axis is time; the axle is different real processes. During debugging of parallel programs, there are some classes of mistakes that are related to a lot of flows (processes) that are being performed simultaneously, and the asynchronous interactions. During debugging of parallel programs [10], [12], [17] there are some classes of mistakes, that are related to a lot of flows (processes) that are being performed simultaneously, and the asynchronous interactions. In addition to basic text debuggers on separate compute nodes, the big effect is reached by the visual resources of debugging. Visualization may be "posthumous", i.e. it will be realized after the program will have finished and all of the data will have collected. Or it may be "online", which are working parallel to program, and such visualization is more preferred.

Mostly, three major aspects can be selected in the design and the development of computer visualization of parallel processes. Primarily, it is the problem of creation, storage, processing and image output. The second aspect is the problem of software. Finally, the third aspect relates to the issues of perception and interpretation of visual image of parallel processes. The main task of computer visualization is the creation of methods of visual presentation for full interpretation of simulation results.

II. Parallel Algoriphm Of Complex Network Generation

The goal of considered algorithm is to model spreading of SIRS (Susceptible – Infected –Recovered (Removed) – Susceptible) process on stochastic Kronecker graph (SKG). SKG is a widely used network generative model proposed by Leskovec et al. [1] which reproduces a graph in a recursive manner using small-size (usually 2x2) initiator matrix. Inventors of SKG also developed KronFit algorithm to obtain an initiator matrix from a sample of a real-world graph by maximum likelihood estimation.

In addition to topological properties' preservation of sample graphs, one of the SKG advantages is the possibility to create graphs with any desired size from the same initiator matrix.

When sizes of studied networks grow, there is a need to use distributed processing to store data and to model interactions between vertices in parallel. In a recent work [2] a parallel algorithm to model SIRS processes on Poisson SKGs was proposed. The algorithm combines generation of network with simulation of interaction between vertices during spreading the process (see Algorithm 1). Formalized description of the algorithm is presented in [2].

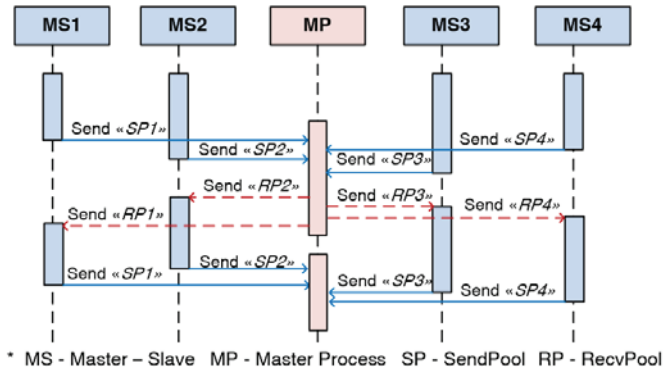


Fig. 1. Sequence diagram of Algorithm 1

Algorithm 1 represents general description of modeling algorithm for a single subnetwork which is attributed to a particular computational process (CP). The mechanism of exchange of information about activated vertices between subnetworks (i.e. communication pattern) can vary. The overall parallel performance of an algorithm is mostly determined by two factors: (i) uniformity of load on CPs; (ii) efficiency of communication pattern. In [2] there was proposed a load balancing algorithm which exploits the properties of initiator matrix and allows to achieve near-optimal load distribution. To avoid all-to-all communication between subnetworks (this pattern significantly reduces scalability of solution), authors of [2] suggest to use Master-Slave communication pattern with varied number of Master processes. Thus, after processing the pool, each Slave process sends its *SendPool* to corresponding Master process, and receives *RecvPool* several times from each of Master.

```
// Pool — a pool of activated vertices, STEPS — a number of
iterations to model, PROB_ACT — the // probability of
activation of a neighbor per iteration, SendPool — a pool of
vertices to be sent to other // subnetworks, RecvPool — a pool
of vertices received from other subnetworks
1: if (step == 1):
2:         receive initially infected vertices to Pool
3: while (step != STEPS):
4:         for each vertex v in Pool:
5:                 if necessary, generate edges of v
6:                 for each e in v.edges:
7:                         if rnd <= PROB_ACT:
8:                                 if e belongs to the same
                  subnetwork, activate e
9:                                 else add e to SendPool
10:        send SendPool
11:        receive RecvPool and activate vertices
12:        step = step + 1
```

Fig. 2. Algorithm of parallel modeling of SIRS on stochastic Kronecker graph

Along with the Master-Slave pattern there can be used other patterns such as bulk synchronous processing (BSP), ring or ring of rings when Slave processes and Master processes form distinct sets of rings. Taking into consideration high labor intensity of analyzing the behavior of message passing algorithms, it has become necessary to illustrate how different patterns work in terms of communication to computation ratio and overhead of communication operations. In this paper, we present a framework for visualization of parallel modeling of dynamical processes on SKG towards this end.

III. Visualization Technique

As results of the parallel computing, we obtain data, the volume of which is difficult to assess. Visualization of such data is very hard task and almost impossible. Thus, there is task to reduce these data. In our research, the information about modeling subnetwork is stored in the memory of the Master-Slave, where this subnetwork is being modeled, at the each iteration. Also, in the memory of Master processes, there is stored information about the exchange of information about activated vertices between subnetworks. In this case, data integration from all computing nodes does not matter.

A. Visualization of parallel processes

The main aim of this study is the visualization of parallel processes. Fig.3 shows a sequence diagram of the visualization process. As we can notice the process of visualization differs from process of modeling (fig.1). We have already known, how and where will happen the modeling of each subnetwork. In that case, we need to visualize each subnetwork simultaneously, then to send *SendPool* from each *Master-Slave* to *Master process* and to visualize its process. During visualization of *Master process* work, all *Masters* are waiting the barrier, after this there starts a new iteration. We do not need to send *RecvPool*, after finish of the visualization *Master process*; it sends the message to all of Masters about the new visualization iteration. Thus, this approach reduces the number of received messages between Masters and allows all of subnetworks start the new iteration of visualization simultaneously. That facilitates the synchronization all of the processes.

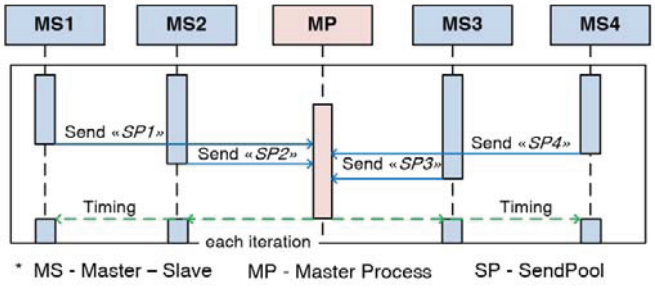


Fig. 3. Sequence diagram of visualization process

B. Visualization of Master –Slave or Subnetwork

When we talk about *Master – Slave* work, we understand that Master-Slave is the compute node, where is subnetwork

modeling. The vertices, that will be included in the subnetwork, have been predetermined on each computer node. Thus, if during modeling a link or an attempt to infection appear between the couple of vertexes and one of them is located on another subnetwork, the node that has tried to connect sends a message to its *Master – Slave*. In this case, during modeling this situation means that the *Master – Slave* records this message in its *SendPool*. But during visualization it means that the size of *Master-Slave* increases. And the moment of message transmission is demonstrated by vertex coloring that transmits the message in Master - Salve's color. And the edge between the vertex and *Master - Salve* is painted the same color (fig.4). After the barrier, new iteration starts with the appearance of new vertexes that were obtained from the other subnetworks.

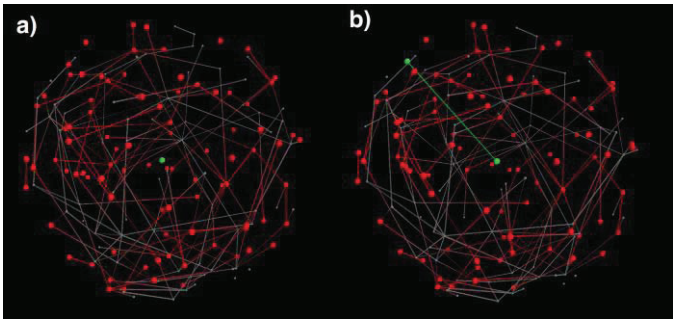


Fig. 4. Visualization of Subnetwork a) work of subnetwork b) send of message

C. Visualization of Master process

Work of *Master process* is portrayed as the network, where *Master process* is the center vertex and it has links with all of *Masters – Slaves.* Each of them has its edges and vertex that meant messages (new links or attempt to infection between subnetworks) in *SendPool*. As soon as the message arrives in the *SendPool*, *Master – Salve* receives a new vertex.

This vertex (message) has color of *Master – Slave*, whom it is addressed. At the end of modeling of subnetworks, each *Master – Slave* send all its vertexes to Master process, this process occurs sequentially. After *Master process* sends all of vertexes to *Masters – Slaves* depending on their color. Thus by the end of visual iteration, each *Master – Slave* has new vertexes of its color.

IV. Implementation

The framework for visualization of parallel modeling of dynamical processes on SKG was implemented by the High level programming language Python with cross-platform tool for 3-D scientific data visualization MayaVi. The framework runs on platforms MS Windows 7 (8 / 10) and Linux where Python, VTK and Mayavi were installed. The minimum system requirements for the proper and correct work have shown in the table 1.

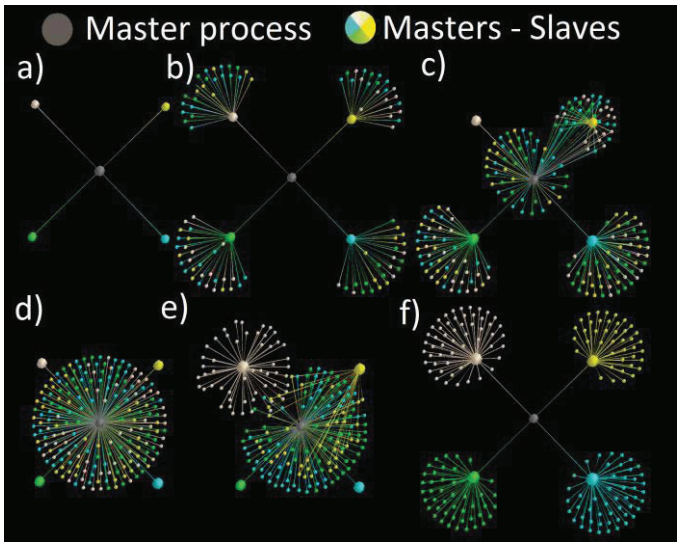


Fig. 5. Visualization of Master process's work a) initial stage; b) add new vertex on Masters - Slaves; c) send SendPool; d) the processing of the data on the Master process; e) send the vertex on Masters – Slaves f) the end of visual iteration.

TABLE I. The system requirements

Graphics Hardware	GTS 250; GeForce 800 GTX; Radeon HD 3870 X2
RAM	2 GB of RAM (4 GB recommended)
Disk Space	1 GB of free disk space for install

The input data are file for each subnetwork, which is containing links (edges) and the process code (fig.6).

```
ID_vertex0  ID_vertex1   "code"
ID_vertex2  ID_vertex3   "code"
...
ID_vertexN "code"
----------------------------------------
codes:
0 - new link
1 - attempt to infection
2 - infection
00 - convalescence
```

Fig. 6. The file format for input data

V. Results

During work, we have been able to demonstrate our approach on modeling of the stochastic Kronecker graph.

Visualization was performed on graphы, which included 1000 (fig.7), 2000(fig.8) and 4000 vertexes.

The network which includes 1000 vertexes is shown on figure 7. As we can see the subnetworks have the scattered structure, because each subnetwork has about 200 vertexes. Thus for networks with number of vertexes less, than 1500, our approach is redundant. And this case, it may be necessary to reduce the number of computer nodes in parallel algorithm or another approach to visualization of subnetworks.

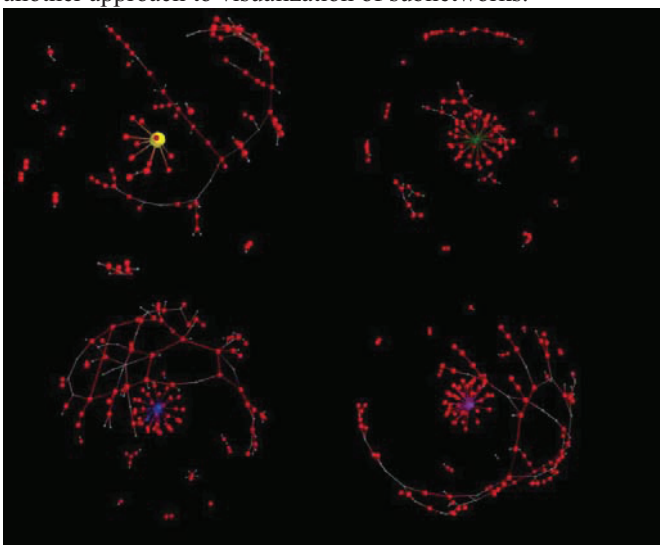


Fig. 7. Parallel visualization of networks with 1000 vertexes.

Figure 8 demonstrates the several stage of graph visualization, the graph has 2000 vertexes [18], fig.8(a) shows the second iteration, where we can see, that *Master – Slaves* are increased in size, this means the *SendPool* are filled with message. Also this act is shown in the center, when we can see new vertexes on each *Masters - Slaves*, which means the same message. The sending of messages to Master process is demonstrated on fig.8 (b)). The reverse operation of sending messages to *Master-Slaves* is shown on figure 8(c). And the last picture (fig.8 (d)) accordingly shows the ending of iteration.

VI. Conclusions And Future Works

We have proposed a new approach for visualization of parallel modeling of dynamical processes on complex network. There is presented the particular case on complex network, it is stochastic Kronecker graphs. In our research, we were using the existing layout algorithms for dynamic graph that based on psychical analogies (force - direct) [19]. As result, we have been able to obtain the full vision of the structure and the process of parallel modeling which clearly helps to monitor and to analyze the work of parallel algorithm. Thus, we can affirm, that the proposed approach to the visualization of parallel modeling of graphs useful for the complex networks with dynamic processes into them.

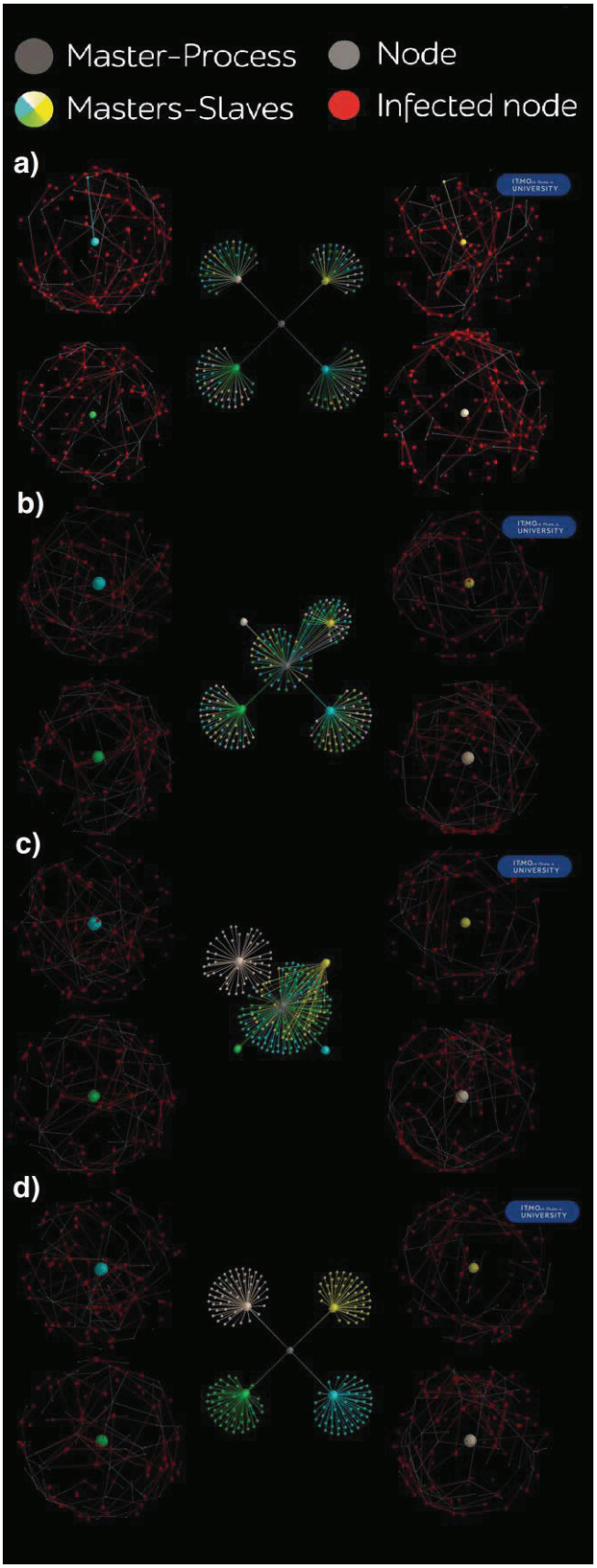


Fig. 8. Visualization Stages

There are several areas for future research. During the visualization we noticed that the force-direct algorithm did not always give the best view of subnetwork. When the size of subnetwork is the sufficiently large, the graph has a spherical shape, which leads the difficult of visual analysis. Also, by increasing the number of *Masters – Slaves* the general view of parallel algorithm becomes congested. Thus we assume that the optimal number of Masters-Slaves for visualization is four.

An important direction of development, in our view, is the visualization of other topologies, for example “Ring Network” or “Grid network”.

ACKNOWLEDGMENT

This paper is financially supported by The Russian Scientific Foundation, Agreement # 14–21–00137 (15.08.2014).

REFERENCES

[1] Leskovec, D.J. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani,. (2010). Kronecker graphs: an approach to modeling networks. *http://eprints.pascal-network.org/archive/00007816/* , 985–1042.

[2] Bochenina, K. K. (2016). A parallel algorithm for modeling of dynamical processes on large stochastic Kronecker graphs. *Procedia Comput. Science (In print.)* .

[3] Ghoniem, M., Fekete, J.D., Castagnola, P. (2004). A comparison of the readability of graphs using node-link andmatrix-based representations. *In Proc. IEEE InfoVis* , pp. 17–24.

[4] Branke, J. (2001). Dynamic graph drawing. *Int. Symp. Graph Drawing* , 228–246.

[5] Diehl, S. G. (2002). Graphs, They Are Changing - Dynamic Graph. *Graph Drawing*, 23-31.

[6] North, S. C. (1995). Incremental layout in dynadag. *Proc. 3rd Int. Symp. Graph Drawing* , 409–418.

[7] Boccaletti, S. (2006). Complex Networks: Structure and Dynamics. *Phys. Rep*, 175-308.

[8] Dorogovtsev, S. N.,Mendes,J.F.F. (2002). Evolution of Networks,. *Adv. Phys*.

[9] Ahrens J. et al. (2001) Large-scale data visualization using parallel data streaming *Computer Graphics and Applications, IEEE*, 34-41.

[10] Kranzlmüller, D., Volkert, J.(2000) Why Debugging Parallel Programs Needs Visualization? *Pros. VMPDP 200*.

[11] Kraemer E., Stasko J. T., (1993) The visualization of parallel systems: An overview *Journal of Parallel and Distributed Computing,* 105-117.

[12] Kirk, D. B., Wen-mei, W. H. (2012) Programming massively parallel processors: a hands-on approach. – *Newnes*.

[13] . Stone, J. M. (1988) A graphical representation of concurrent processes *ACM,* 226-235.

[14] Socha D., Bailey M. L., Notkin D., (1988) Voyeur: Graphical views of parallel programs. *ACM,* 206-215.

[15] Adhianto L. et al. (2010) HPCToolkit: Tools for performance analysis of optimized parallel programs *Concurrency and Computation: Practice and Experience*. 685-701.

[16] Allen, M. B. (1999)Parallel programming. – Upper Saddle River, *NJ : Prentice hall.*

[17] Wu, X. (2012) Performance evaluation, prediction and visualization of parallel systems. *Springer Science & Business Media*.

[18] http://escience.ifmo.ru/research/view/20

[19] Fruchterman, T.M. J., Reingold, E. M. (1991) Graph drawing by force – directed placement. // *Pract. Exper*. 1129–1164.

An Object-Oriented Model of Case-Based Reasoning System Using Situations Tree

Ruslan K. Akhmadulin
Department of Cybernetic Systems
Tyumen Industrial University
Tyumen city, Russia
ark@tsogu.ru

Igor N. Gluhih, Ivan Y. Karyakin
Information System Department
Tyumen State University
Tyumen city, Russia
igluhih@utmn.ru, ivan_nyc@mail.ru

Abstract— **The article is dedicated to development of an object-oriented model of an intelligence system based on case-based reasoning (CBR). An upgrade of the decision tree has been introduced to represent the situations hierarchy and related decisions in the knowledge base. The construction of the situations tree has been described. The problem of representing the situations tree in terms of classes and objects has been solved and the method of storage of the situations tree in the database has been defined. As a result, on the basis of the object-oriented paradigm the proposed approach allows creating domain-independent CBR systems that are shells for filling with case-oriented knowledge to solve specific problems.**

Index Terms - case-based reasoning, object-oriented model, decision tree.

I. INTRODUCTION

In modern intelligence systems of decision support – decision support system (DSS) – one of the most popular concepts of reasoning is case-based reasoning (CBR). The basic idea of CBR is to use knowledge about similar problems and their solutions, already familiar from previous experiences, to search for solutions to new problems [1-2]. The universality and efficiency of this approach allows using it in the development of activity and decision-making support systems in various subject areas: medicine [3, 4], construction and industry [5-7], software engineering [8], financial sector [9, 10], development of recommender systems [11], corporate knowledge bases [12], complex objects management automation [13-15], etc.

Cases in CBR represent recorded in the knowledge base descriptions of known situations and the decisions which were made in these situations and yielded positive results or were deemed rational by the experts forming the knowledge base [2]. To formalize the representation a couple P = <situation (Sit), resolution (R)>, is introduced which can be supplemented with other auxiliary elements, for example, the results of the decision [2]. The very knowledge base (KB) of such a system is called cases base.

In reasoning a current situation (problem) is identified and case P* is searched for which meets some criteria of closeness to this situation. Next, the solution from the P* couple is applied. This may be accompanied by adaptation of the solutions [1, 2, 16], which takes into account possible differences between the current situation and that, contained in KB. If in the course of solution adaptation and application new results for a new problem have been obtained, a new case is included into KB.

Among the most important research tasks in the development of CBR systems there are the tasks of searching for and selection of cases for KB, for which two main approaches are used [1, 2]. The first one involves search and selection of cases based on calculation of situation attributes proximity in space using some metrics; the second one is sequential analysis of the attributes using a decision tree. And although the second approach is less universal and flexible, more time consuming to edit or augment the knowledge base, it has certain advantages. Thus, the occurrence of the decision tree already specifies the arrangement of the cases in KB and the very algorithm of output.

In this work, we introduce an upgrade of the decision tree to represent the situations hierarchy and related decisions in KB. On this basis an object-oriented model of CBR-systems, invariant to the subject area, has been developed; it is applicable to creation of reasoning problem-oriented systems in various subject areas.

II. PRELIMINARIES

For an object-oriented model of a CBR system knowledge representation and reasoning are used on the basis of the situations tree (ST) [17]. When building the ST a ratio ρ = «supertype-subtype» (is-a) on a variety of situations is introduced and a formal description of a situation through a variety of its attributes is specified:

$$Sit = \{A_i \mid i \in M^k\}, \quad (1)$$

where Mk is a set of indices of attributes, the capacity of which depends on the level k in the situations tree.

A_i is an i-th situation attribute: $A_i = <ID_i, D_i>$, ID_i is an attribute identifying code (name); D_i is a set of attribute values.

In general case it is assumed that the attribute value can be of any type, i.e. any known quantitative or qualitative scales are used.

When building a ST, a number of conditions are adopted which regulate the creation of cases base and allow arranging reasoning [18]:

- the situations of each lower level are a refinement (extension of the definition) of the situations of the upper level, i.e. $M_k \subseteq M_{k+1}$. In other words, a situation of the lower level inherits the set of situation attributes of the upper level.
- every situation in the ST is a subtype of only one situation-prototype; it can serve as a prototype for an unlimited number of situations of the lower level;
- the situations of the same level are described by the same set of attributes. The difference between them is provided by different values of these attributes;
- when comparing the same attributes of situations X and Y the concept of distance in general case is not introduced, but a conclusion about the equality-inequality of their values can be made due to belonging to the same subset (or subrange) of values $d \in D^k_i$: $x \in d, y \in d \Rightarrow x = y$, $x \in d, y \notin d \Rightarrow x \neq y$.
- each situation of the ST has its decision. The decision of situation-prototype is applicable for all subtypes; but it differs from them in less detail and a higher level of abstraction. Thus, each point of the ST corresponds not only to its own situation, but in general, to couple P = <Sit, R>.

Graphically the ST is built up as a graph of dependence ρ on the set of possible situations (Figure 1):

$$G_{Sit}(\{S^k_j\}, V), \quad (2)$$

where $\{S^k_i\} \Leftrightarrow \{Sit^k_i\}$ is a set of points one-to-one corresponding to a set of situations; j = 1, 2, ..., N^k is the index of the number of a situation in the k-th level, N^k is the number of selected situations on the k-th level.

V is a set of arcs with $v(S^k_j, S^{k+1}_j) \in V$, if dependence $\rho(Sit^k_j, Sit^{k+1}_j)$ is realized, and there is no arc from the lower level to the upper one, as well as between the elements of the same level.

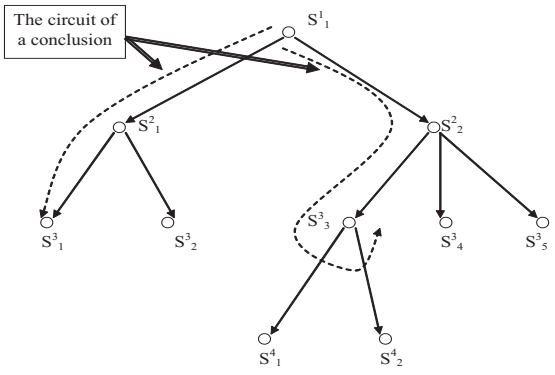


Figure 1. Situations tree

During the operation of the CBR-system the movement is from the root of the tree downwards. Since each level contains its set of attributes, this determines what questions should be asked to identify the current situation Sit_{cur} at the given k-th level. The closest $Sit^{k*} \in \{Sit^k_i\}$ is selected according to the determined attribute values Sit_{cur}, which determines the further path of movement along the tree. Reasoning ends with the end point (example in Figure 1 – point S31). Since a case with its decision corresponds to this point, it is applied as desired.

The arrangement of a ST allows reasoning in conditions of lack of information about the values of certain attributes taking into account the above conditions, when each point corresponds to its decision. The example in Figure 1 is reasoning S^3_3, which ends the algorithm of searching for a decision with unknown values of attributes present at the 4th level of the description of the situations, which does not allow recognizing Sit_{cur} as S^4_1 or S^4_2.

III. Object-Oriented Model

For further development of an object-oriented model of CBR-system and software implementation the following problems have been solved in the article:

- representation of a situations tree in terms of classes and objects;
- definition of the method of storage of situations trees in the database.

In solution of these tasks the results of the works devoted to representation of the systems based on knowledge in the form of object models [19-22] were used

For software implementation of the situations base in terms of object-oriented approach a common class ("layout") was created which allows describing any case, represented in the system, as a class instance. In this class not only its attributes are stored, but the reference to a "parent" case that allows creating a situations tree.

To store the case, class Case is created. Such attributes as the Name of the case, the Reference to the parent case, the Reference to the solution, and the List of attributes with values are implemented in this class. Among the methods of the class a method of Equivalency checking is implemented, which allows determining whether the situation corresponds to that under consideration.

The list of attributes is an arbitrary number of couples Attribute-Value. Each such couple is represented as a class instance of the Attribute case. This class implements not only the fields and properties that allow storing an Attribute and its Value, but also the method that correctly compares the current attribute value with the compared value for compliance.

Another class used to describe the situations tree is the Solution. The Instances of this class store possible variants of the solutions to the situations that are referenced by the instances of Case class. The main characteristics of a class instance are the Name, the result Type, and the result Text.

The implementation of a situations tree using object approach has simplified the process of searching for a similar case and an appropriate solution. Now for this purpose the methods implemented in class Case are used, including Equivalency checking. As a result, search for a case in the tree

corresponding to the desired situation is reduced to alternate method call starting from the tree root and further on.

To store cases and solutions a knowledge base is implemented, which contains all the required information about the situations tree in terms of object-oriented approach. The physical basis for the situational knowledge base is a database that stores cases with the list of values of their attributes, as well as the solutions reached in these situations.

The database contains four main entities, shown in figure 2.

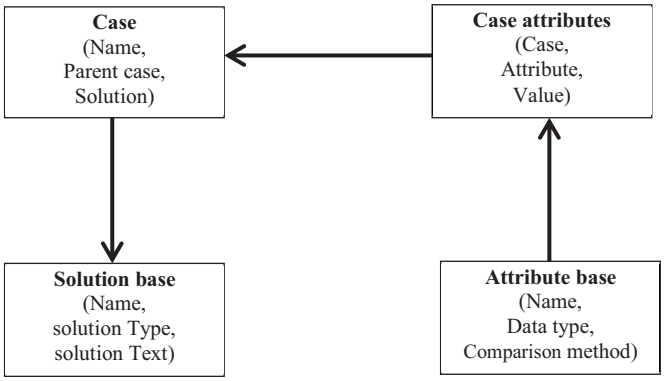


Figure 2. Database layout

First, it is Cases. Each case contains both its own name, and a reference to the parent case. This approach allows representing all the cases in a tree form where each case of the next level adds new attributes to its "parent" cases. Case-root of the tree has no reference to its parent.

Each case can have a reference to a Solution. All solutions are stored in the entity of Solution base. Each solution is represented in one of possible forms, for example in the form of text or program code.

To store case attributes and their values, another entity - Case Attributes - is designed. Each record of this entity stores a Reference to the case, a Reference to an attribute, Field to store the value of an arbitrary data type and an additional field to store the Second value in case the attribute value is specified by the range.

The very attributes are stored in a centralized entity of Attributes Base. It contains a description of each attribute: The Type of its values, and the Equivalency checking method: strict equality, larger or less than a predetermined value, entering the interval, etc.

IV. Results And Conclusions

Based on the object-oriented paradigm the proposed approach allows creating domain-independent CBR systems, which are shells of these systems, filled with knowledge about the cases for solving specific problems.

One of the important advantages of this approach is the possibility of developing basic case libraries for a wide class of problems in a certain subject area. The mechanisms of inheritance and overriding, typical for object-oriented programming, with the use of these libraries allows developers without knowledge implementing further configuration of their own system which takes into account specific features of certain problems.

Thus, there is a new class of systems-shells, based on knowledge, where knowledge bases are not empty, but contain knowledge generally valid for this or that subject area. They can be used by means of refining and addition when creating their advising or recommender systems, in which the situations are redefined due to new attributes, take into account local characteristics and contexts for decision-making in specific circumstances.

V. References

[1] J.L. Kolodner “An introduction to Case-Based Reasoning. Artificial Intelligence Review”, num. 6, 1992, pp.3-34.

[2] I.D.Watson, F.Marir “Case-based reasoning: A review. The Knowledge Engineering Review”, vol. 9, No. 4, 1994, pp. 355-381

[3] H.J. Gómez-Vallejo, B. Uriel-Latorre, M. Sande-Meijide, B Villamarín-Bello., R. Pavón, F. Fdez-Riverola, D. Glez-Peña “A case-based reasoning system for aiding detection and classification of nosocomial infections. Decision Support Systems”, vol. 84, April 2016, pp. 104-116.

[4] P.R. Varshavskiy “Mekhanizmy pravdopodobnykh rassuzhdeniy na osnove pretsedentov (nakoplennogo opyta) dlya sistem ekspertnoy diagnostiki”, 11-ya natsionalnaya konferentsiya KII-2008: Trudy konferentsii. V 3-kh t., T. 2, M: Lenand, 2008, s. 321 — 329.

[5] N. Naderpajouh, A. Afshar A case-based reasoning approach to application of value engineering methodology in the construction industry. Construction Management and Economics, № 26(4), 2008, pp.363–372.

[6] M.M. Mohamed, M.A. Rasha “A case-based reasoning approach for estimating the costs of pump station projects”, Journal of Advanced Research, vol. 2, 2011, pp. 289–295.

[7] Ji Xia, Gang Chen, Peng Tan, and Cheng Zhang. “An online case-based reasoning system for coal blends combustion optimization of thermal power plant”, International Journal of Electrical Power & Energy Systems, vol. 62, 2014, pp. 299-311.

[8] Rodrigo G.C. Rocha, Ryan R. Azevedo, Ygor Cesar Sousa, Eduardo de A. Tavares, and Silvio Meira. A case-based reasoning system to support the global software development. 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, KES2014. Procedia Computer Science, №35, 2014, pp. 194 – 202.

[9] M.T. Elhadi Bankruptcy support system: taking advantage of information retrieval and case-based reasoning. Expert Systems with Applications, vol.18, issue 3, 2000, pp. 215-219.

[10] Kyong Joo Oh, and Tae Yoon Kim. Financial market monitoring by case-based reasoning. Expert Systems with Applications, vol.18, issue 3, 2007, pp. 789-800.

[11] I. N. Glukhikh Adaptation in the recommendation systems with case based reasoning. In Advanced Studies in Science: Theory and Practice (Vol. 5). London, UK: The Global partnership LLC. doi: 10.17809/0557468427-43

[12] I.N. Gluhih, V.N. Piankov, A.R. Zabolotnov Situational models in corporative knowledge base on geological-technical measures' know-how. Neftyanoe khozyaystvo - Oil Industry, №6, 2002, pp. 45.

[13] V.I.Vasiliev, V.A. Pestrikov, A.S. Krasko Intellektualnaya podderzhka prinyatiya resheniya v ekstremalnykh situatsiyakh na osnove vyvoda po pretsedentam. Izvestiya UFU. Tekhnichesliye nauki, 2008. Vyp.8, s.5-13.

[14] P.R. Varshavskiy, A.P. Yeremeyev Modelirovaniye rassuzhdeniy na osnove pretsedentov v intellektualnykh sistemalh prinyatiya resheniy. Iskusstvenny intellect I prinyatiye resheniy, №2, 2009, s. 45-57.

[15] D.A. Pospelov Situatsionnoye upravleniye. Teoriya I praktika. M.: Nauka, 1986, 288 s.

[16] Susan Craw, Nirmalie Wiratunga, Ray C. Rowe. Learning adaptation knowledge to improve case-based reasoning. Artificial Intelligence, vol. 170, 2006, pp. 1175–1192.

[17] I. Gluhih Predstavlenie znanij i vyvod reshenij v situacionnyh bazah znanij//Vestnik Tjumenskogo gosudarstvennogo universiteta. – 2006. - № 5. – P. 265-270.

[18] I. Gluhih Intellectualnye informacionnye systemy. – Moscow, Academy, 2010, 112p.

[19] R. Bergmann, A. Stahl Similarity measures for object-oriented case representations. In Proceedings of the Fourth European Workshop on Case-Based Reasoning. - Berlin: Springer, 1998, pp.25-36.

[20] Bin He, Zhongqiang Deng and Haifeng Lv. Object-oriented Knowledge Modelling for Conceptual Design of Mechanisms. International Journal of Database Theory and Application, vol.6, No.5, 2013, pp.67-84.

[21] Dattatraya Vishnu Kodavade, and Shaila Dinakar Apte. A Universal Object Oriented Expert System Frame Work for Fault Diagnosis. International Journal of Intelligence Science, № 2, 2012, pp. 63-70.

[22] Nabil Mohammed Ali Munassar, A. Govardhan. Comparison between Traditional Approach and Object-Oriented Approach in Software Engineering Development. International Journal of Advanced Computer Science and Applications, vol. 2, No. 6, 2011, pp. 70-76.

Comparative Analysis of Feature Selection Algorithms in Construction of Fuzzy Classifiers

I.V. Gorbunov, S.R. Subhankulova, I. A. Hodashinsky
Department of Complex Information Security
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
giv@keva.tusur.ru, sophi1059@gmail.com, hodashn@rambler.ru

A. E. Yankovskaya
Tomsk State University of Architecture and Building, National Research Tomsk State University, National Institute Tomsk Polytechnic University, Tomsk State University of Control Systems and Radioelectronics,
Tomsk, Russia
ayyankov@gmail.com

***Abstract*— The goal of the paper is the analysis of the fuzzy classifiers effectiveness, which are built by different algorithms of feature selection according to wrapper algorithms. The search of the informative features is provided on basis of the greedy algorithm (*GrA*), the discrete genetic algorithm (*GA*), the discrete mine blast algorithm (*MBA*$_D$).**

***Index Terms* — Feature selection, fuzzy classifiers, greedy algorithm, discrete genetic algorithm, discrete mine blast algorithm.**

I. Introduction

The stage of feature selection is the first stage in the building of a classificator however not all the features are necessary for the problem to solve. Some of the original features may be irrelevant and less informative, impairing accuracy of the qualification and increasing computational complexity of the classificator.

The mainframe algorithm of the feature selection does not exist. There are the two basic approaches to the solution of the problem. The first approach suggests the inclusion of the classificator in process into the process of feature selection, which is known as the wrapper approach. The second one doesn't depend on the classification algorithm, it is known as the filter approach [1].

The goal of the paper is the analysis of the fuzzy classifiers effectiveness, which are built by different algorithms of feature selection according to wrapper algorithms.

II. Terms of Reference

There is the summary shape is given $\{(\mathbf{x}_p, c_p)| p = 1, 2, \ldots, N\}$, where each object (examination) is defined by the feature vector (of the input values) – $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and the class mark, which the object belongs to. The binary vector $\mathbf{f} = (f_1, f_2, \ldots, f_n)$, where $f_i = 0$ means that i-th feature does not take part in the classification, is put in accordance with the feature vector x.

The fuzzy classificator is defined by the rules of the following type:

R_{ji} : IF $x_1=A_{ji1}$ AND $x_2=A_{ji2}$ AND $x_3=A_{ji3}$ AND … AND $x_n=A_{jin}$ THEN class=c_j,

where c_j is the identificator of the j-th class, $j \in [1, cl]$, A_{jik} is a fuzzy term, characterizing k-th feature in the ji-th rule Rji, i =1,2, …b $|Rj|$], , Rj is the array of the rules, which are taking the examination to the class with the c_j classificator.

In the process of the fuzzy classification the object refers to each class with the definite degree of belonging, calculated as in the following:

$$\beta_j(\mathbf{x}) = \sum_{R_{ji}} \prod_{k=1}^{n} f_i \cdot A_{jik}(x_k), \; j = 1,2,\ldots,cl.$$

The class mark is the output vertex of the classificator, which is found as in the following:

$$\text{class} = c_{j^*}, \; j^* = \arg\max_{1 \le j \le cl} \beta_j .$$

The fuzzy classificator can be represented by the function $c = \varphi(\mathbf{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector, defining the rule base.

On the given table of examination, we identify the unit function

$$delta(p, \boldsymbol{\theta}, \mathbf{f}) = \begin{cases} 1, & \text{if } c_p = \varphi(c_p, \boldsymbol{\theta}, \mathbf{f}) \\ 0, & \text{else} \end{cases}, p = 1,2,\ldots,N$$

,

then the criterion of the classification evaluates in the following way:

$$E(\boldsymbol{\theta}, \mathbf{f}) = \frac{\sum_{p=1}^{N} delta(p, \boldsymbol{\theta}, \mathbf{f})}{N} .$$

The goal of the building of the fuzzy classificator is the search of such parameters of the classificator $\boldsymbol{\theta}$ and such characteristics of $\mathbf{f}$ which lead the value of the criterion $E(\boldsymbol{\theta}, \mathbf{f})$ to the maximum.

The optimization of the parameters of the classificator θ is provided by means of the continuous mine blast algorithm. The search of the informative features is provided on basis of

the greedy algorithm, the discrete genetic algorithm, the discrete mine blast algorithm, the discrete bee algorithm.

III. The Mine Blast Algorithm

The Mine blast algorithm — MBA [2]) is based on the examination of a mine explosion on a mine field. Flying debris fall on the other mines, which provokes the chain of blasts. Each debris is characterized by the direction and distance between the collision with another mine. Minefield clearance is supposed its complete blasting. It requires to find a mine with the maximum blast effect, which is situated in the optimum point **θ**.

The continuous mine blast algorithm foe the parameters of fuzzy classificator optimization (MBA_R) is presented below [3].

Inputs. N_s is the number of the debris for each mine; the maximum number of iterations.

Output. **θ*** is the vector of the optimum parameters of the fuzzy classificator.

1. To set the start point of the blast θ0. To calculate the angle of debris dispersion φ=360/Ns. To calculate the distance of the debris flight:

$$\mathbf{d_0}=\mathbf{UB}-\mathbf{LB},$$

where **LB** is a vector formed of corresponding minimum value of the $\boldsymbol{\theta_0}$ parameters, **UB** is a vector formed of the corresponding maximum values of the $\boldsymbol{\theta_0}$ parameters.

2. To calculate the following distance:

$$\mathbf{d}_n = \mathbf{d}_{n-1} * randn,\ n = 1,2,\ldots,N_s,$$

where *randn* is a number distributed by a normal law

To calculate the coordinates of a blasted mine:

$$\boldsymbol{\theta}_{e(n)} = \mathbf{d}_n * randn.$$

3. To create debris, i.e. to form a new vector:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{e(n)} + \boldsymbol{\theta}_{n-1}.$$

4. To check satisfiability of limits. If the limits aren't satisfied, go on the 2nd step.

5. To calculate the value of the performance function of the solution $\boldsymbol{\theta}_n$.

6. The best intermediate solution is replaced by the solution obtained, if the performance function of the solution will be better than the previous one.

7. If the maximum number of iterations is not reached, to go to the 2nd step or exit from the algorithm.

IV. Mine Blast Discrete Algorithm

On the initial stage, the parameters of the algorithm are initialized: $\mathbf{X}_0$ is the starting point of the blast (feature vector of the object being classified), the number of debris N_s, which fly apart after the mine blast and the number of iterations *maxiter*. Every component of the initial vector equal (1) (it is considered that all features are informative). Further, the angle of variations φ is calculated, the flight of the debris r, the components of the blasted mines and the components of new debris.

Further, it is defined if the feature is an informative one. Among the vectors received it is defined the best workaround with the least classification error value. Further, the process repeats internationally until the number of iterations reaches the set point.

The pseudocode of the algorithm MBA_D is given below.

Inputs: $\mathbf{X}_0, N_s$, *maxiter*.
Output: value: $\mathbf{X}_{best}$.
$\mathbf{X}[0] := (x_1, x_2, \ldots, x_n)$; *iter*:=0;
while (*iter* < *maxiter*)
  $\mathbf{r}[0] := \text{rand}(\mathbf{max}, \mathbf{min})$;
  $\varphi := 360 / N_s$;
  for *i* from 1 to N_s
    $\mathbf{r}[i] := \mathbf{r}[i-1] * randn$;
    $\mathbf{X_e}[i] := \mathbf{r}[i] * \cos(\varphi)$;
    $\mathbf{X}[i] := 1/(1+\exp(-\mathbf{X_e}[i]))$;
    for *j* from 1 to *n*
      if (rand(0,1) < *x*[*j*])
      then *x*[*j*]:= 1;
      else *x*[*j*]:= 0;
    end of the for
  end of the for
    if (*E* (**X**[*i*])<E($\mathbf{X}_{best}$))
    then $\mathbf{X}_{best}$:= **X**[*i*];
    iter := *iter* + 1;
end of the while
out $\mathbf{X}_{best}$:= Search_best(**X**[*i*]).

V. Experiments

The experiments have been performed by considering 11 real-world datasets (http://www.keel.es). To carry the different experiments out, a five-fold cross-validation model has been applied.

During the experiment at first, the algorithms of feature selection were launched (GrA, GA, MBA_D, GrA + MBA_D). The sets of selected features were the result of the work of the algorithms. Further, the ranging was provided and one feature set was selected to work with the continuous algorithm MBA_R. In the table 1 the following descriptions are given: 1) the characteristics of the data sets; 2) the value of the per cent of the correct classification in training (tra) and test (tst) sets, which were get as a result of fuzzy classificators layout by the algorithms mentioned above; 3) the list of non-informative features modified by the algorithms GrA, GA, MBA_D, for each data set; 4) the number of the rules for each data set classification.

VI. Conclusion

In this paper, we compare feature selection techniques and assess the effect of selected feature on learning and predicting ability. Four feature selection techniques are used to evaluate their effectiveness using fuzzy classifier.

The benchmarking analysis of the four given methods of the construction of fuzzy classificators showed their quite high efficiency and possibility to use them for practical problems solution.

TABLE I. COMPARISON OF RESULTS OBTAINED BY APPLYING METHODS

Datasets			monk-2	glass	wiscon sin	wine	cleveland	vehicle	bands	sonar	pima	new-thyroid	iris
Number of features			6	9	9	13	13	18	19	60	8	5	4
Number of instances			432	214	683	178	297	846	365	208	768	215	150
Number of classes			2	7	2	3	5	4	2	2	2	3	3
			% correct classified										
Algorithms	MBA_R	*tra*	92,67	70,09	97,02	98,73	60,65	49,44	70,13	78,46	78,99	98,65	97,77
		tst	92,3	62,71	95,52	94,59	54,99	45,84	65,3	65,57	74,87	95,81	94,4
	MBA_D + MBA_R	*tra*	92,05	69,76	96,6	99,52	61,23	50,56	73,18	76,92	78,81	98,97	97,9
		tst	91,2	63,12	94,7	96,83	55,55	47,39	66,96	66,18	74,66	96,74	95,33
	GrA +MBA_R	*tra*	52,78	70,12	96,41	99,05	60,81	54,1	73,16	80,29	78,58	98,09	-
		tst	52,78	60,18	94,97	95,61	57,54	49,72	69,13	68,01	75,81	95,81	-
	GA + MBA_R	*tra*	90,91	69,74	96,38	99,38	60,9	53,32	71,85	79,62	78,39	99,02	97,97
		tst	89,85	61,48	94,98	96,17	54,58	50,05	67,02	70,32	76,63	96,93	96,13
	GrA + MBA_D + MBA_R	*tra*	-	69,65	-	99,18	60,74	-	72,45	66,56	-	-	-
		tst	-	61,65	-	95,16	55,42	-	68,3	64,88	-	-	-
			Non-informative features										
Discrete algorithms	GrA		1, 3-6	1, 5, 6, 9	2, 4-6, 8, 9	2, 3, 5, 6, 8, 10, 12	5,7,11, 12	2-9, 11-14, 16	3, 7, 11-13, 15, 17-19	1-5, 9-12, 14, 15, 17-27, 29, 30, 32-35, 37-42, 44, 53-60	3, 4	3, 4	-
	MBA_D		1,3,4,6	8,9	6	6, 8, 9	2, 6, 7, 9	6, 16	3, 6, 12-15	1-16, 18-35, 37-60	3, 4	4	2
	GA		1, 3, 4, 6	5, 8, 9	4, 6, 9	2, 3, 5, 6, 8-10	2,5,6,7, 9,12	1, 3-8, 11-14, 17	7, 8, 13-15, 18, 19	2-4, 9, 12, 13, 19, 22-25, 27, 28, 33, 34, 37, 38, 41, 42, 47, 49, 51-53, 55-57	3-5, 8	3, 4	1, 2
Number of rules			2	7	2	3	5	4	2	2	2	3	3

ACKNOWLEDGMENT

We appreciate the support of the Russian Foundation for Basic Research (16-07-00034, 16-37-60051, 16-07-00859, 16-47-700350 r_a) for this work.

REFERENCES

[1] V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzos, Feature Selection for High-Dimensional Data, Springer, 2015. 163 p.

[2] A. Sadollah, A. Bahreininejad, H. Eskandar, M.Hamdi, "Mine blast algorithm: A new population based algorithm for solving constrained engineering optimization problems," Applied Soft Computing. 2013. Vol. 13. pp. 2592-2612.

[3] I.A. Hodashinsky, S.R. Subhankulova, "Parameters identification of fuzzy systems based on the mine blast algorithm»," Information Science and Control Systems. 2015. Vol. 2. pp. 89-98. (In Russian)

Ordered Multidimensional Model Construction of Relational Source for Integral OLAP-modeling

Anna Korobko, Ludmila Nozhenkova
Department of Applied Computer Science
Institute of Computational Modelling of SB RAS Name
Krasnoyarsk, Russia
lynx@icm.krasn.ru, expert@icm.krasn.ru

***Abstract*—Internet is a rich source of information, but it consists of miscellaneous data fragments. Analytical "surfing" on World Wide Web opens extraordinary prospects for attain and maintain competitive advantage of enterprises and for scientific researches. Exploratory OLAP is one of the main agenda of analytical processing of heterogeneous data sources. The author proposes an original approach to exploratory OLAP based on virtual data consolidation (without physical gathering) based on conceptual analytic model in multidimensional terms. The algorithm of discovering multidimensional concepts in normalized relational data sources based on functional dependencies analysis is represented.**

Exploratory OLAP; heterogeneous data; conceptual analytical model; integral OLAP-model

I. Introduction

Proposed by dr. Codd rules of OLAP technology [1] define requirements to ideal software (tool) for on-line analytical processing. These demands are just a guideline and existed products comply with the requirements in order to modern computer science possibilities. Some of them are embodied at a first approximation only [2]. For example, OLAP tool should provide data accessibility for the user. It is understood that the tool should wield data from heterogeneous sources in order to perform any conversions necessary to present a coherent view. It is only in recent times that Exploratory OLAP is a difficult but crucial task for further development of the on-line analytical processing. Great variety of open accessed informational resources and modern business challenges provide significant interest to developing OLAP tools with respect to interactive data exploration from the different miscellaneous sources. Enterprises seek ways to enrich body of analytical processing (internal data) by integrating open-access collateral information to attain and maintain competitive advantage.

The key demand of OLAP technology is mostly multidimensional conceptual view. This rule has given rise to creating multidimensional data model (MDM) and data warehouse ideology. The multidimensional model intuitively represents data under the metaphor of the cube whose cells store qualitative characteristics of business domain. Fairly, such approach allows us to represent data in a user friendly way, but considering internal data only analytics encounter large sparse cubes. These cubes are difficult-to-process, difficult-to-understand and knowledge mining from such plenty of data is impossible at all.

Nowadays, due to explosive growth in developing ergonomic and native software the crucial demand to application is user-friendly (human-centric) interface complementary to functional requirements. The computer capacity and results of fundamental research allow to create systems with due attention to human cognitive ability and learning skills. Computer science development is aimed to latent support and to perform routine task automatically, allowing user to create.

Developing methods of heterogeneous data integration on conceptual level (without physical loading) is a topic issue to interactive data exploration. So, the OLAP tool (not the end-user) should be concerned with where the physical data comes from. The keystone of this research field is generating correspondences between relevant elements (concepts) of diverse datasets (relational database, XML documents, Web sites etc.) within an integral model.

Using authority method of integral OLAP-modeling [3] provides essential multidimensional conceptual view of information, user-friendly governing of data and heterogeneous data consolidation. To implement the proposed method algorithms of automatic conceptual multidimensional model construction should be infer for each integrated data source (regardless of the origin).

The paper is structured as follows. Section II describes basic concepts of the authority method of integral OLAP-modeling. Section III presents the original approach to exploratory OLAP based on virtual data consolidation (without physical gathering) based on integral OLAP-model. In section IV the algorithm of discovering multidimensional concepts in normalized relational data sources based on functional dependencies analysis is represented. Section 5 illustrates implementation of proposed algorithm for certain example.

II. The integral OLAP-model

The large sparse cubes are common case either for internal data (closed data set) or extended information space. But it's more crucial for assembled data from the different miscellaneous sources. Joining several data sources into the comprehensive analytical model, we mostly get sparse cube.

The method of Formal Conceptual Analysis (FCA) [4, 5] can help us to split sparse cube into a set of compact (solid, consistent) cubes. What is more, in accordance with theory of FCA the set of compact cubes can be ordered by partial ordering relation, i.e. they form a lattice.

Author have proposed the method of constructing integral OLAP-model of the domain as a lattice of multidimensional cube-concepts based on integration of multidimensional model and Formal Concept Analysis. The integral OLAP-model includes all possible combinations of facts and dimensions of the domain and allows the decision maker to discover unexpected analytical patterns in a set of all analyzed concepts (facts and dimensions).

The integral analytical model of the domain is a result of Formal Concept Analysis of multidimensional concepts. A set of analyzed concepts includes a set of cube facts (measures) F = {f1, f2, …, fm} and a set of dimensions D = {d1, d2, …, dn}. The numeric characteristics are a set of facts and the analysis aspects are a set of dimensions.

Formal context K is defined as a triplet (F, D, R) that consists of the set F, the set D and the relation $R \subseteq F \times D$. The elements of set F are the facts; the elements of set D are the dimensions. The relation of comparability R between F and D (described by fRd) indicates when fact $f \in F$ can be processed with dimension $d \in D$. The formal context can be represented as a binary matrix where the facts are the rows, the dimensions are the columns and relation of comparability is a cross at intersection between a row and a column.

Formal cube-concepts are formed based on the formal context K. For set $A \subseteq F$ and for set $B \subseteq D$ it is defined that: $A' = \{d \in D \mid fRd \text{ for all } f \in A\}$ (all dimensions in D shared by the facts of A); $B' = \{f \in F \mid fRd \text{ for all } d \in B\}$ (all facts in F can be processed with dimensions of B). The formal cube-concept is defined by derivation operators as pair (A, B) with $A \subseteq F$, $B \subseteq D$, $A = B'$, $B = A'$. It means that A is a set of equidimensional facts, which are processed with all dimensions of B. The set of facts A forms the extent and the set of dimensions B form the intent of the formal cube-concept (A, B). The formal cube-concept is an analytical multidimensional cube which is complete with respect to addition of the equidimensional facts and the compatible dimensions.

The set of all formal cube-concepts is ordered by the subcube-supercube relation. For two cube-concepts (A1, B1) and (A2, B2) this order is identified as: $(A1, B1) \leq (A2, B2)$ when $A1 \subseteq A2$ and $B2 \subseteq B1$. It means that the extent of a parent cube includes the extent of a child cube and the intent of the child cube includes the intent of the parent cube. In this case, (A1, B1) is called a subcube and (A2, B2) is called a supercube. The set of all concepts together with the subcube-supercube relations forms a conceptual lattice L of the multidimensional cubes.

The conceptual lattice of the multidimensional cube-concepts is an application of the integral OLAP-model of the domain. The integral OLAP-model covers all possible analytical queries of the domain. The features of cube-concept lattice provide the manipulation of all analyzed objects intuitively and support of the analytical experiment.

III. Exploratory OLAP based on the integral OLAP-model

The most of proprietary OLAP tools allow user to connect to diverse miscellaneous sources and to create OLAP cubes regardless of the data origin. There are two scenarios: (i) creating a catalog of data sources without complementary accesses. (ii) implementing special ETL (Extract Transformation Load) procedures for each resource to integrate one into data warehouse. Originality of proposed way of the exploratory OLAP is to provide on-fly "surfing" on a union of the assessable data without extracting, transformation and loading. The suggesting approach is not a substitution for the data warehouse. The new paradigm doesn't deal with preprocessed consistent data but leans on current snapshots of integrated data sources. It can be used for rapid hypothesis testing, corroboration brilliant idea and support of analytical investigations. For example, it would be interesting to match your enterprise performance indicators and published information about competitor or to validate whether or not staff efficiency is conditional on weather. The information which is interesting for user can be loaded into data warehouse for further considerable analytical processing.

The integral OLAP model fits graceful for an application of the exploratory OLAP. The integral model provides multidimensional data representation. On the one hand, it is close to the way of man thinking; on the other hand, it supports performance improvement as its simple structure allows us to support adaptive manipulation.

Various solutions have been proposed in theory and practice for discovering OLAP dimensions and facts in structured and semi-structured data. A survey of research achievements on delivering of multidimensional conceptual schemas from relational sources can be found in [6]. At the same paper the novel method of multidimensional design based on end-user requirements analysis is proposed. According to Winter et al. [7] theoretical approaches in this field can be classified in supply-driven [8, 13], demand-driven [9] and hybrid methods [10]. However, most of them must be carried out manually and only a few automate the process. A semi-automated methodology for multidimensional design of XML sources is proposed in [11]. In such areas as Social Network Analysis content-driven discovery of measurable facts and dimensional characteristics is based on applying data mining and other techniques to semi-structured data [12]. The existing approaches may be said to correspond to their goals and can be feasible combined with the author proposition for efficient multidimensional schema discovering.

Great variety of researches in this field and its successful tests argue that the multidimensional data model is a good foundation for constructing comprehensive OLAP-model for a jointed set of heterogeneous sources. The same problems are solving for constructing domain data warehouse that accumulate data from various operational systems. In that case process of additional data source assembling lean on ETL procedures which are developed manually. And it is the best way to get consistent well-matched data for deep purposeful analysis. However, it doesn't look like an on-line analytical

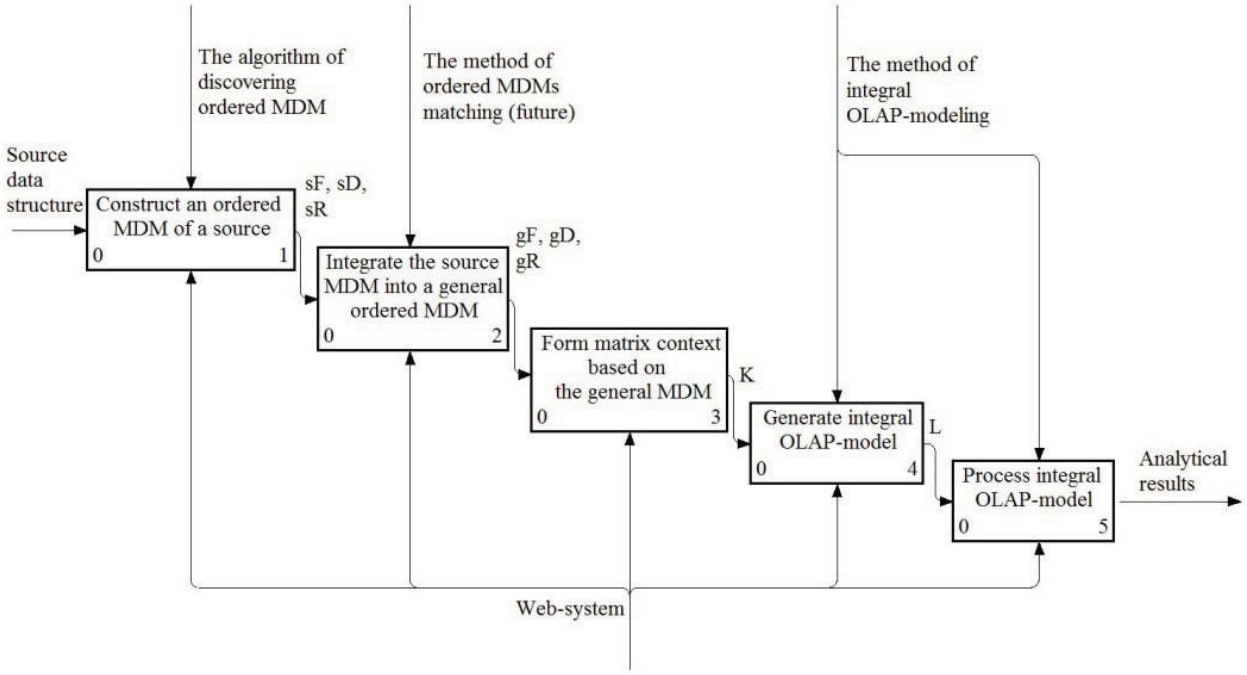


Fig. 1. IDEF0 diagram of exploratory OLAP based on the integral OLAP-model

surfing as well as it leads to significant time consumption and demand involving highly technically trained professionals. The novel approach to automated hot insertion of heterogeneous open-accessed sources in an existed integral OLAP-model in order to avoid drawbacks of hard ETL procedures needs to be developed. Main activities of exploratory OLAP based on the integral OLAP-model are represented on Fig.1.

On the assumption of that we already have an integral OLAP-model let's consider process of integrating a new data source into the one. At first the data source structure should be analyzed to define the multidimensional model concepts – facts (sF) and dimensions (sD) – paying attentions to the relation of comparability sR between sF and sD. The second activity "Integrate the source MDM into a general ordered MDM" is one of the most demanding because of the fact that it implies structural and semantic matching of ordered multidimensional models of sources. The output of second activity is the united sets of facts gF, dimensions gD and ordering relation gR. The next issue is to transform a general MDM into binary matrix view – the incident matrix of multidimensional model graph. The output of the activity "Form matrix context based on the general MDM" is the context K which can be used for the integral OLAP-model constructing (process 4 on Fig.1). Getting a cube-concepts lattice we can go forward to analytical decision making support based on integral OLAP-model of gathered heterogeneous data sources.

It presents a number of severe science-driven difficulties:

- discovering particular ordered multidimensional model from the heterogeneous data sources according to ubiquitous terms (concepts);
- particular ordered multidimensional models matching – search of dimensions describing the same analytical aspects;
- discovering hierarchical chains of dimensions;
- user-friendly governing and querying support to treat the conceptual integral OLAP-model;
- executing integral analytical query – dividing it into particular sub-queries, sub-query execution and joining sub-queries execution results.

According to the represented diagram it's better to start from development of ubiquitous standards of the ordered multidimensional model constructing depending on data format. There is provided an investigation (studying) of the normalized relational database case.

IV. The Algorithm of Ordered Multidimensional Model Construction for a Normalized Relational Data Base

The proposed algorithm is intended to prove the authority approach to discovering the ordered multidimensional model and the way of constructing graph of MDM concepts. A normalized relational data base is subject of current research. To illustrate the algorithm execution we consider an example. Figure 2 represents studying relational data base schema. There are no partial dependencies on complex key and transitive dependencies in the normalized source.

Regarding the set of functional dependencies as the multi graph, the attributes of data base can be treated as graph nodes and functional dependency relation is represented as edges. Figure 3 illustrates the functional dependencies graph of studying relational data base schema. Each node of the graph

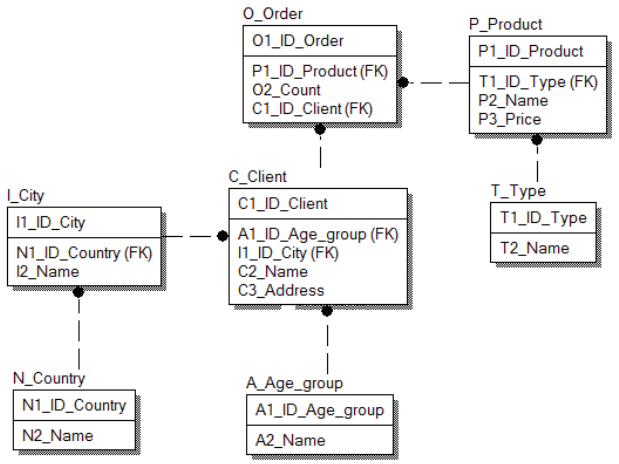


Fig. 2. The relational data base schema

corresponds to schema attribute (A1 – ID Age group, A2 – Age group name, etc.) and an edge corresponds to a functional dependency (for example, O1 functionally determines O2 C1 P1). It means that the order identifier uniquely determines the client, product and count. It should be emphasized that the grey circles are the numeric attributes of relational schema tables.

In this section the algorithm of discovering ordered multidimensional model in a normalized relational data base is detailed in pseudo-cod. Briefly, the algorithm is aimed to produce the multidimensional model based on a FD graph (Fig. 3) in such a way that new nodes represent the multidimensional concepts (i.e. dimensions and facts) and the edges are interpreted as instances of the relation of comparability according to terms of the integral OLAP-model. Function algorithm() is the main function of the algorithm and it is called first. The algorithm has two global variables: F – a set of facts of the sought multidimensional model, R – the relation of comparability between facts and dimensions what is needed for integral OLAP-model construction in order to the authority method. A lifetime of global variables is equal to the algorithm execution time. The main function of the algorithm has one input – a set of considered nodes V of the functional dependencies graph. As output, the function produces D, F

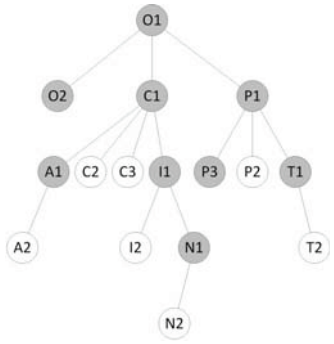


Fig. 3. The functional dependencies multi graph of the relational data base schema

```
global: R, F

function algorithm()
input: V
output: D, F, R
begin
   D=∅
   for each v from V do
   if getParent(v) is empty do
      D=D ∪ getDimensions(v);
   endfor
   return (D, F, R);
end
```

and R – the ordered multidimensional model. In the main function each node of the FD graph without parents is analyzed first (if getParent(v) is empty). It means that we consider the data base primary keys which are not used as foreign keys. Thus, the tables with potential facts are detected. In our example we have only one such node – it's O1.

The intermediary recursive function getDimensions() is called for the detected top nodes of the functional dependencies graph. It is represented in pseudo code. The function has one input – the analyzing node v∈V. As output, the function

```
function getDimensions()
input: v
output: D
begin
   D=∅
   sV = getSibling(v);
   for each x from sV do
      if getChild(x) is not empty do
         D=D ∪ getDimensions(getChild(x));
      else if typeof(v) is not numeric do
         D=D ∪ {x};
      endif
   endfor
   for each x from sV do
      if (typeof(x) is numeric) and
      (getChild(x) is not empty) do
         f = count(x);
         F=F ∪ f;
         for each d from D do R=R ∪ (f, d);
      else if (typeof(x) is numeric) and
      (getChild(x) is empty) do
         f = all(x);
         F=F ∪ f;
         for each d from D do R=R ∪ (f, d);
      endif
   endfor
   return D;
end
```

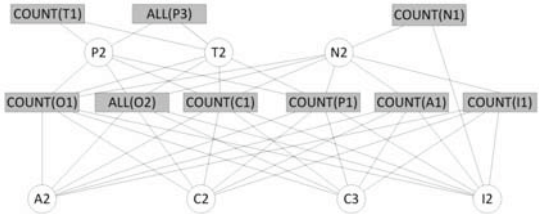


Fig. 3. The ordered multidimensional model of the studying relational schema

produces D – a set of dimensions of future ordered MDM to provide recursion as well as the intermediary function. The mentioned function getSibling() returns a set of sibling nodes which have the same parent as the input node. It is needed to recognize neighbour facts, dimensions and to define comparability relation between them. It is of importance that according to FD graph features the defined fact is analytically comparable with all dimensions found in a lower level of itself and sibling's children. The function getChild()returns one of the input node children.

The foreign keys give us opportunity to dive into the graph depth to derive more multidimensional concepts from the relational data base. That is the case of the recursive function calling. Nonnumeric attributes of the table creates the dimensions which are added to the set D.

The result of algorithm execution for studying case is represented in Fig. 4. There are 2 types of nodes: a circle is a dimension and a rectangle is a fact. An edge between multidimensional concepts means they can be processed together in analytical query. The incident matrix of the ordered multidimensional model is interpreted as the formal context K from the section II. As you can see use of the ordered multidimensional model for constructing the integral OLAP-model and exploratory analysis support gives us a great variety of analytical combinations based on structural features of the data source.

V. Conclusion

In summary, proposed approach to exploratory analytical processing of heterogeneous data sources has a number of benefits: (i) it doesn't require loading and physical assembling of data. (ii) it is based on human centric approach to provide intuitively manipulation of conceptual integral multidimensional model. (iii) it intends to be automatic.

The paper should be considered as a first attempt to develop theoretical foundation of the exploratory OLAP technology in accordance with method of conceptual integral OLAP-modeling.

Acknowledgment

The reported study was funded by RFBR according to the research project №16-07-01001.

References

[1] Codd E.F., Codd S.B., Salley C.T. Providing OLAP. On-line Analytical Processing to User-Analists: An IT Mandate. Codd & Associates; 1993.

[2] Nozhenkova L.F., Shidurov V.V. OLAP-technologies of operational analytical support of administration. Information technologies and computer systems, 2010, No 2, pp.15-27.

[3] Penkova T. Korobko A. Method of constructing the integral OLAP-model based on formal concept analysis. Frontiers in Artificial Intelligence and Applications, Vol. 243, 2012, pp. 219-227, doi:10.3233/978-1-61499-105-2-21.

[4] Wille R. Restructuring Lattice Theory: an approach based on hierarchies of concept. Reidel, Dordrecht-Boston, 1982.

[5] Ganter B., Wille R. Formal Concept Analysis: mathematical Foundations. Springer-Verlag. Berlin Heidelberg New York, 1999.

[6] O. Romero and A. Abelló, "Automatic validation of requirements to support multidimensional design," Data & Knowledge Engineering, vol. 69(9), 2010, pp. 917-942.

[7] R. Winter, B. Strauch, "A method for demand-driven information requirements analysis in data warehousing projects," Proceedings of the 36th Annual Hawaii International Conference on System Sciences, IEEE, 2003, pp. 231-239.

[8] M. Golfarelli, D. Maio, and S Rizzi, "The dimensional fact model: A conceptual model for data warehouses," International Journal of Cooperative Information Systems, 7.02n03, 1998, pp. 215-247.

[9] P. Giorgini, S. Rizzi, M. Garzetti, Goal-oriented requirement analysis for data warehouse design, Proc. of 8th Int. Workshop on Data Warehousing and OLAP, ACM Press, 2005, pp. 47–56.

[10] C. Phipps, K.C. Davis, "Automating data warehouse conceptual schema design and evaluation," Proc. of 4th Int. Workshop on Design and Management of Data Warehouses, vol. 58, 2002, pp. 23 –32.

[11] B. Vrdoljak, M. Banek, and S. Rizzi, "Designing Web Warehouses from XML Schemas," Proc. Fifth Int'l Conf. Data Warehousing and Knowledge Discovery (DaWaK '01), 2003, pp. 89-98.

[12] S. Mansmann, N.U. Rehman, A. Weiler and M.H. Scholl, "Discovering OLAP dimensions in semi-structured data," Information Systems, vol. 44, 2014, pp. 120-133.

[13] M.R. Jensen, T. Holmgren, T.B. Pedersen, "Discovering multidimensional structure in relational data," 6th Int. Conf. on Data Warehousing and Knowledge Discovery, LNCS, vol. 3181, Springer, 2004, pp. 138–148.

Information's semantic search, classification, structuring and integration objectives in the knowledge management context problems

Yury Kravchenko
Department of CAD
Southern federal university
Rostov-on-Don, Russia
krav-jura@yandex.ru

Elmar Kuliev
Department of CAD
Southern federal university
Rostov-on-Don, Russia
elmar_2005@mail.ru

Ilona Kursitys
Department of CAD
Southern federal university
Rostov-on-Don, Russia
i.kursitys@mail.ru

Abstract — The article is devoted to setting objectives and developing ways to solve problems of knowledge semantic search, classification, structuring and integration in relation to knowledge management issue. Information management is viewed as a set of processes systematic acquisition, synthesis and sharing of knowledge. Knowledge management automation problem as continuous knowledge verification process to identify patterns in order to create and meet demand for new knowledge, directly related to the solution of information's semantic search, classification, structuring and integration problems. One of the major scientific challenges in the field of information technology today is the development of information analysis and processing mechanisms in heterogeneous sources with the aim of empowering information systems logical analysis of information capabilities and generate conclusions that will form the basis of knowledge accumulation and processing execution procedures. Since, according to scientists, the decision of information overflow actual problem will shift from data storage and processing to the knowledge accumulation and processing. Despite the pronounced specificity of subject areas, ontology should be built as a chain of interrelated processes that will provide integrated nature of intellectual knowledge management system. As a source of knowledge operational information systems' database is most easily connected through a mechanism for creating data warehouses. Integrating knowledge from different sources may be based on the ontology requirements for the development of which will be a pre-formed sheets.

Index Terms — Semantic search, ontology, classification, structuring, integration, knowledge management systems, information processes, decision support.

I. Introduction

The problem of information overflow appears in an environment with intense information flows exchange. In such conditions the network of heterogeneous information sources pass through the "growing point", when the results of knowledge collecting and processing become less attractive, since the users meet information overflow challenges.

Let us consider knowledge management systems as the development of information systems, which gained definite technological results in solving problems of effective information storing, processing and delivery to the user under the rules or on a formal request.

The suggested hypothesis is an assumption that semantic identification of key information can be effectively carried out on the basis of the semantic concepts taxonomy tree as a systematization of complex reality areas and hierarchical knowledge to define and arrange terms and its synonyms. The difficulty of taxonomy semantic tree building problem is to determine effective ways of semantic search, classification, structuring and integration of information.

II. Semantic knowledge search

In the most general context information search is a complex problem which includes two subtasks:

1) to compare user's vision of required knowledge with the content from available distributed heterogeneous knowledge sources;

2) on the basis of this comparison to build the information object with finite set of properties, which values are retrieved from sources mentioned above.

Such problem is difficult to formalize, so it is worth to deal with simplified subtasks of semantic search with preset restrictions to user's requirements and to information objects that will be generated.

The major tendencies of information search systems are [1-3]:

- from formal search to semantic search;
- from unified search to personified search;
- from individual search to shared search;
- from closed search to manageable search.

Semantic search is a type of search, when comparison and building the information object mentioned above are carried out on the semantic level. Semantic search differs from the traditional one due to the usage of knowledge about the searched object, sources and search domain, and the ability to retrieve knowledge rather than data.

In most cases the subject of search is some data about any object. Thus, user has some pieces of information about this object, its properties and structure, and the missing data is expected to be retrieved after search is done.

Information object (IO) is a model of some real or virtual object of a domain, which defines structure, attributes,

constraints and behavior of this object. The structure of IO can be represented with the use of classes of corresponding ontology as the background. In the context of semantic search the IO can have previously set complex structure, formalized as a class of corresponding ontology. The taxonomy of the IO allows the user to describe in proper form what information is needed and what the user knows about it. Therewith, there is a problem of searching the ontology that represents the structure of the IO, the knowledge about which the user needs.

The ontology O represents a semiotic system $O=<P,V,R,C>$, where P denotes a set of concepts (classes); V is a set of concept examples; R is a set of predicates – types of relations; C is a set of relations that assigns the following types of relations between entities:

1. Partial order on sets P and R, which assigns relations is-a – "subclass-superclass" – "up-down".

2. Relations between concepts that have the form of a triple $<p_1 - r_1 - p_2>$, where $p_1, p_2 \in P$; $r_1 \in R$.

3. Relations between occurrences thich have the form of a triple $<v_1 - r_1 - v_2>$, where $v_1, v_2 \in V$; $r_1 \in R$.

4. Relations between predicates which have the form of a triple $<r_1 - r_i - r_2>$, where $r_1, r_2, r_i \in R$.

A closeness estimate between a piece of knowledge and a query is a numeric value, which presents a degree of likeness between them. The closeness estimate can be considered as semantic if and only if it is evaluated on the basis of documents' and query's semantics [4-9].

The idea of semantic search includes description of search query in the form of set of triples. Suppose that a query q which contains a set of triples $T(q)$. The result of search in knowledge source will be presented as a set of knowledge elements $E=\{e_i \mid i \epsilon [1,k]\}$, where k is a number of knowledge elements e_i. Besides, the semantic metadata of the set of knowledge elements T(e) must meet the following condition of semantic similarity $sim(T(q),T(e))$ to the query description: $T(q):sim(e,q)=sim(T(q),T(e))>\varepsilon$, where $sim(e,q)$ means similarity of a query q to the element e, ε denotes a threshold value of relevance. The search results are ranked in accordance with its semantic similarity to the query.

Semantic search allows us to accumulate, retrieve, update and integrate knowledge from different domains on the basis of its interdisciplinary relations reveal and usage.

III. KNOWLEDGE CLASSIFICATION.

Let us introduce the knowledge classification problem's mathematical model. Suppose that X is a set of knowledge elements description, Y is a set of classes' names. There exists an undefined objective dependence presented in a form of a function $y^*:X \rightarrow Y$, which values are defined on the finite training set $X^m=\{(x_1,y_1),\ldots,(x_m,y_m)$ only. It is required to develop an algorithm $a:X \rightarrow Y$ which is able to classify a random object $x \epsilon X$.

Consider the classification problem in context of preceding and following problems of cyclical script for knowledge management (Fig.1).

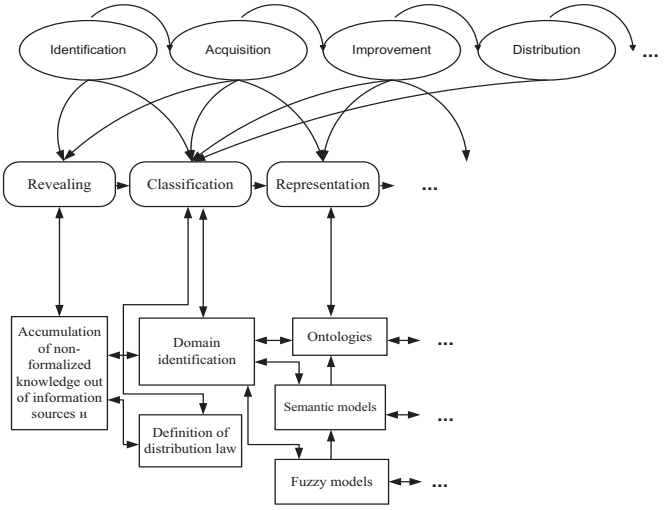


Fig. 1. The position of the classification problem in context of knowledge management problems

Knowledge management processes are based on the following principles: identification, acquisition, improvement, distribution, usage and storage. The knowledge classification problem is preceded by the revealing problem responsible for the retrieving of non-formalized knowledge out of heterogeneous sources. The following problem is a knowledge representation problem, i.e. knowledge formalization.

Some examples of knowledge classification are:

Knowledge classification by Machlup:

- practical knowledge;
- intelligent knowledge (including scientific, humanitarian and cultural);
- common knowledge;
- religious knowledge;
- unwanted knowledge;

Classification of organizational knowledge by Tonak and Takeuch:

- explicit (systematized) knowledge – can be represented as words and figures and easily transferred and exchanged in the form of accurate data, scientific formulas, orderly procedures or universal principles;
- implicit (unsystematized and non-formalized) – is somewhat hard to reveal and to express. [5, 10-17].

Let us consider knowledge identification as the first stage of defining the system of interdisciplinary relations between the used domains. The main purpose is to reveal knowledge that is already acquired. Firstly, the existing data is divided into explicit and implicit (non-formalized). Implicit knowledge becomes explicit in the process of externalization. The next step is to define the degree of knowledge specification and importance. It is connected with knowledge acquisition – a process of knowledge source selection, utility estimation and knowledge refinement, supporting the consistency between knowledge intake and its demands.

Knowledge improvement is made during the semantic search in heterogeneous distributed sources and encyclopedic

help systems based on ontological and case-models, and models of successful search query precedents. Knowledge improvement can be considered as a subtask of new knowledge creation, providing conditions for creative activity, generation of ideas, knowledge exchange and integration.

IV. KNOWLEDGE STRUCTURING

The difficulty of knowledge accumulation and processing involves the estimation of presence or absence of significant consistency characteristics in knowledge elements structure from different domains and in determination of relations between them. The hypothesis suggests, that this complex problem that can be solved with the use of knowledge structuring at a certain stage.

Knowledge structuring is a process of dividing knowledge elements into stable groups and subgroups. The following principles of knowledge structuring are:

- Knowledge must be divided into groups and subgroups in accordance with definite system-relevant attributes.
- Formed groups and subgroups must be logically connected and arranged in the appropriate order.

The extent of saved and copied information exceeded human cognitive ability long ago. Thus, every subject of knowledge accumulation and processing is challenged with two relevant problems:

- to exclude the information noise from analyzed sources;
- to counter the information manipulation.

The candidate solution of these problems is the use of knowledge structuring [6,7,18-20].

Let us introduce a number of definitions for the further problem consideration. Thinking denotes the reflection of the main object and events attributes in human brain. The most important attribute is called the essential (system-relevant) attribute. Knowledge in common sense means subjective reflection of reality in the form of concepts, definitions, and relations between them [7,21,22]. The hypothesis claims that a concept has the character of universal reflection of objects' and real-world events' essential attributes, while the representation is just descriptive image of an object which is composed of unessential attributes and of individual character. The concept definition is considered as logical action that allows us to disclose concept content and involves essential attributes only.

Knowledge systems is based on the assignment of relations in a set of knowledge elements from the same domain or between different domains at the interdisciplinary level.

Let us formulate the knowledge structuring problem statement. All the system-relevant attributes of knowledge elements from the same domain can be divided into *m* classes. Now, we can form a set of essential attributes of the system relevance

$$F = \{F_1 \cup F_2 \cup \ldots \cup F_m\}.$$
$$F_1 = \{f_{11}, f_{12}, \ldots, f_{1(i-1)}, f_{1i}\}, \quad (1)$$
$$F_2 = \{f_{21}, f_{22}, \ldots, f_{2(j-1)}, f_{2j}\},$$
$$F_m = \{f_{m1}, f_{m2}, \ldots, f_{m(k-1)}, f_{mk}\},$$

where $f_{11}, f_{12}, \ldots, f_{1(i-1)}, f_{1i}$ are elements of set F_1, that assign 1st class of system-relevant attributes for knowledge elements from some domain; $f_{21}, f_{22}, \ldots, f_{2(j-1)}, f_{2j}$ are elements of set F_2, that assign 2nd class of system-relevant attributes for knowledge elements from some domain; $f_{m1}, f_{m2}, \ldots, f_{m(k-1)}, f_{mk}$ are elements of set F_m, that assign m-class of system-relevant attributes for knowledge elements from some domain.

Let us assign a set of existing essential attributes

$$Q_z = \{Q_{11} \cup Q_{12} \cup \ldots \cup Q_{nm}\} \quad (2)$$

for each knowledge element analyzed q_z ($z = 1 \ldots n$), where $Q_{11} \subset F_1, Q_{12} \subset F_2, Q_{nm} \subset F_m$.

Then, the expression for defining the knowledge element consistency to system-relevant requirement of domain can be represented in the following form:

(3)

Objective function will be presented as

$$F. \quad (4)$$

In other cases the knowledge element incompletely corresponds to requirements of system-relevant knowledge which makes it possible to modify the knowledge element or to exclude it from the search results [8,9].

To make a formal representation of described problem statement model we use the binary directed graph:

$$G=<P,F,I,L>, \quad (5)$$

where P is a set of knowledge elements; F – a set of required system-relevant attributes; I – a set of arcs to characteristics of a certain knowledge element; L a set of arcs to desired system-relevant characteristics for the corresponding domain.

The knowledge structuring model is illustrated on the abstract example (Fig.2). This example can be explained as follows:

1) The knowledge element P1 has attributes $\{F_{11}, F_{1(i-y)}, F_{2j}, F_{m1}\}$, (arcs:$I_{11}, I_{12}, I_{13}, I_{14}$), where *y* is undefined integer value. Attributes $\{F_{11}, F_{1i}, F_{21}, F_{mk}\}$, (arcs:$L_{11}, L_{12}, L_{13}, L_{14}$) are system-relevant for considered domain. Thus, the expression for defining the consistency of knowledge element to system-relevant domain requirement have a form of $M_1=I_1 \cap L_1=F_{11}$, and the purpose of the search or refinement will have a form of $S_1=L_1 \setminus M_1=\{F_{1i}, F_{2l}, F_{mk}\}$.

2) The knowledge element P2 has attributes $\{F_{1i}, F_{m1}, F_{mk}\}$ (arcs:I_{21}, I_{22}, I_{23}). Attributes $\{F_{1i}, F_{2j}, F_{m1}\}$, (arcs :L_{21}, L_{22}, L_{23}). are system-relevant for considered domain. Thus, the expression for defining the knowledge element consistency to system-relevant requirement of domain have a form of $M_2=I_2 \cap L_2=\{F_{1i}, F_{ml}\}$, and the purpose of the search or refinement will have a form of $S_2=L_2 \setminus M_2=F_{2j}$.

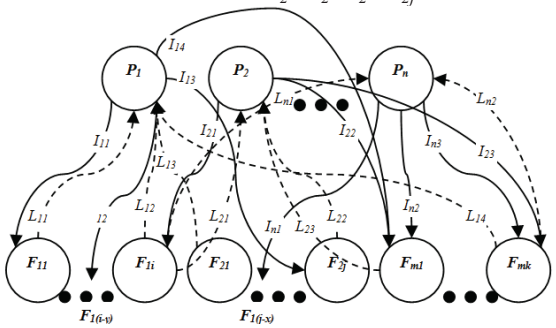


Fig. 2. The knowledge structuring graph model based on estimation of system-relevant attributes presence

3) The knowledge element P_n has attributes $\{F_{1(j-x)}, F_{m1}, F_{mk}\}$ (arcs:I_{n1}, I_{n2}, I_{n3}). Attributes $\{F_{1i}, F_{mk}\}$, (arcs:L_{n1}, L_{n2}) are system-relevant for considered domain. Thus, the expression for defining the knowledge element consistency to system-relevant requirement of domain have a form of $M_n=I_n \cap L_n=F_{mk}$, and the purpose of the search or refinement will have a form of $S_n=L_n \setminus M_n=F_{li}$.

The problem of knowledge accumulation is difficult and time-consuming. Uncontrolled growth of information flows results to lack of resources needed for processing of information flow. One of the steps of solving this problem is knowledge structuring that unambiguously identifies relations on sets of concepts and definitions.

The main idea of knowledge structuring is the opportunity to build a multilevel net of concepts connected by definitions. High-level concepts can be defined when relations on low-level concepts are set.

V. Knowledge Integration.

The knowledge integration problem is referred to a set of development subtasks: knowledge bases of decision making and precedents modules; sets of object, ontological, fuzzy, semantic and analytical models to implement decision making; modules of model selection forming on the basis of knowledge base mathematical and simulation modeling.

To develop interdisciplinary knowledge the integration of already developed ontologies of different domains is required. Let us formulate the problem of many ontologies system integration:

(6)

Where O_i is the ontograph, i is the domain index number, N *is* quantity of domains.

The union is considered as conceptual system integration of ontological graphs (OG) and their relations. The knowledge scope V is estimated through formal-ontological representation parameters. if represented by ontograph without considering types of relations and difficulties of interpretation functions V will take on a value of OG's vertex number. In case of a simple tree structure the expression for defining the knowledge scope in domain will have the following form:

(7)

Where $Z_{p,d}$ denotes the degree of incidence of a vertex d; $\overline{1,\mathrm{P}}$denotes the number of OG levels; $\mathrm{d} = \overline{1,\mathrm{D_p}}$ denotes the number of vertex of OG's p-level.

Considering relation types and difficulty of the interpretation function results to OG with measured vertex and edges. In such case, the expression for defining the knowledge scope will have the following form:

, (8)

Where γ_d и $\delta_{d,j}$ means values of weight functions of relations and interpretation.

The main results of conducted research assume the further work on the development of formal methodology for design the domain ontology, algorithms and procedures for knowledge system integration with the use of bioinspired search methods and algorithms.

VI. Experimental Research

Let us use semantic networks as the structure that contains the base of distributed heterogeneous knowledge sources, divided into objects in accordance with selected attributes with clearly defined relation hierarchy.

Before building the semantic network, we need to decompose resource knowledge into several structural units to improve the efficiency of their relation processing procedures. Any quantity of decomposition units can be assigned. The optimal decomposition units number is considered to be 6 [2].

That kind of organization allows us to distinguish researching material detail levels (Fig.3). The first level contains root node of the network, it generalizes the network content. Each new level discloses the content of preceding level, revealing the whole domain for the researcher and showing interdisciplinary connections with other nodes.

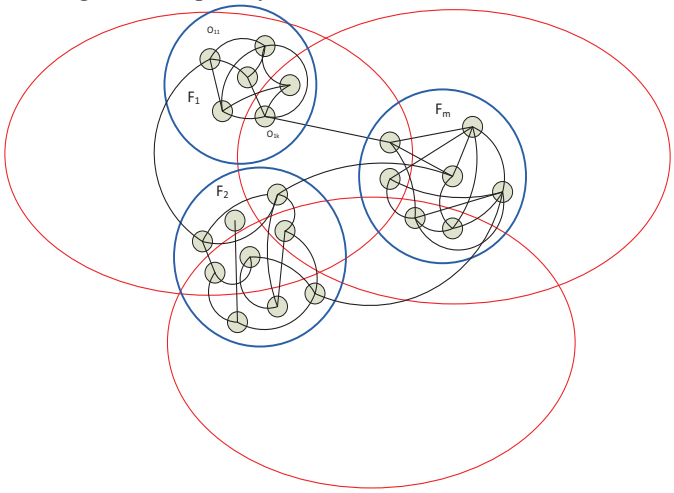


Fig. 3. The abstract model of the semantic network with several decomposition levels.

The network involves a set of domains $F\}$, where $z = \overline{1,m}$, m is an integer constant denoting the domains quantity, and $\}$, where k_z is an integer constant denoting the number of knowledge elements in F_z. In case of assigned relations between objects of the same domain, we obtain n, where n is the number of relations between , z is an integer constant denoting the index number of researched domain. If set all the relations between domain objects, including interdisciplinary ones, we will obtain $\forall$ p, where p is the number of relations between the network,.

VII. Conclusion

This paper develops methods of solution the problems of semantic search, knowledge classification, structuring and integration in the context of knowledge management problem.

The information flows management is considered as a complex of processes of the semantic acquisition, synthesis, exchange and usage of knowledge. The confidence estimation which gives arguments and evidence of this statements is based

on the usage of metadata and metastatements in knowledge management rather that information management technologies.

Concrete scientific results are shown as systematized complex of actions required to solve the problem of investigation and search of relations in distributed knowledge sources in order to increase the probability of new knowledge generation.

ACKNOWLEDGMENT

The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

REFERENCES

[1] D. Amerland, Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence / D.Amerland. – Que Publishing, 2013. – 230 p.

[2] V.V. Bova, Y.A. Kravchenko, V.V. Kureichik, Decision Support Systems for Knowledge Management // Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol. 3. – Springer International Publishing AG Switzerland, 2015. – P.123-130.

[3] V.V. Bova, Y.A. Kravchenko, V.V. Kureichik, Development of Distributed Information Systems: Ontological Approach // Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol. 3. – Springer International Publishing AG Switzerland, 2015. – P. 113-122.

[4] A.N. Dukkardt, E.V. Kuliev, V.V. Kureichik, A.A. Novikov, Automated formation of the interactive tasks for the computer-aided training (2015) 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, art. no. 7338632, pp. 604-608.

[5] N. Guarino, Towards a Metotodology for Ontology - Model Engineering [Electronic resource] / N. Guriano, C.A. Welty // Proceeding of the ECOOP-2000 Workshop on Model Engineering (eds. by Bezivin J. and Ernst J.). – 2000. – Mode of access: http://www.metamodel.com/IWME00/articles/guarino.pdf.

[6] A. Gangeni, An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies [Text] / A. Gangeni, D.M. Pisanelli, G. Steve // Data & Knowledge Engineering. – 1999. – Vol. 31. – P. 183-220.

[7] S. Rodzin, O. Rodzina, New computational models for big data and optimization (2015) 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, art. no. 7338504, pp. 3-7.

[8] Y.A. Kravchenko, V.V. Kureichik, D.Y. Zaporozhets, D.V. Zaruba, Information and knowledge integration based on simulation modeling (2015) 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, art. no. 7338508, pp. 22-24.

[9] Y.A. Kravchenko, V.V. Kureichik, Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014". – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.

[10] A.A. Lezhebokov, D.V. Zaruba, D.Y. Zaporozhets, Three-dimensional content in mobile applications for educational information systems (2015) 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, art. no. 7338567, pp. 299-304.

[11] V.M. Kureichik, S.I. Rodzin Evolutionary algorithms: genetic programming // Journal of Computer and Systems Sciences International. 2002. – T. 41, № 1. – C. 123-132.

[12] L.A. Gladkov, N.V. Gladkova, A.A. Legebokov, Organization of knowledge management based on hybrid intelligent methods (2015) Advances in Intelligent Systems and Computing, 349, pp. 107-112.

[13] A.N. Dukkardt, A.A. Lezhebokov, D.Y. Zaporozhets, Informational system to support the design process of complex equipment based on the mechanism of manipulation and management for three-dimensional objects models (2015) Advances in Intelligent Systems and Computing, 347, pp. 59-66.

[14] V.V. Bova, V.V. Kureichik, A.A. Lezhebokov, Integration of ontologies in scope of model and conceptual semantics: Modified approach (2015) 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, art. no. 7338537, pp. 156-160.

[15] S.M. Kovalev, V.V. Kureichik, V.M. Kureichik, S.V. Sokolov, Optical fuzzy logic systems in problems of adaptive simulation of weakly formalized processes (2011) Journal of Computer and Systems Sciences International, 50 (3), pp. 462-471.

[16] Qing He, Xiu-Rong Zhao, Ping Luo, Zhong-Zhi Shi Combination methodologies of multi-agent hyper surface classifiers: design and implementation issues // Second international workshop, AIS-ADM 2007, Proceedings. – Springer Berlin Heidelberg, 2007. – pp. 100-113.\

[17] A. De Nicola, M. Missikoff, R. Navigli A software engineering approach to ontology building // Informatioh systems, vol. 34, pp. 258-275, 2009.

[18] N. Guarino, D. Oberle, S. Staab What is an Ontology // Handbook on Ontologies. – Springer, 2009, pp. 1-17.

[19] X.-S. Yang, A new metaheuristic sat-inspired algorithm [Text] / X.-S.Yang // Nature Inspired Cooperative Strategies for Optimization (NISCO'2010), Berlin: Springer, 2010, Vol. 284. pp. 65-74.

[20] J.Sarraipa, et al., Semantic Enrichment of Standard-based Electronic Catalogues // 13th IFAC Symposium on Information Control Problems in Manufacturing, 2009.

[21] L. Kerschberg, W. Kim, A. Scime, Personalizable semantic taxonomy-based search agent. USA: George Mason Intellectual Properties, INC (Fairfax, VA). – 2006.

[22] L. Kerschberg, H. Jeong, W. Kim, Emergent Semantic in Knowledge Sifter: An Evolutionary Search Agent based on Semantic Web Services // Journal on Data Semantics. – VI. LNCS, vol. 4090, pp. 187-209. Springer, Heidelberg (2006).

Multi-objective particle swarm optimization for ontology alignment

Semenova A.V.
Computer sciences and information security institute
Southern Federal University
Taganrog, Russia
alexaforum@rambler.ru

Kureychik V.M.
Computer sciences and information security institute
Southern Federal University
Taganrog, Russia
kur@tgn.sfedu.ru

***Abstract* – In computer science design and implementation of high-tech areas in the modern society is accompanied by increasing the role of ontological knowledge base. Accumulation of shared ontologies is seen as a mechanism of unlimited knowledge acquisition about the world. However, the problem of integration, matching and alignment of ontologies is not solved yet. The problem of ontology alignment is to find such a structure and permissible parameters that provide the optimal values for one or more quality criteria. It should be noted that today there are many methods to compute the similarity between two discrete elements of different ontologies. Integration of up-to-date similarity computation techniques allows obtaining a versatile and accurate result. One of approach is based on the weights. Typically, the weights are assigned manually or by specific approaches. The main shortcoming of existing approaches is the lack of optimality. This article proposes a new combined approach for ontology alignment based on Latent Semantic Indexing and multi-objective particle swarm optimization method. For objective functions two criteria were chosen: the accuracy and recall. To obtain an optimal population the method of local search was employed to replace the worst of the population in the new generation. Experimental research of the suggested approach confirms the effectiveness of it.**

***Index Terms* – ontology alignment, particle swarm optimization, knowledge bases, concepts, components, multi-objective optimization.**

I. Introduction

Nowadays problem of big data processing is very urgent and of great importance in information science. Variety of different information systems inevitably leads to the problem of providing their interaction. Ontological approaches for data modeling as well as semantic data storage and processing technologies have been developed for operations with quasistructured data. The development of semantic technologies has led to the expansion of their scope including structured data life cycle of complex engineering objects. Such approaches allow to provide efficient communication among huge number of different people and computers as part of the increasingly complex innovative engineering projects.

Generally speaking, by ontology we understand a system of concepts of a domain, which is represented as a set of components associated by different relations. Ontologies are used for formal specification of concepts and relationships that characterize a specific area of expertise. The advantage of ontologies as a means of knowledge representation is their formal structure that facilitates their computer processing. Ontologies can be defined as an "explicit, formal specification of a shared conceptualization" [1].

Most formal ontologies declare individuals/instances, classes/concepts, attributes, as well as relations. Classes (or concept) are the basic categories, which can be arranged in taxonomy. Each class describes a group of individual instances, which are combined under common set of properties. The concepts may have different relations with another concepts and instances. The formal definition of ontology is the following expression [1]:

$$O = (C, P, R, A), \quad (1)$$

where

C – denotes concept (or classes) set for a specific domain;

P – set of concepts attributes. Property is a component of the relation p(c,v,f) , where c∈C – ontology concept, v – property value, associated with c and f defines restrictions for facets in v. One of restriction is a type (tf), capacity (cf), and range (rf).

R = {r | r⊆ C ×C × R t} – set of binary relations between concepts in C. There is the following variety of relation types: 1:1, 1:many, many:many. The basic set of relations are: $synonymOFF, kindOFF, partOFF, instanceOFF, propertyOFF$.

A – axioms' set. Axiom is a rule (real fact) that specify cause-and-effect relationship.

Another specific point of view on the concept of ontology is considering the ontology as a logical descriptive theory. One of these definitions is, for example, the definition given in [2]: "ontology – a logical theory of considering of destined meanings of formal vocabulary. From the point of view of logic descriptions ontology may be defined as a tuple" [2]

$$O = \langle S, I \rangle, \quad (2)$$

where S – set of statements belonging to the level of ontology scheme O, and I – statements, constituting a set of components. Ontology scheme S is a terminological component of ontology.

Shared ontologies accumulation is seen as a mechanism of unlimited accumulation of knowledge about the world. Ontologies are a specification of knowledge in a formal language in which the fixed group of experts agreed about name conversions and their properties in the specific domain. So the problem of heterogeneous data integration occurs. In view of possible (and very likely) the presence of structural, lexical and semantic conflicts in various data sources, formation of an ontological knowledge base is a complex multi-step process. Therefore, the problem of integration of ontologies, providing finding semantic similarities between the elements of heterogeneous ontologies, is relevant in the field of information technologies [3].

There are different approaches for solving the problem of data heterogeneity in ontologies. However, they can generally be classified into one of two classes of solutions. The first class is a centralized solution, in which the basic ontology already built. Most of the known techniques for ontology matching are based on rules (rule-based), or on the methods of machine learning. Both approaches have drawbacks, such as ignoring developer's information contained in the ontology class instances. A more serious drawback is the way in which these techniques treat the semantic components of ontologies while matching process [4].

By ontology alignment we understood the establishment of correspondences (or matches) between two or more heterogeneous ontologies and their storage / operation or a set with established correspondences, which is the result of a comparison of ontologies.

Process of ontology alignment typically consists of the following steps (Figure 1).

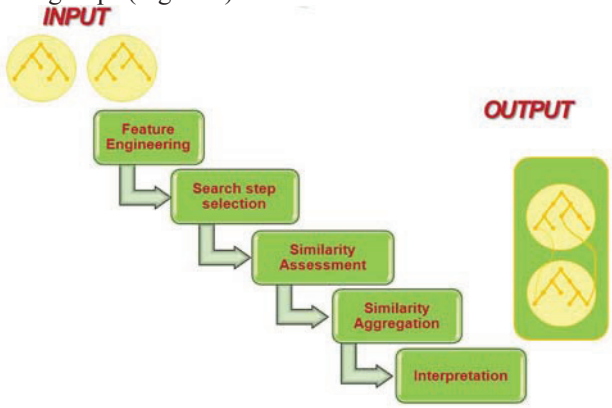


Fig. 1. Process of ontology alignment

Let's consider a formal definition of ontology alignment, borrowed from [5]. Ontology alignment process may be expressed by a function based on the E-set of all components such as e∈E and based on the set of potential O ontologies is a partial function [5]

$$align: E \times O \times O \to E \qquad (3)$$

It should be noted that $align_{O_1O_2}$ holds for $align(e, O_1, O_2)$. Once an alignment between different ontologies O_1 and O_2 is performed, then component e is aligned with component f, wherein $align(e) = f$. A pair of components (e, f) that is not aligned yet and for which suitable alignment criteria still need to be checked is called a "candidate alignment".

Now consider an alternate way for formalization of the process of ontology alignment.

Let $O_1 = \langle S_1, I_1 \rangle$ and $O_2 = \langle S_2, I_2 \rangle$ be ontologies formalizing semantics of a certain domain P. Thus the ontologies O_1 and O_2 are the definite interpretations of the P. The domain P is considered as a set of facts f. So ontologies O_1 and O_2 are equivalent if and only if $\forall f \in P\ int_{O1}(f) \equiv int_{O2}(f)$ (where $int_{Ox}(f)$ is an interpretation of a fact f pertaining to ontology O).

Suppose Δ be an abstract measure of difference of interpretations

$$\Delta = \Delta(P, O_1, O_2), \Delta \in [0,1], \cdot \qquad (3)$$

and Δ equals to 0 if and only if $O_1 \equiv O_2$ and increase monotonically with number f ∈ P such that $\neg(int_{O1}(f) \equiv int_{O2}(f))$. So ontology alignment process lies in finding of a set of all possible pairs of $int_{O1}(f) \circ int_{O2}(f)$ where $\circ \in \Re = \{\subset, \subseteq, \equiv, \supset, \supseteq\}$.

Hence the alignment of two ontologies is a set $A = \{int_{O1}(f) \circ int_{O2}(f)\}$, where Δ=inf(ranΔ) and ranΔ is a range space of for the O_1 and O_2.

To find well-defined optimal alignment is very difficult task. The heterogeneity and ambiguity of data descriptions makes unrealistic the scenario in which those optimal alignments for many pairs of components will be considered as "best alignments by any of the existing alignment algorithms. Consequently the optimal alignment is often unknown, because there is no gold standard as a reference.

The problem of ontology alignment is to find such a structure and allowed parameters that provide the optimal values for one or more quality criteria. Under criterion we understood system performance value regarding the goal. Thus each objective meets the criteria by which the degree of achievement of this goal can be estimated. Such optimization is called multi-objective. Typically, the optimization cannot be achieved simultaneously by all criteria, so the question which of the criteria preferable remains open. Numerical simulation of multi-objective tasks requires large computational cost and the very methods of multi-objective optimization regarding to in relation to ontology alignment are not sufficiently developed. So, researches in this direction have been actively carried out only in the last ten years [6].

In this paper we consider the multi-objective optimization approach for ontology alignment. Simultaneous analysis by two quality criteria allows demonstrating superiority of our approach over some currently available approaches of ontology alignment. The time complexity of the algorithm is quadratic.

II. Single Objective Optimization for Ontology Alignment

Ontology alignment problem is solved by finding optimal weights of semantic similarity measures of concepts. So, consider ontology alignment problem as optimization task. In

[7] a combined approach for ontology alignment based on swarm intelligence is suggested. Visualization this approach for optimal ontology alignment is depicted on the Fig. 2.

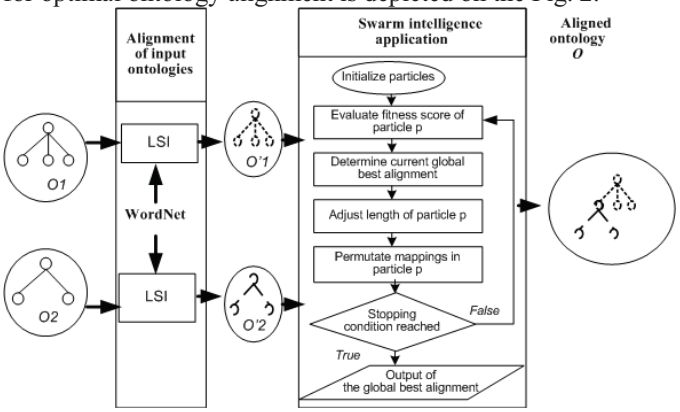


Fig. 2. Stages of optimal ontology alignment

At the beginning, an alignment between different input ontologies O_1 and O_2 is performed based on WordNet. WordNet is a semantic network for English developed at the Princeton University. The dictionary consists of four major networks for significant parts of speech: nouns, verbs, adjectives and adverbs. The basic unit in the WordNet dictionary is not a single word, and the so-called synonyms (synsets), combining the words with a similar meaning and are essentially network nodes. Words that have several meanings are included in several synsets and can be attributed to a variety of syntactic and lexical classes. Tokens of different parts of speech are stored separately and the description for each part of speech have different structures. So definitions of informal ontological concepts are mapped to formal standard concepts of intermediate ontology WordNet. Then Latent Semantic Indexing was applied. Latent semantic indexing (LSI) is a data processing method in a natural language, analyzing the relationship between the collection of documents and terms occurring in them, associating with certain factors (subjects) all documents and terms. The LSI method based on the principles of factor analysis, in particular the identification of latent connections studied phenomena or objects. In the classification / clustering the method used to extract context-dependent values of lexical units by means of statistical processing of large text blocks [8].

Consider operation sequence to optimize the alignment of ontologies maximizing the number of matches in the resulting ontology by application of swarm intelligence.

Optimization method using swarm intelligence (Particle Swarm Optimization, further PSO) based on the modeling of the behavior of a population of particles in the space of parameters of optimization problem. This method was proposed by Kennedy. The proposed method is attractive and easy to implement in that during the computation of the gradient is not used. The method can be used for many tasks, including the training of neural networks, the search task minimum functions and tasks typical of genetic algorithms. Aligning ontologies by discrete particle swarm method comprises a number of N particles (of the population). Thus, we have a number of iterations I. Commonly at each iteration the velocity vector is applied to each particle, which determines its new position in the parameter space. This evolution is due to controlled random repeated iterations within each particle.

Swarm intelligence adapted for ontology alignment provides achievement of the following two objectives [7]:

1. Identification of the best suitable alignment.
2. Maximize the number of matches in the alignment.

Swarm intelligence algorithm consists of three steps: initialization, iteration of particle swarm calculation and procedure for computing a new particle configuration and its update [7].

The calculation of an alignment starts with an initialization. In this initialization step, each particle is initialized with an arbitrary number of similarities. It also includes analysis, i.e. calculation of fitness value for each conformity and initial confirmation of personal best alignment.

First, each particle has a position and speed and the position and the speed of each particle be updated. The updated speed vector is used to produce a new particle X .

Let's $\vec{X}_p$ be a particle that is represented as a vector, p$\in$ {1,…,N}, where $c_{(p,k)}$ (for each j$\in${1,⋯,k}) is a similarity. By similarity $c_{(p,k)}$ we understand a pair of components (e_1, e_2) meaning that e_1 and e_2 refer to the same real-world component [9].

Total number of similarities in an alignment in space one-to-one may be at most n where $n = \min\{|C_1|, |C_2|\} + \min\{|P_1|, |P_2|\}$.

Dimension of particle (k) is variable and k$\in${1,…,n}. Let $F(\vec{X}_p)$ be the fitness function of A alignment for a particle $\vec{X}_p$. Each particle holds the configuration of the best alignment of dimensionality $l \in \{1, \dots, n\}$ (*pBest*) according to F that is designated as [11]

$$\vec{B}_p = \{d_{(p,1)}, d_{(p,2)}, \dots, d_{(p,l)}\} \quad (4)$$

where for each $j \in \{1, \dots, l\}$, $d_{(p,l)}$ is a conformity. According to [7], the dimension l of the best alignment (pBest) does not necessary match to the dimension k of its current configuration.

The best parameter configuration for any particle after swarm computation is denoted by [11]

$$\vec{G} = \{d_1, d_2, \dots, d_m\} \quad (5)$$

where for each $j \in \{1, \dots, m\}$, d_j is a conformity. Its dimension is m$\in${0,⋯,n}.

Now define the particle fitness vector using the matrix [11]

$$\vec{F}_p = \begin{pmatrix} f_{(p,1)} & f_{(p,2)} & \cdots & f_{(p,k)} \\ c_{(p,j1)} & c_{(p,j2)} & \cdots & c_{(p,jk)} \end{pmatrix} . \quad (6)$$

Moreover, fitness $f_{(p,\mu)}$ is confirmed on each correspondence $c_{(p,j\mu)}$. The vector is ranked according to its fitness value.

Now define the speed vector as an array of dimension 2 × k [11]:

$$\vec{V}_p = \begin{pmatrix} v_{(p,1)} & v_{(p,2)} & \dots & v_{(p,k)} \\ c_{(p,j1)} & c_{(p,j2)} & \dots & c_{(p,lk)} \end{pmatrix} \quad (7)$$

where $v_{(p,\mu)}$ is mapped on each correspondence $c_{(p,l\mu)}$. Vector is ranked by $v_{(p,\mu)}$. Values $v_{(p,\mu)}$ are used to increase the likelihood of those matches that are to be ranked in the particle, which is also available in the personal and absolute best alignments. Initially, for each $c_{(p,l\mu)}$ and $v_{(p,\mu)}$ set 1. This initialization is performed for all new correspondence joined to the particle during their evolution. Then $v_{(p,\mu)}$ is upgraded. Two parameters $\in \mathbb{R}^+$ and $\gamma \in \mathbb{R}^+$ are used for this purpose. If $c_{(p,l\mu)}$ belongs to $\vec{B}_p$, then β is added to $v_{(p,\mu)}$. If $c_{(p,l\mu)}$ belongs to $\vec{G}_p$, then γ is added to $v_{(p,\mu)}$. Then, every $v_{(p,\mu)}$ is multiplied by a universal random value $\phi_\mu \in (0,1)$.

Now calculate constant set which will not be substitute during a random re-initialization wherein k∈(0,1) [11]

$$F_{(p,k)} = \{c_{(p,j\mu)} | \mu \in \{1, \dots, k \cdot k\}, j_\mu - \text{ordering as in } \vec{F}_p\}, \quad (8)$$

$$V_{(p,k)} = \{c_{(p,l\mu)} | \mu \in \{1, \dots, k \cdot k\}, l_\mu - \text{ordering as in } \vec{V}_p\} \quad (9)$$

Consequently, the sets $F_{(p,k)}$ and $V_{(p,k)}$ contain such matching particles that $k \cdot k$ are well evaluated and have the highest rank according to their values $v_{(p,\mu)}$, respectively. Define the static set

$$K_{(p,k)} = F_{(p,k)} \cap V_{(p,k)}, \quad (10)$$

containing such matching, which are part of both sets $F_{(p,k)}$ and $V_{(p,k)}$.

For more strict convergence with respect to optimal alignment additional secure set of correspondences is applyed, which will never be replaced in the particle:

$$S_{(p,\sigma)} = \{c_{(pj\mu)} | f_{(p,\mu)} < \sigma\}, \quad (11)$$

where σ∈ (0,1) - is the threshold for fitness matching which are to be included in a secure set. Since there is a possibility of "jamming" in the local optimum for the alignment, value of σ must be chosen as small as possible. Each update algorithm first calculates the new length of the particle k' 'in accordance with the process of self-adaptation. Then, the particle retains the set $S_{(p,\sigma)} \cup K_{(p,k)}$, replaces the remaining matching $k' - |S_{(p,\sigma)} \cup K_{(p,\sigma)}|$ by new random ones. This behavior ensures the convergence of each particle to the optimal (5), as static set is continuously increasing and deviation due to random re-initialization are less drastic with the evolution of the swarm.

The execution of the algorithm is an iterative, guided by development of particle swarm. Each particle can be evaluated and updated in parallel. It is the benefit of this approach [7].

The population in Particle Swarm Optimization is initialized with random particles and the candidate solutions or particles move around the search area with the aim to achieve optimal alignment.

Results obtained during swarm intelligence calculations lead to different manipulations with concepts of ontologies:

- if the concepts are equivalent, they represent the same concept in the ontology, therefore they should be merged;
- if one ontology concepts is a generalization of the corresponding concept of other ontology, they should be presented as a concept and subclass respectively. Matching attributes are removed from the subclass, because they will inherit characteristics of the superclass;
- if one concept of integrated ontology corresponds to several objects of the main ontology, then an adjustment of the resulting ontology at the structural level is performed. The concepts of one ontology are a refinement of a corresponding concept of other ontology. Such concepts should be submitted as the subclass and class, respectively. Matching attributes are removed from the subclass, because they are inherited from the superclass;
- if two concepts of ontology are equivalent partially, they have common subclass. It is their generalization. Corresponding attributes are removed from the subclass, because they will inherit the characteristics of the generalized concept.

This strategy increases the search dimension, which is essential for finding the optimal solution. In the proposed approach, the mapping of less formal descriptions of concepts to the standard ontology concepts is performed. The advantage of the proposed approach is to minimize subjectivity introduced by man, by the method of calculation of weight coefficients particle swarm instead of setting a priori expert in the subject field. The use of the proposed approach to combining ontological knowledge allows us to solve the problem of automatic alignment and direct association of ontologies. However, it produces only a one solution for ontology alignment, because it is designed as a single objective optimization problem. This fact has motivated us to develop an approach where swarm intelligence is modeled as a multi-objective optimization problem to reach several optimal solutions ontology alignment.

III. Multi-objective optimization for ontology alignment

Calculating the optimal measure of similarity for alignment is a difficult task, because its solution requires the use of reliable and efficient methods that allow getting comprehensive and accurate alignment. Consider the swarm intelligence method for ontology alignment, using multi-objective optimization approach. Multi- objective optimization or parallel programming is the process of simultaneous optimization of two or more conflicting objective functions in a given domain. Multi-criteria optimization task is formulated as follows [12]:

$$\min_{\vec{x}} \{f_1(\vec{x}), f_2(\vec{x}), \dots, f_k(\vec{x})\}, x \in X$$

where the integer k⩾2 is the number of objectives and the set X is allowable set of decision vectors. Commonly the allowable set is determinated by constraint functions. To

compute vector-valued objective function the following expression is applied [12]

$$f: X \to \mathbb{R}^k, f(x) = \left(f_1(x), \ldots, f_k(x)\right)^T$$

If some objective function is to be maximized, it is equivalent to minimize its negative.

Multi-objective optimization task is to find a vector of target variables satisfying cash constraints and optimizing the function of the vector whose elements correspond to the objective function. These functions form a mathematical description of the satisfactory test. Hence, to "optimize" means finding a solution, in which the objective function would be acceptable to the director task [5].

Consider the set of data, wherein the data lines are different similarity coefficients and columns - the relations between two different ontologies. For subsequent combining these similarity coefficients in one metric optimum weights were obtained. The proposed approach is possible to find the set of weights that meet the criteria of similarity which allows obtaining an optimal alignment. In the process of evaluation of swarm intelligence generalized function was calculated

$$f_{integ}(O1_i, O2_i) = \sum_{k=1}^{7} w_k \times F_k(salign_{ij}) \quad (12)$$

where $\sum_{k=1}^{7} w_k = 1$.

If $f_{общ}(O1_i, O2_i)$ exceeds a threshold then $salign_{ij}$ is a valid alignment. In such a way all valid alignments are defined. Subsequently, using these valid alignments and reference alignments objective functions are calculated.

The approach consists of the following stages.

A. Initialization

The population is called a swarm, and it is composed of m number of appropriate solutions or particles. Each particle has n positions or cells comprising n weighting coefficients corresponding to n different similarity measures. An example of coding model particles having seven cells which are converted into seven weighting coefficients (normalized value of the cell) for seven similarity measures is depicted on Fig.3.

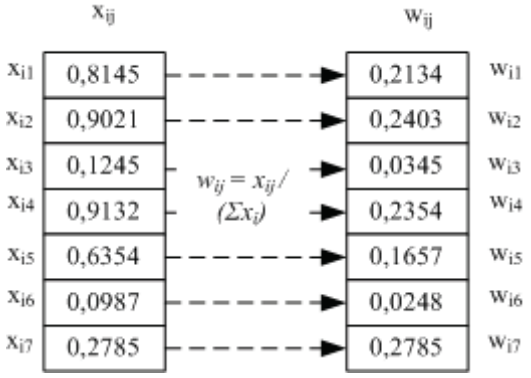


Fig. 3. i-th particle with seven positions is converted to seven weights according to the expression $w_{ij} = \frac{x_{ij}}{\sum_{i=1}^{7} xi}$, $0 \le w_{ij} \sum \le 1$

Initially, for each cell of particles the value from 0 to 1 is randomly selected. Once selected primary swarms, calculated the corresponding values of fitness. The initial velocity of the particles of each cell is zero. The inputs to the proposed method are a swarm of 50 and weighting factors c1 and c2. The threshold value is chosen to be 0.5. The algorithm performed within 30 iterations. The time complexity of the algorithm is square.

B. Objective function

C) The proposed approach works with multiple objective functions: the accuracy and recall of the search. Accuracy is the criterion of correct alignment found in the resulting alignment. Recall is the criterion of finding the right alignment found from a given reference alignment. The criterion of "accuracy" is calculated by the following formula [13]:

$$P = \frac{|A| - |A \cap R|}{|A|}, \quad (13)$$

Recall criterion is calculated according to the formula:

$$R = \frac{|A| - |A \cap R|}{|A|}, \quad (14)$$

As proposed multi-objective Particle Swarm Optimization is implemented as minimization problem so first objective is computed as (1-precision) and second objective is computed as (1-recall).

C. Next Generation Swarm is Produced by Evaluating the Position and Velocity

Each cell or position represents the weight (normalized value of the cell) with respect to the similarity measure. The cells inside the particles contain values from 0 to 1, and the speed of each gene are given zero values. Using the information obtained in the previous step, the position and velocity of each particle of each cluster are updated. Each particle keeps track of the best position it has reached, which is also called pbest. In terms of multi-criteria approach, the position is selected for pbest, whose adaptation of the particle dominates the other devices. And the best position among all particles called global best or gbest. In fact, when the particle moves to a new position at a rate that its position and velocity changes in accordance with equations 15 and 16 [14] [14]:

$$v_{ij}(t+1) = w \times v_{ij}(t) + c_1 \cdot r_1 \cdot \left(pbest_{ij}(t) x_{ij}(t)\right) + \\ + c_2 \cdot r_2 \cdot \left(gbest_{ij}(t) - x_{ij}(t)\right) \quad (15)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (16)$$

where t – a time stamp, *j-th* cluster of *i-th* particle. Velocity $v_{ij}(t+1)$ is calculated by usage of previous velocity $v_{ij}(t)$, *pbest* и *gbest*. Then a new position $x_{ij}(t+1)$ is obtained by adding new velocity with current position $x_{ij}(t)$; c_1 and c_2 are set to 2, r_1 and r_2 are random values from the range from 0 to 1.

After applying non-dominated sorting and crowding distance sorting to the archive, a Local Search is conducted for obtaining the better approximation of weights regarding optimal alignment. The Local Search algorithm is described in algorithm. In the Local-Search algorithm, the best particle replaces the worst particle of the new generation.

IV. Results of Multi-Objective Swarm Intelligence Optimization

The approach optimizes multiple objectives i.e. accuracy and recall is simultaneously optimized. Precision is a measure of correct alignment found from output alignment and recall is a measure of correct alignment found from a given reference

alignment, hi information retrieval particle from the archive. After applying non-dominated sorting and crowding distance sorting to the archive, a Local Search is conducted for obtaining the better approximation of weights regarding optimal alignment. The Local Search algorithm is described in algorithm 3. In the Local-Search algorithm, the best particle replaces the worst particle of the new generation [15].

The proposed algorithm has been applied on a randomly created synthetic dataset. Let us assume two ontologies with the form as depicted on Figure 4.

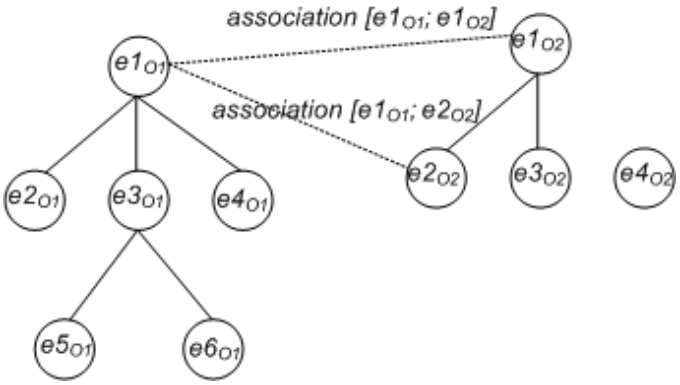


Fig. 4. Example of ontology alignment

It is evident from the figure that ontology O_1 has six components and ontology O_2 has four components. Each component of ontology a has link with every other components of ontology b. As there are four links for every component of ontology a. Hence a total of twenty four pair-wise links are presented by associations. Although only the associations [$e1_{O1}$; $e1_{O2}$] and [$e1_{O1}$; $e2_{O2}$] are shown on Figure 4. The associations are given weight by the similarity value computed from the corresponding components. Then we randomly generate a similarity versus association matrix where seven similarity measures and twenty four associations are considered. The data matrix contains values between 0 and 1. It is assumed that an association is a correspondence if the mean of the seven similarity measures regarding the associations exceeds a threshold value 0.8.

Visualization for the example of ontology alignment is shown on Fig. 5.

From the equations 10 and 11 it is clear that high f-measure as well as low fall-out is always giving the best alignment solution. The maximum F-score generating candidate solution should have highest precision and highest recall.

We've compared the suggested approach with single objective optimization by accuracy and recall (Table 1).

Efficiency of the suggested approach for the criterion accuracy is 0,81428 (high). Again with respect to the f-measure the table shows that our method outperforms other single objective versions. The fall-out for the proposed method is 0 which is less than good for on Single objective (acquisition) and Single objective (recall). Therefore, the proposed method establishes its efficiency

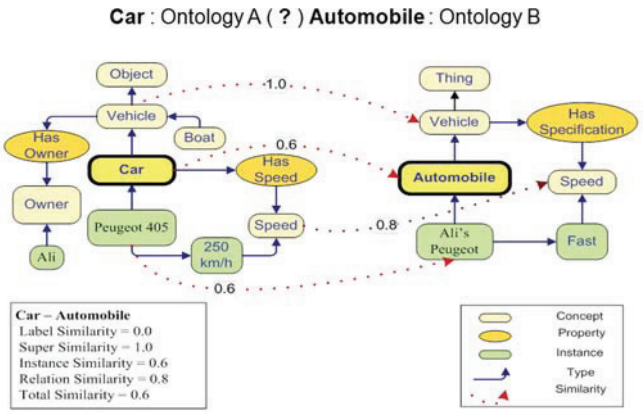


Fig. 5. Visualization of ontology alignment

TABLE I. RESULTS OF EXPERIMENTS

Approach	accuracy	recall	f-measure
Optimization by two objective functions (multi-objective)	0,81428	1,00	0,8976
Optimization by one objective function (accuracy)	0,7142	0,86	0,7803
Optimization by one objective function (recall)	0,3333	1,00	0,0073

V. CONCLUSION

The problem of ontology alignment is to find such a structure and allowed parameters that provide the optimal values for one or more quality criteria.

First, we've studied a single target particle swarm optimization for ontology alignment. Particle Swarm Optimization (PSO) is based on a number of search techniques, inspired by the behavior of natural swarms to solve optimization problems [9]. In Particle Swarm Optimization, particle filters candidate solutions (in our case candidate orientations). Population of these particles is called a swarm. A swarm comprises N particles to move to a D-dimensional search space. By dimensionality we mean the number of similarities in the current alignment. If dimensionality is equal to 0, then a alignment of particle is an empty. The object of the particle swarm optimization is followed by random particle population of the search for optimal solutions initialized by updating the generations.

This strategy increases the search dimension, which is essential for finding the optimal solution. In the presented combined approach the dimension of the individual particles is updated at each iteration. The population size is much larger than the number of possible particle. The above approach has several advantages in terms of good accuracy, refinement of alignments over time and minimizing the human subjectivity. However, it produces only a single solution for ontology alignment, because it is designed as a single objective optimization problem.

For this reason the approach based on multi-objective swarm intelligence approach for generating a local search for

weight vectors was implemented. Then, using these weights different similarity measures are aggregated to improve the ontology alignment problem. Here is an artificial dataset was used to analyze the performance of the proposed technique. Therefore, a comparison between the proposed study and earlier versions is performed. The simultaneous analysis of two quality criteria enables to demonstrate efficiency of our approach for ontology alignment. The algorithm is characterized by square time complexity.

ACKNOWLEDGMENT

This work is performed in Computer sciences and information security institute of Southern Federal University and supported by the Russian Science Foundation grant 14-11-00242.

REFERENCES

[1] Euzenat J., Ontology Matching, Berlin Heidelberg (DE): Springer-Verlag, 2007.

[2] Abolhassani "On Ontology Alignment Experiments," Webology, т. 3, № 3, 2006.

[3] Gargantilla, J. G.-P. A. "OntoWeb: A Survey on Ontology-Based Applications" Deliverable 1.6, 2004.

[4] Euzenat, "Proceedings of the 4th International Workshop on Ontology Matching," в Results of the Ontology Alignment Evaluation Initiative, 2009.

[5] Jurgen Bock. Ontology Alignment using Biologically-inspired Optimisation Algorithms. Dissertation, Karlsruher Institut for Technologie (KIT) Fakultet for Wirtschaftswissenschaften, 2012

[6] Kopaygorodskiy A.N. Primenenie ontologij v semanticheskikh informatsionnykh sistemakh. Ontologiya proektirovaniya №4(14), 2014, p.90-98

[7] Semenova A.V., Kureychik V.M. Application of Swarm Intelligence for Domain Ontology Alignment. Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16), 2016

[8] Sowa J., Knowledge Representation: Logical, Philosophical, and Computational Foundations., Pacific Grove, 2000.

[9] Kennedy, R. E. J. "Proceedings of IEEE International Conference on Neural Networks," в Particle swarm optimization, Perth, Australia, 1995.

[10] Semenova A.V., Kureychik V.M. Domain ontology development for linguistic purposes. 9th International Conference on Application of Information and Communication Technologies (AICT), 2015

[11] Bock, Jurgen and Jan Hettenhausen "Discrete Particle Swarm Optimization for Ontology Alignment." Journal of Information Sciences Vol-19, No-2 pp. 152-173, 2012.

[12] Gladkov L. A., Kurejchik V. V, Kurejchik V. M. i dr. Bioinspirirovannye metody v optimizatsii: monografiya. - M: Fizmatlit, 2009. - S. 384.

[13] D. Sunchez, M. Batet, Ontology-based semantic similarity: A new feature-based approach. Journal of Biomedical Informatics 44, 118-125, 2012

[14] Lapshin, V. Ontologii v komp'yuternykh sistemakh, Moskva: Nauchnyj mir, 2010.

[15] Gladkov L.A, Kurejchik V.V., Kurejchik V.M. Diskretnaya matematika: Teoriya mnozhestv, algoritmov, algebry logiki: Uchebnoe posobie /Pod red. V.M. Kurejchika. - Taganrog: Izd-vo TTI YUFU, 2009. - 312 p.

[16] Kurejchik V. M., Lebedev B. K., Lebedev O. K. Poiskovaya adaptatsiya: teoriya i praktika. - M: Fizmatlit. - p. 272.

Evaluation of Fuzzy Petri nets with the tool TransPlaceSim

Michal Kucharik
Department of Informatics, Faculty of Natural Sciences,
Constantine the Philosopher University in Nitra
Tr. A. Hlinku 1
940 11 Nitra, Slovakia
michal.kucharik@ukf.sk

Zoltán Balogh
Department of Informatics, Faculty of Natural Sciences,
Constantine the Philosopher University in Nitra
Tr. A. Hlinku 1
940 11 Nitra, Slovakia
zbalogh@ukf.sk

***Abstract*—The tool TransPlaceSim was created for secure, easy to use and easy to extend tool to create, simulate and evaluate Petri net models. The paper deals with a new form of evaluation added to this tool which is evaluation of fuzzy Petri nets. In the first part, there are descriptions of Petri nets, fuzzy logic and possibilities of using fuzzy logic in teaching. Consequently, the tool TransPlaceSim is introduced and how to create fuzzy models in it, while the study is applied on an example from the field of teaching. In the end it describes the way this tool evaluates fuzzy Petri nets.**

***Index Terms*—Petri nets, fuzzy logic, modeling, simulation, Petri net tool.**

I. Introduction

For the past few years, the Petri nets (PN) have been successfully applied in various fields. As a tool for system simulation Petri nets are one of the most used methods for modeling and simulating system processes in various areas. They are useful in modeling as they combine discrete events and continuous events in an easily comprehensible manner. On one side, they are appropriate for modeling of discrete parts of systems and to provide discrete abstraction of continuous behavior. On the other hand, generalized stochastic Petri nets can dynamically describe a system and they have good detailed simulating ability [1].

Classic Petri nets are defined as structure N = (P, T, A), where P is a set of places, T is a set of transitions and $A \subseteq (P \times T) \cup (T \times P)$, where $(\forall t \in T)(\exists p, q \in P)(p, t),(t, q) \in A$.

PN represents an oriented bipartite graph containing two types of knots: places and transitions, which are mutually linked by arcs. Each transition may have a non-negative number of input places and a non-negative number of output places. Each place contains a non-negative number of tokens and has a designated maximum capacity. Each arc is reviewed by a non-negative number that represents its weight.

The transition can be activated only if the number of tokens in entry places is greater or equal to weight of arc and the number of tokens plus the arc´s weight is less or equal to capacity of exit place [2] [3].

Fuzzy Petri Nets (FPN) are described as a kind of computer system model with strict mathematical theory support that can describe synchronisation, parallel and fuzzy data. Indeterminancy and fuzziness of real world can be described reasonably and conduct uncertain inferences [4].

FPN is an extension of PN, a bidirectional graph with places and transitions like PN. In FPN, markings are associated with a real value between 0 and 1. FPN is a promising tool for modeling of educational processes and expert systems and is suitable for representation and reasoning of fuzzy knowledge. Authors Z. Balogh and C. Klimeš in their work Modelling of education process in LMS using Petri nets structure have defined FPN structure as a 10-tuple:

FPN = (P, T, D, I, O, f, α, β, λ, HFT), where $P = \{p_1, p_2, ..., p_m\}$ is a finite set of places, $T = \{t_1, t_2, ..., t_n\}$ is a finite set of transitions, $P \cap T = \emptyset$ corresponding to the execution of FPRs, $D = \{d_1, d_2, ... d_m\}$ is a finite set of propositions of fuzzy production rules (FPRs), $P \cap T \cap D = \emptyset$, $|P| = |D|$, d_i (i = 1, 2, ..., n) denotes the proposition that interprets fuzzy linguistic variables, such as "normal", "thorough", "difficult", "extraordinary", as in our model, O is an output incidence matrix, $f = \{\mu_1, \mu_2, ..., \mu_n\}$, where μ_i denotes the CF of R_i, which indicates the reliability of the rule R_i, and $\mu_i \in [0, 1]$, $\alpha : P \rightarrow [0, 1]$ is the function that assigns a token value between 0 and 1 to each place, $\beta: P \rightarrow D$ is an association function, a bijective mapping from places to propositions. $\lambda: T \rightarrow [0, 1]$ is the function which assigns a threshold λ_i between 0 and 1 to a transition t_i, and $HFT = \{hft_1, hft_2, ..., hft_n\}$ is a finite set of HFT, where I(hft) = 1 and O(hft) = 3 [2].

A. Fuzzy logic

The set theory states that a subset A of a set X is defined by a characteristic function c_A that maps elements of X to elements of set {0, 1}. Statement truth $x \in A$ is determined by the ordered pair $(x, c_A(x))$, where $c_A(x)$ is a characteristic function that determines whether x belongs to set A or not. In fuzzy theory, a subset A from a set X is characterized by the membership function m_x that maps elements of X to values from interval [0, 1]. The grade of the statement truth $x \in A$ is determined by a set of ordered pairs $(x, m_x(x))$, where $m_x(x)$ is the grade of membership of x in set A. Fuzzy logic defines a concept of sets with imprecise boundary between the membership and the non-membership elements [4].

Fuzzy knowledge systems are composed of conditional IF-THEN rule groups. For a designer who understands the system it is easy to write the rules and he might add as many rules as it is needed for an adequate system description. Unlike in binary logic, in fuzzy logic the truth is expressed as a level of veracity. Fuzzy systems rely on a member function which calculates the value between 0 and 1. The level of expression veracity is determined by the value between 0 and 1 [5].

Fuzzy IF-THEN rule is a concept used to describe logic dependence between variables in the following way:

IF X1 is A1 AND ... AND Xn is An THEN Y is B

where X1..Xn and Y are variables and A1..An a B are values of the variables. They are often specified linguistically.

The point of the fuzzy logic is to map a group of input to a group of outputs while the main mechanism is the group of IF-THEN rules. All rules are evaluated in a parallel way and random order. The rules are important because they refer to variables and values they acquire. Before building the system that interprets the rules, it is essential to define the concepts that are going to be used and define the adjectives that describe them [5].

II. Tool TransPlaceSim

The program was written in C++ programming language and created with IDE MS Visual Studio 2013. The program has used DirectX 11 API for graphics. A reliable, easy to use and easy to expand tool was needed to be created to simulate and evaluate Petri nets. The basic parts of Petri nets are places, transitions and arcs. To insert these elements into a model, the keys P for places, T for transitions were needed to be pressed. The mouse has been clicked where elements had to be inserted. To connect elements with each other, the key A was pressed and a place or a transition was clicked to start creating an arc. Then the element was clicked to which this arc should point.

It was possible to zoom and to shift the view of the model with the middle mouse button. With tabs View and Settings one could change the appearance of the model and its environment. As standard, a name display and a current number of tokens to capacity had switched on. For transitions it was its name and for arcs its weight.

By clicking with mouse on a specific element (place, transition or arc) properties window would open in the upper right corner.

Properties for places were its name, start token count, actual token count, capacity, position (x, y) and sliders which allowed setting the place name position and place name distance.

For transitions its name, time type, position and sliders which allowed setting the transition name position and transition name distance. Time types determine whether transition should be activated when it met the activation conditions. Immediate transition has been activated. Deterministic transition has been activated only if no other transitions could be activated in that step. Stochastic transition has been activated only if its random time dropped to zero in that step.

For arcs it was the weight that specifies how many tokens should have been removed or added to places, type of arc, information about what place and transition this arc connects and sliders which allowed setting the arrow data position and arrow data distance. Arc types can be standard, inhibitive and test types. A standard one removed or added tokens according to the arc´s weight after transition activation. Inhibitive arc could point only from place to transition, did not remove tokens from place and allowed transition activation only when its weight was higher that the number of tokens in a place. Test arc was used only to test whether a transition can be activated, but did not remove nor add tokens to places.

Thanks to these element properties it was possible to create nearly all Petri net types in this program except for colored Petri nets.

When selecting Simulation tab, simulation menu was displayed. By pressing the Reset simulation button the actual token number of each place was set to starting number of tokens. By pressing the Step forward button, one simulation step has been completed. Simulation started by pressing the Play button. The running simulation has been highlighted with a red window border. Simulation has been stopped by pressing the Stop button. The slider served to set the simulation animation speed.

During simulation, transitions have been activated according to activation rules derived from Petri nets definition and according to transition time types.

A. Model evaluation

The tool TransPlaceSim enabled several evaluation methods of created models. By clicking on Formalisation tab, menu with all possible evaluations has been displayed. The new evaluation added to this tool was Display Fuzzy conditions. With the evaluation it has been required to create a correct model and to specify input and output places. Then the program has created IF-THEN rules for this model according to the model and defined input and output places.

Different articles have described different representations of fuzzy Petri nets. Some used transitions as conditions where input places were input variables, arcs represented values of these variables, output places were output variables and output arcs were represented by values of these variables [6]. In the program TransPlaceSim a different representation has been used. The group of places represented the input variables, and the transitions that used these places as input represented the values of these variables. The second group of places represented the values that this model may acquire [1].

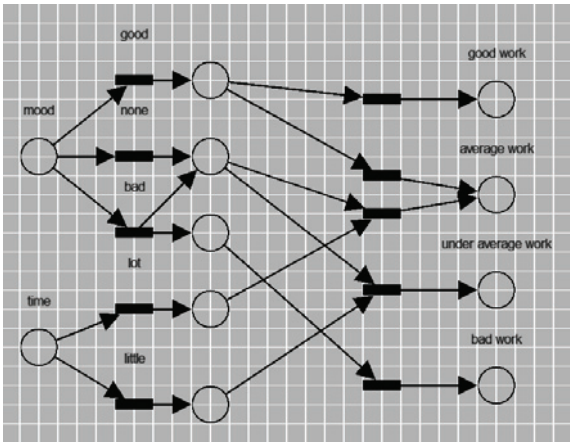


Fig. 1. Example of a fuzzy Petri net.

Thanks to this approach it was possible to create a multi-level fuzzy models that may contain not only AND rules but also OR rules and thanks to inhibitive arcs negations as well (NOT, NAND and XOR) [10].

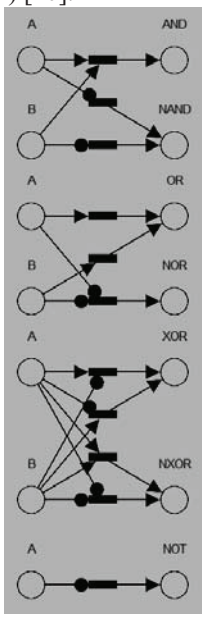


Fig. 2. Example of logic functions with Petri nets

To get fuzzy rules from the model it had to meet several conditions.

- Output places may not have output arcs.
- No loops may be presented in the model.
- Arc´s weight can only be 1 and arcs can be only standard or inhibitive.
- Place capacity must be 1.
- Transitions cannot be stochastic or deterministic.

B. Calculation of possible input combinations

As mentioned above, when evaluating fuzzy models all possible input combinations were written. This process could become slow because with a greater number of input variables and values they may acquire, the number of combinations was exponentially increasing.

A number of all possible combinations has been the result when the numbers of all values have been multiplied. These variables can acquire $k = \prod x_i$. Where x_i is number of all transitions that came out of i-th place.

III. Evaluation of Educational Processes: Case Study

In the article Possibilities of modeling web-based education using IF-THEN rules and fuzzy Petri nets in LMS and in the book Modelovanie procesov pomocou Petriho sietí is an example of Fuzzy Petri net that describes modeling of educational processes specifically modeling of students´ knowledge that we used as an example [11] [12].

Serial machines are suitable for the description of the teaching process behavior, however, they have several limitations. Petri nets that originated on the ground of modeling possibilities of serial machines are used for these given purposes.

Formal description complemented by a visual graphic depiction is one of the advantages of Petri nets, therefore, we would use them to model teaching processes. Petri nets allow designing an accurate specification of teaching process and also allow removing ambiguity, vagueness and contradiction from the model. Moreover, Petri nets have also defined mathematical base that can be used in different software tools for teaching process analysis [7].

These articles describe the creation of a model from designed rules. In TransPlaceSim we can get back rules from a model.

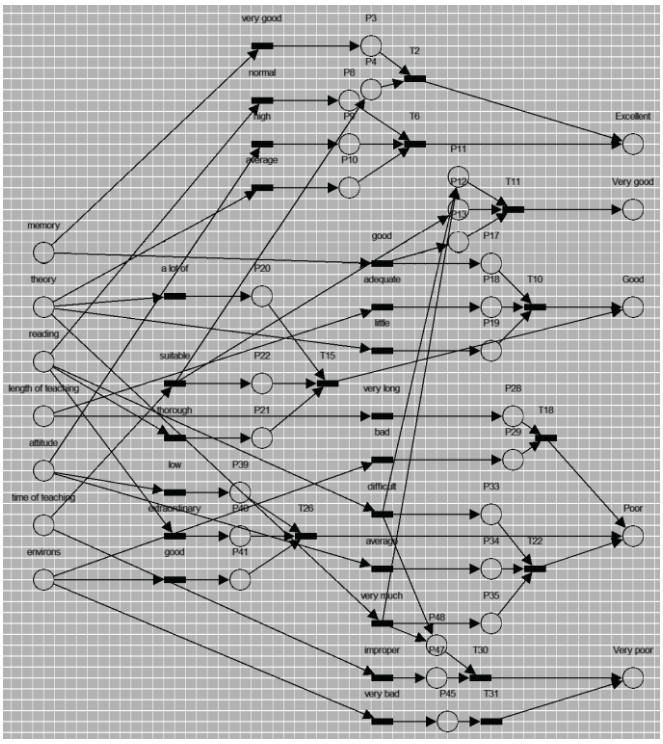


Fig. 3. Fuzzy Petri net with Knowledge output

In the example, the creation of IF-THEN rules has been mentioned on which authors have created a model of a fuzzy Petri net. Output of this model was the knowledge which has

been defined with 10 rules. When the model has been created in TransPlaceSim and fuzzy condition evaluation was used, output for all possible combinations of input values were gained and with this, the outputs that authors have not defined.

Knowledge defined in example [5]:

IF (environs = very bad AND concentration = low) THEN knowledge = very poor

IF (environs = bad AND length of learning = very long) THEN knowledge = poor

IF (environs = good AND reading = extraordinary AND attitude = low) THEN knowledge = poor

IF (time of teaching = suitable AND reading = thorough AND theory = very much) THEN knowledge = good

IF (reading = difficult AND theory= a lot of AND attitude = average) THEN knowledge = poor

IF (memory = good AND length of learning = adequate AND theory = adequate) THEN knowledge = good

IF (time of teaching = suitable AND memory = good AND theory = adequate) THEN knowledge = very good

IF (time of teaching = very short AND reading = difficult) THEN knowledge = very poor

IF (motivation = high AND memory = very good AND time of teaching = suitable) THEN knowledge = excellent

IF (reading = thorough AND attitude = adequate AND theory = adequate) THEN knowledge = excellent

After evaluation in TransPlaceSim one of the conditions was:

IF (motivation = high AND memory = very good AND reading = difficult AND attitude = average AND theory = very much AND time of teaching = suitable AND length of teaching = very long AND environs = very bad AND concentration = low) THEN (1.00*(Excellent AND Poor AND Very poor)).

The result has been achieved because some conditions did not exclude each other. Combination of these rules could result in all of the outputs already learned. Another strange result has been achieved for:

IF (memory = good AND reading = normal AND attitude = average AND theory = average AND time of teaching = suitable AND length of teaching = adequate AND environs = bad) THEN (1.00*(Unspecified)).

The term unspecified means that the combination of inputs did not result to any output or that network got stuck before a token reached one of the output places. This state can occur if the author of the model does not think of all possible input combinations.

All possible combinations of input values resulted from the number of input variables and values they may acquire.

Memory = {very good, good}

Theory = {average, a lot of, little, very much}

Reading = {normal, thorough, difficult, extraordinary}

Length of teaching = {adequate, very long}

Attitude = {low, average, high}

Time of teaching = {suitable, improper}

Environs = {good, bad, very bad}

If in this case all values were multiplied, input variables would acquire ($k = \prod x_i$), number of combinations was possible k = 2*4*4*2*3*2*3 = 1152. With this number of possible combinations calculation took 3 min 9 sec.

A. Evaluation algorithm

The program first needed to find out whether the model met the conditions to consider it to be a fuzzy model. Then it calculated the maximum number of steps required to reach the final level of fuzzy net. The number of steps was required for a method that calculated the probabilities of individual steps. The program then created all possible combinations of input values. Finally, for every combination of input values the combinations were written as well as the output results for these combinations. Pre-processing and evaluation of data can be solved by different algorithms [13].

IV. Discussion

The object of the research was a creation of a system model which is one of the most spread activities of people from the beginning of their existence. Generally, it can be said that our vision of the world is a model of the reality that is the outside world which can be explored. When modeling, we refer to accessible information. The created model represents the formalized knowledge about the modeled system from the exploration point of view and it usually covers that part of the description which is needed for the given purpose. The model always draws upon a certain subset of our knowledge that is incomplete. It is possible for us to model only what we are able to understand and describe. Then simulative modeling represents a process of knowledge transformation from a mechanically non-acceptable representation to a representation acceptable by a computer.

When simulating arbitrary processes with Petri nets, occasionally, there is a need to display the state which might or might not occur. To create such a network, it is possible to draw upon the standard logic and fuzzy logic. The advantage of a fuzzy logic application into Petri nets is an easy comprehensibleness and a sophisticated mathematical tool, a relatively simple design and a modularity of process modeling solution.

The tool TransPlaceSim to model Petri nets has been expanded by IF-THEN rules evaluation. The application has checked all the conditions of the model in order to consider it to be a fuzzy model. The tool calculated which number of steps was needed to end the fuzzy nets. From the number of steps needed to end the nets, the likelihood of individual state achievement has been calculated. It has resulted in all possible combinations of output values. It is possible to use the tool in all specific processes, for instance the process of students´ education, or students´ knowledge based on certain input information. Rules might be possible to use in designed expert systems.

V. Conclusion

There are several tools for PN simulation but none of them has formalisation ability. The aim of this paper was to describe the creation of modeling and simulating tool for Petri nets and its Fuzzy module through which it would be possible to model,

simulate, evaluate and create IF-THEN rules based on defined criteria. It also describes the specific example of student´s knowledge modeling and its states in which it may be found. Based on the achieved results, it can be noted that the tool TransPlaceSim was functional and it was creating rules from the designed IF-THEN model. In the future, there are plans to develop the model by other possibilities in order to model and consequently simulate educational processes.

In the future we would like to expand this program for more options and functions so that we can model and simulate educational processes. We would use TransPlaceSim to create a model of educational environment Moodle used on our university. We would simulate student usage of this environment, use Fuzzy functions to monitor their test results and then improve courses, tests and overall curriculum. First course that this should be used on is the course Operating systems.

ACKNOWLEDGMENT

This publication is published thanks to the financial support of the project UGA VII/8/2016 named: Modelovanie edukačných procesov pomocou Petriho sietí; Educational process modeling using Petri nets.

REFERENCES

[1] C. Klimeš and Z. Balogh, *Modelovanie procesov pomocou Petriho sietí*. Nitra: Univerzita Konštantína Filozofa v Nitre, 2012.

[2] Z. Balogh and C. Klimeš, "Modelling of education process in LMS using Petri nets structure," in *IADIS International Conference e-Learning 2010, Part of the IADIS Multi Conference on Computer Science and Information Systems 2010, MCCSIS 2010*, Freiburg, 2010, pp. 289-291.

[3] Z. Balogh, M. Turcani, and M. Magdin, "The possibilities of using Petri Nets for realization of a universal model of educational process," in *2013 IEEE 14th International Conference on Information Reuse and Integration, IEEE IRI 2013*, San Francisco, CA, 2013, pp. 162-169.

[4] D. Wu and H. Yan, "Structure model of CDIO ability evaluation expert system based on fuzzy petri net," in *2012 International Conference on Information Technology and Management Innovation, ICITMI 2012* vol. 263-266, ed. Guangzhou, 2013, pp. 2164-2167.

[5] C. G. Looney, "Fuzzy Petri nets for rule-based decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 18, pp. 178-183, 1988.

[6] R. I. Hamed, S. I. Ahson, and R. Parveen, "Designing genetic regulatory networks using fuzzy Petri nets approach," *International Journal of Automation and Computing,* vol. 7, pp. 403-412, 2010// 2010.

[7] A. Sengupta and T. K. Pal, *Fuzzy Preference Ordering of Interval Numbers in Decision Problems* vol. 238. Berlin: Springer-Verlag Berlin Heidelberg, 2009.

[8] V. Novák and S. Lehmke, "Logical structure of fuzzy IF-THEN rules," *Fuzzy Sets and Systems,* vol. 157, pp. 2003-2029, 8/1/ 2006.

[9] Y. Hsin-Tai and L. Shie-Jue, "Inference and improper knowledge detection on fuzzy rules with enhanced high-level fuzzy Petri nets," in *Fuzzy Information Processing Society - NAFIPS, 1998 Conference of the North American*, 1998, pp. 135-139.

[10] Š. Koprda and M. Magdin, "Fuzzy adaptive controller design for control of air in conditioned room," in *11th International Conference on Intelligent Computing, ICIC 2015* vol. 9226, A. Hussain, D. S. Huang, and K. H. Jo, Eds., ed: Springer Verlag, 2015, pp. 318-329.

[11] Z. Balogh and M. Turcani, "Possibilities of Modelling Web-Based Education Using IF-THEN Rules and Fuzzy Petri Nets in LMS," in *Informatics Engineering and Information Science, Pt I*. vol. 251, A. AbdManaf, A. Zeki, M. Zamani, S. Chuprat, and E. ElQawasmeh, Eds., ed, 2011, pp. 93-106.

[12] Z. Balogh, M. Magdin, M. Turcani, and M. Burianova, *Interactivity Elements Implementation Analysis in E-Courses of Professional Informatics Subjects*, 2011.

[13] M. Munk, Ľ. Benko, M. Gangur, and M. Turčáni, "Influence of ratio of auxiliary pages on the pre-processing phase of web usage mining," *E a M: Ekonomie a Management,* vol. 18, pp. 144-159, 2015.

Application of Data Mining Technology to Optimize the City Transport Network

Irina Makarova, Rifat Khabibullin, Vadim Mavrin, Danila Tikhonov, Denis Davletshin, Ksenia Shubenkova
"Service of transport systems" department
Kazan Federal University
Naberezhnye Chelny, Russia
kamIVM@mail.ru

***Abstract*—The article considers the experience of OLAP-technology in solving optimization problems of the city transport system. The authors proposed a conceptual decision support scheme to optimize the transport-road network parameters. The developed system provides the possibility of intelligent analysis and decision-making based on OLAP-technologies and simulation models.**

***Index Terms*—intelligent analysis, OLAP, transport network, simulation modeling.**

I. Introduction

Urban mobility is rapidly becoming one of the greatest challenges facing developed and developing countries. Transport is estimated to be responsible for nearly a quarter of global energy-related CO2. There are also rising concerns about its impact on the quality of urban life, including social inequities, and about the effects of its pollution on health and buildings.

Overall demand for transport activity (for both passenger and freight) is growing rapidly, and it is predicted to roughly double between 2005 and 2050. The global vehicle fleet is set to multiply three or four-fold in the next few decades, with most of this growth set to occur in developing countries. There is a growing consensus on the need for more sustainable patterns of transport activities [1]. Achieving effective and long-lasting improvements in road safety has been attained in a number of countries that have adopted a broad approach addressing many dimensions of road safety.

As it is indicated by the WHO Global Status Report on Road Safety 2015, plateauing of the road traffic deaths despite a 4% increase in global population and a 16% of motorization in the past few years shows that implementation of road safety measures has led to saving lives. It was stated that almost half of all deaths on the world's roads are among those of the least protected – motorcyclists (23%), pedestrians (22%) and cyclists (4%). As it is pointed out in the WHO report, not enough attention is paid to the needs of pedestrians, cyclists and motorcyclists, although the number of road traffic deaths among them achieves 49% worldwide [2].

Considerable growth of the city traffic intensity leads to a deterioration of traffic conditions, leads to an increasing number of road accidents and accident victims, and, generally, to the rapid deterioration of the city transport service quality. According to statistics, over 60% of all accidents occur in cities and in other settlements. Moreover, 30% of all accidents concentrate at the city. Urban traffic control schemes often have to be modified in response to the following factors: emerging of new attraction centers and residential areals, reconfiguration of road-street networks and public transport route network. The most important thing is to consider these factors at a constant growth of motorization. In these cases, the rational administrative decision can be received by means of intelligent transportation systems, including, simulation models and OLAP-technologies [3].

II. Decision Support System for City's Transport System Management

Since the quality of decisions in managing of large systems depends on the quality of information, adequately selected methods of its analysis, and effective tools, this calls for creating of decision support systems (DSS).

Good governance not only improves the economic performance of the system, but also need for solving social problems by improving public transport services and to reduce the negative impact of the transport complex on the environment.

The structure of DSS essentially depends on the kind of objectives to be tackled, on the available data, information and knowledge and also on the system's users. Therefore a DSS consists of three main parts: (1) a data system for collecting and storing of information obtainable from internal and external sources; as a rule it is a data storage, (2) a dialogue system affording the user to set the data to be selected and methods for their processing, and (3) a system of models, i.e. ideas, algorithms and procedures permitting to process and analyze the data. The user has a certain experience, is aware of the situation and is motivated by certain considerations when selecting the data. The data are processed using various procedures, ranging between mere search to statistical analysis and nonlinear optimization.

Since decision making is based on the real data of the object under management, both analysis and adopting of strategic decisions require aggregate information available from a specially created data storage (DS). DS are formed for integrating, actualizing and coordinating of the operative data from heterogeneous sources in order to arrive at a single, non-

controversial view of the object as a whole. DS contain the information collected from several operative databases of an on-line transaction processing systems (OLTP). The core of a road situation control center is a multi-dimensional intelligent data model (an OLAP cube) which collects, stores and formalizes the road network parameters (Fig. 1 a,b). An OLAP cube is a term that typically refers to multi-dimensional array of data. The term cube here refers to a multi-dimensional dataset, which is also sometimes called a hypercube if the number of dimensions is greater than 3.

In database theory, an OLAP cube is an abstract representation of a projection of a relational database management system RDBMS (relational database management system) relation. Given a relation of order N, consider a projection that subtends X, Y, and Z as the key and W as the residual attribute. Characterizing this as a function, F(X, Y, Z) → W, the attributes X, Y, and Z correspond to the axes of the cube, while the W value into which each (X, Y, Z) triple maps corresponds to the data element that populates each cell of the cube.

Storing of information as an OLAP cube and its subsequent processing will make it possible to precisely assess the dynamics of street road network (SRN) parameters in different dimensions (number of transport vehicles, road section, season of the year, average speed, availability of traffic lights, etc.). By analyzing the information on varying road parameters within the day time, week day, and month it is possible to forecast probable changes in road situations in the future.

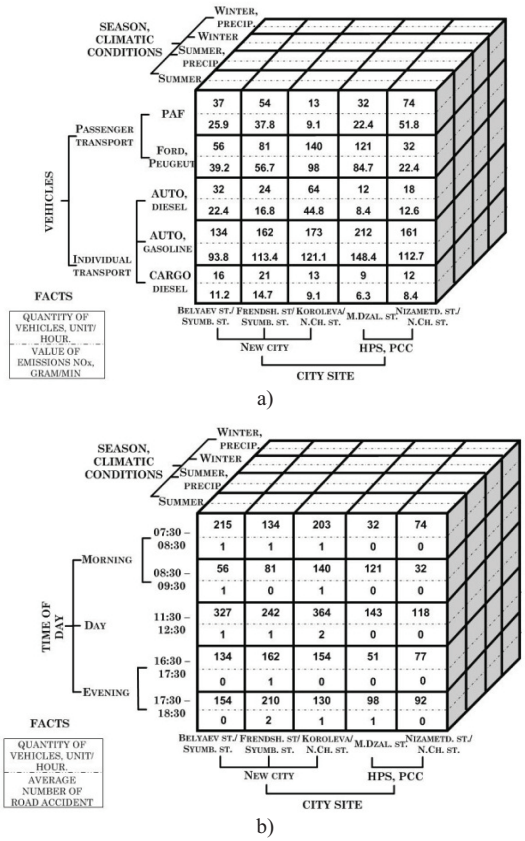


Fig. 1. OLAP cube: a) the "kind of transport" dimension; b) the "time of day" dimension.

Besides, modeling of variants of possible solutions with varying system parameters allows selecting the optimal parameters and creating a database of best solutions at fixed parameters of the transport flow and external environment. Such bases serve for operative decision making in the case of transport emergencies. The intellectual core of the DSS is often composed by simulation models, which affords not only to make a qualitative analysis of the processes but to investigate the consequences of variations and select the variant satisfying all preset limitations to obtain the system parameters optimal for the preset conditions [3].

The software, developed for this purpose, allows to construct models reflecting the processes as they actually occurred and then to conduct a series of virtual experiments by setting the model time. In so doing the process can be investigated both in a single test and by setting a certain multitude of tests, thus selecting the optimal process parameters. The results will be determined by their random character. To select the optimal management of transport systems by DSS, series of simulation models experiments are conducted. Such experiments allow predicting the consequences of modifications taking into account many factors that reflect the trends of economic system development of the city and region [4, 5].

III. OLAP-Technology When Processing Big Data

For safe functioning of transport system the infrastructure, in particular the street-road network (SRN) needs advance development, or, at least, it should correspond to changes of automobilization level. As the international experience shows, the problem of formation of many kilometers congestions and traffic jams often is result of irrational management of transport system.

It is very hard to predict traffic intensity changes caused by reconstruction of sections of roads, construction of new centers of passengers' attraction and new residential districts. The reason is that it is necessary to consider a large number of stochastic factors. In most cases, the reconstruction of the SRN is carried out without taking into account the growth of traffic density and intensity flow. Therefore such measures as expansion of the road network, construction of new turns and additional lanes do not have any positive effect and in some cases lead to a deterioration of the situation on the roads. A comprehensive solution of these problems can be provided by the use of intelligent transport systems (ITS), both through information services, and through the means and methods to perform data mining and make decisions based on it.

It is evident that the ITS plays an important role in optimization of the transport system operations [6]. It promotes the sustainability of the environment [7], diminishes the negative impact of the transport complex on the environment and also the energy consumption [8]. Nowadays the ITS has turned into tool in transport planning being used for surveying [9], decreasing of traffic congestions [10], and planning of shared vehicle use [11].

To identify origins of adverse situations in SRN the full-scale researches are conducted, after then the data are analyzed.

At the same time fast access to data is provided by means of OLAP-technology. Multidimensional data model (OLAP) is used for the accumulation and processing large amounts of operational information that is a basis for making management decisions.

The study of the transport flow parameters allows determining the condition of the road network and the problems that must be solved to improve the situation on city roads: congestion of traffic lanes; average flow rate in the section of the street-road network; traffic density; flow rate; ecological parameters of the environment near the roadways and highways.

IV. Interface's Design for Olap-Requests Implementation

Data storage in the form of an OLAP-cube allows quickly and with a high accuracy receiving dependence of parameters of the movement on different factors (the number of vehicles, a road section, a season, average speed, existence of the traffic lights, etc.).

The set of attributes, that are the part of an OLAP-cube, is conditionally divided into 2 groups: (1) Dimension table that are used as criteria for the analysis and form multidimensional space of an OLAP-cube. Sections (hyperplanes) of a cube are received by means of fixing of values of measurements. Each section represents the certain data query including aggregations, and (2) Fact table – functions which serve for formation of queries.

The elements of a dimension can be organized as a hierarchy, a set of parent-child relationships, typically where a parent member summarizes its children. Parent elements can further be aggregated as the children of another parent.

Conceiving data as a cube with hierarchical dimensions leads to conceptually straightforward operations to facilitate analysis. Slicer is a term for a dimension which is held constant for all cells so that multi-dimensional information can be shown in a two-dimensional physical space of a spreadsheet or pivot table. Each cell of the cube holds a number that represents some measure of the research areas, such as: traffic intensity, the number of accidents and average speed of vehicle flow.

The two basic hierarchical operations when displaying data at multiple levels of aggregations are the “drill-down” and “roll-up” operations. Drill-down refers to the process of viewing data at a level of increased detail, while roll-up refers to the process of viewing data with decreasing detail [12]. Besides, such operations are applied as: slice - this operation selects one particular dimension from a given cube and provides a new sub-cube; dice – this operation selects two or more dimensions from a given cube and provides a new sub-cube; pivot – this operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. In our case as the set of attributes were taken such indicators as: city, area and street, time of year and time of day. These attributes allow most clearly showing the situation that can occur in areas of SRN.

Multidimensional Expressions (MDX) is a query language for OLAP databases, which provides a specialized syntax for querying and manipulating the multidimensional data stored in OLAP cubes [13]. MDX has become the standard for OLAP systems.

```
select{
[Season].[Times].[All].[Spring] on colums
[City].[Area].[Street].[All].Children on rows
}
from [intensity]
```

The following example shows a query that uses the select statement. This query returns a result set that contains the road traffic intensity in the chosen season and days, and also on the selected streets and avenues (Fig. 2).

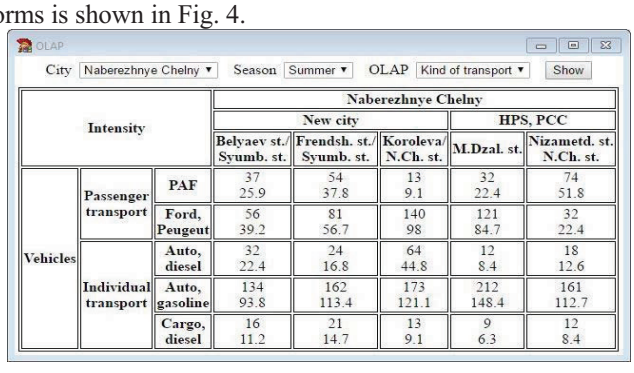
V. The Module Architecture and Its Implementation

One of the main conditions of successful management is up-to-date and qualitative information. For work with data the special interface has been created. It provides information input and its administration. Besides, users can create queries and obtain the necessary data for the intellectual analysis. The architecture of this system is presented in Fig. 3.

For implementation of OLAP cube the hybrid option has been chosen. Hybrid OLAP is a combination of both ROLAP (Relational OLAP) and MOLAP (Multidimensional OLAP). It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allows storing the large data volumes of detailed information.

System administrator is responsible for input and update of information. He refreshes data in dimension tables, and also adjusts the fact table, if new queries are necessary for users. The data for multi-dimensional analysis you can obtain directly from the database. The special utility of formalization of the chosen data array is used for this purpose. Reporting Services (SSRS) allows creating reports for a large number of data sources. This service has a full range of tools for creation, management and delivery of reports. SSRS has API interfaces that help developers to make integration or to expand possibilities of data processing and reports in user applications. The Reporting Services tools are completely integrated with the SQL Server tools and components [14].

This architecture was a basis for creation of the program module which allows output data on all possible measurements of an OLAP-cube for the intelligent analysis. One of such forms is shown in Fig. 4.



Intensity			Naberezhnye Chelny				
			New city			HPS, PCC	
			Belyaev st./ Syumb. st.	Frendsh. st./ Syumb. st.	Koroleva/ N.Ch. st.	M.Dzal. st.	Nizametd. st. N.Ch. st.
Vehicles	Passenger transport	PAF	37 25.9	54 37.8	13 9.1	32 22.4	74 51.8
		Ford, Peugeut	56 39.2	81 56.7	140 98	121 84.7	32 22.4
	Individual transport	Auto, diesel	32 22.4	24 16.8	64 44.8	12 8.4	18 12.6
		Auto, gasoline	134 93.8	162 113.4	173 121.1	212 148.4	161 112.7
		Cargo, diesel	16 11.2	21 14.7	13 9.1	9 6.3	12 8.4

Fig. 2. Example of formation of query.

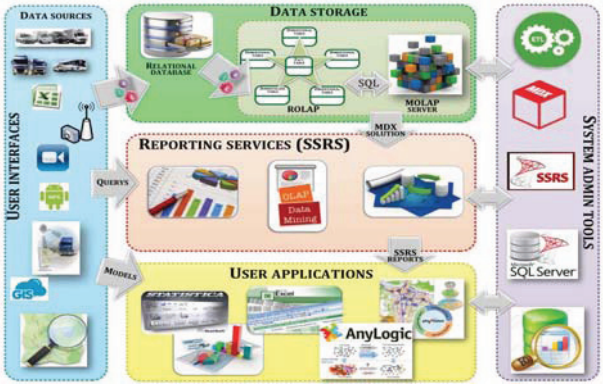


Fig. 3. The system's architecture.

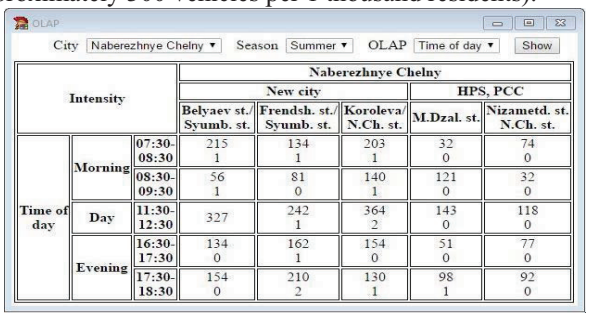
In upper part of a form it is possible to choose attributes which are necessary for receiving a data slice. Such attributes are: city, site and season. Selections of attributes can be varied for display of the necessary data slice. You can save these data for later use in the simulation model.

You can also use the technique of data analysis, which is based on an integrated use of technology online analytical mining knowledge and data mining methods, allowing expanding the analytical capacity of the system for decision support [15, 16]. The use of intelligent data analysis methods in parallel with the OLAP allows obtaining some new results.

There is the possibility of using a variety of Data Mining methods in relation to the collection of records to find similar values [17]. It will allow, for example, if the array of reports on certain periods of time is created, to find common features between the current situation and a situation which was already found earlier. So, you can evaluate efficiency of decisions in the past and predict possibilities of future projects. It will allow increasing quality of decision-making in the current time. For example, after the choice of year, month or date - you can look what was a situation in a certain place and to predict it in the future.

VI. The Choice of Data and Carrying out Experiments on the Simulation Model

An example of such solutions can serve as the project to develop the ITS for the city of Naberezhnye Chelny, the second largest and population city of the Republic of Tatarstan, a major industrial center, an important transport junction of the Republic. As far as Naberezhnye Chelny is the developing city, it is characterized by the increased level of motorization (approximately 300 vehicles per 1 thousand residents).


OLAP

City: Naberezhnye Chelny | Season: Summer | OLAP: Time of day | Show

Intensity			Naberezhnye Chelny				
			New city			HPS, PCC	
			Belyaev st./ Syumb. st.	Frendsh. st./ Syumb. st.	Koroleva/ N.Ch. st.	M.Dzal. st.	Nizametd. st. N.Ch. st.
Time of day	Morning	07:30-08:30	215 1	134 1	203 1	32 0	74 0
		08:30-09:30	56 1	81 0	140 1	121 0	32 0
	Day	11:30-12:30	327	242 1	364 2	143 0	118 0
	Evening	16:30-17:30	134 0	162 1	154 0	51 0	77 0
		17:30-18:30	154 0	210 2	130 1	98 1	92 0

Fig. 4. Application form.

According to its purpose and transport loads, one of the "problem" areas of Naberezhnye Chelny city it is possible to allocate the intersection of the North-Eastern suburb of Druzhba Narodov (Fig. 5). Average speed of the traffic on this section is 35-40 km/h. The intersection has a complex configuration in view of the fact that Avenue of Druzhba Narodov is at a virtually perpendicular position relative to three major avenues (Moscow, Mira and Syuyumbike), at the same time it has two intersections with Prospekt Syuyumbike, forming a triangular junction configuration. Besides, there are three turns on the section. Thus, the plot has six crossing points with a counter-flow traffic.

According to statistical information of traffic safety department, this is a place of concentration of road accidents. In peak time this site does not cope with a transport flow: congestion occur, often there are serious traffic jams. Drivers spend a lot of time for maneuver on turns because of a huge flow of vehicles. These dangerous problems arise in view of the non-optimal organization of transport flows. The situation was aggravated in connection with input of a new site of a tram route network. Application of the OLAP technology to work with data on movement parameters, allows making the detailed analysis of the stretch of road (Fig. 6).

Results of experiments on model have shown that the geometry of the studied stretch of road negatively affects the traffic characteristics, because it is not corresponding to traffic flow parameters. To improve the situation, we proposed a new version of optimized configuration of the intersection of Avenues Syuyumbike and Druzhba Narodov. It was offered to organize a circular motion that allows to reduce number of conflict points (elimination of turns) on this road section. Besides, we have suggested to apply at the adjacent intersection (Avenues Druzhba Narodov and Mira) traffic light regulation with alternative number of phases, i.e. set of the main and intermediate traffic light steps.

The analysis of results of research of this site has shown that the most acceptable option is the organization on a site of a circular motion. Table 1 summarizes the traffic light stage changes with the adaptive control, depending on the traffic density. The research results of the studied crossroad are shown in Table 2. These data demonstrate that parameters of traffic of studied stretch of road could be greatly improved.

VII. Conclusion

The research findings have demonstrated a considerable influence of motor transport to urban quality of life. It has been established that the problem should be approached comprehensively.

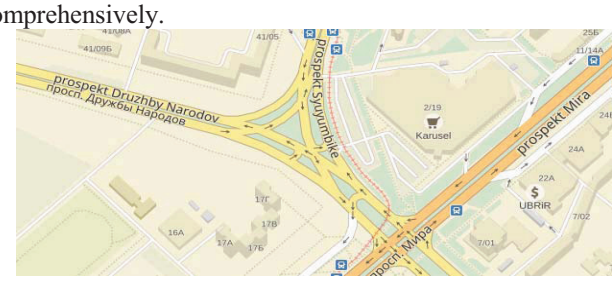
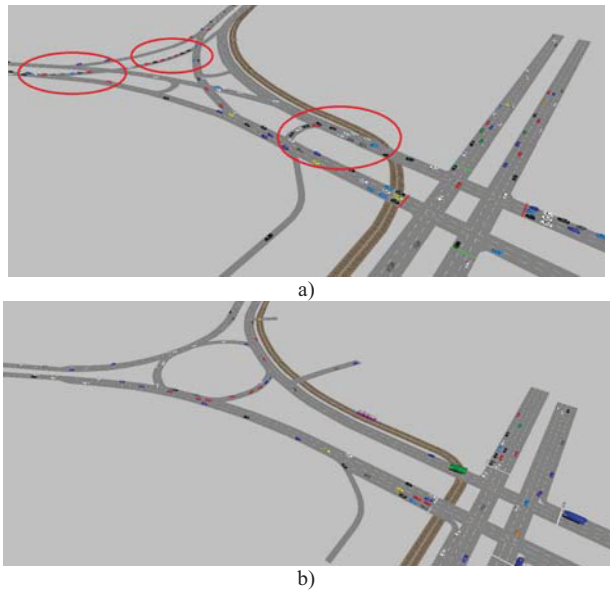


Fig. 5. Scheme of the simulated junction of the road network.



a)

b)

Fig. 6. Simulation model a) before optimizations; b) after optimizations.

Design and technological solutions to improve the environmental performance of vehicles should be combined with the rational management of transport complex. The growing volumes of data increase the importance of analyses, most of which are built on OLAP principles. Application of data mining and simulation modeling to optimize the transport network will allow for more timely and efficiently to solve problems related to road network. It has been established that constructive and technological decisions and transport complex rational management should be approached comprehensively. Simulation modeling can help to identify the optimal parameters for the transport flow and find rational managerial solutions.

TABLE I. TRAFFIC LIGHT PHASE CHANGE ON THE CROSSROAD DEPENDING ON THE TRAFFIC DENSITY

Traffic Density of the Site	Total Stage Duration, sec.	Red Signal (Basic Tact), sec.	Green Signal (Basic Tact), sec.	Red and Yellow Signal, sec.	Yellow Signal, sec.
95%	85	41	38	3	3
82%	83	37	40	3	3
74%	82	35	41	3	3
61%	81	32	43	3	3

TABLE II. CALCULATED PARAMETERS OF STUDIED STRETCH OF ROAD NETWORK

Designation of Parameter	Value Prior to Any Changes	Value after Changes
Average speed along studied stretch of road, km/h	35	43
The number of stops per unit time, pcs.	6	2
Traffic flow density, % of road area	92	67
Average duration of travel on the stretch of road, min.	4	1,4

REFERENCES

[1] Sustainable Transport and Air Pollution. http://www.unep.org/resourceefficiency/Policy/ResourceEfficientCities/FocusAreas/SustainableTransportandAirPollution/tabid/101667/Default.aspx

[2] World Health Organization. Global status report on road safety, 2015: Summary. Geneva: WHO Press. 2015. 16 p.

[3] I. Makarova, et al., "Urban transport system management in the context of region sustainable development strategy", in Transport Problems, vol. 8, 2013, pp. 107-111.

[4] I. Makarova, et al., "Intellectualization of transport systems for the benefit of safety and the sustainable development of territories", in Journal of International Scientific Publications: Ecology&Safety, vol. 7, 2013, pp. 189-199.

[5] I. Makarova, et al., "Increase of City Transport System Management Efficiency with Application of Modeling Methods and Data Intellectual Analysis", in Intelligent Transportation Systems – Problems and Perspectives. Springer, 2015, pp.37-80.

[6] J. Xia & M. Chen "Defining Traffic Flow Phases Using Intelligent Transportation Systems-Generated Data", in Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, vol. 11, 2007, pp. 15-24.

[7] F. Zhou & S. Jun "Deploying an Intelligent Transportation System in Chongming County", in Shanghai Journal of Urban Technology, vol. 17, 2010, pp. 39-51.

[8] K. Gkritza et al., "Intelligent Transportation Systems Applications for the Environment and Energy Conservation (Part 1)", in Journal of Intelligent Transportation Systems: Technology, Planning, and Operations. vol. 17, 2013. pp. 1-2.

[9] M. R. Tayyaran et al., "Impact of telecommuting and intelligent transportation systems on residential location choice", in Transportation Planning and Technology, vol.26, 2003, pp.171-193.

[10] R. Harb, et al., "Two Simplified Intelligent Transportation System-Based Lane Management Strategies for Short-Term Work Zones", in Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, vol.15, 2011, pp. 52-61.

[11] T. Gärling, et al., "Adaptation of Private Car Use in Response to Travel Demand Management Measures: Potential Roles of Intelligent Transportation Systems", in Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, vol. 8, 2004, pp. 189-194.

[12] M. Ceci, A. Cuzzocrea, D. Malerba "Effectively and efficiently supporting roll-up and drill-down OLAP operations over continuous dimensions via hierarchical clustering", in Journal of Intelligent Information Systems, 2015, vol. 44, pp. 309-333.

[13] A. Ghrab, et al., "A framework for building OLAP cubes on graphs", in Lecture Notes in Computer Science, 2015, vol. 9282, pp. 92-105.

[14] Reporting Services Reports (SSRS). URL: https://msdn.microsoft.com/en-us/library/bb522712.aspx

[15] K. J. Cios, Data Mining: A Knowledge Discovery Approach. Springer Science+Business Media, LLC, 2007. - 600 p.

[16] E. F.Codd, "Providing OLAP to User-Analysts: An IT Mandate", Codd & Associates, 1993. 27 p.

[17] M. M. Awan, M. Usman "Intelligent analysis of data cube via statistical methods", in The 10th International Conference on Digital Information Management (ICDIM), 2015. pp. 20-27.

A Statistical Approach to Mining the DM Strategy for OWA Operators

Resmiye Nasiboglu
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
resmiye.nasiboglu@deu.edu.tr

Baris Tekin Tezel
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
baris.tezel@deu.edu.tr

***Abstract*— The most important thing in using the ordered weighted averaging (OWA) operator which can easily describe the mental model of an Decision Maker (DM), is to characterize OWA weights. Determination of OWA weights cannot provide a characterization by itself. If we want to generalization and reusability of the OWA weights to aggregate various sized objects, we have to be determine more general form. In this paper, we propose a new approach for learning a stress function, which can be characterized as a DM strategy of OWA operator. For this aim the Kolmogorov–Smirnov test for similarity probability density functions is used.**

***Index Terms*—OWA operator, Stress function, Probability density function, Goodness of fit test.**

I. Introduction

The ordered weighted averaging operator (OWA) which was initiated by Yager [1], provides a parametrized class of mean operators. OWA as an aggregation operator has been widely applied in many applications and researched [2]-[8]. An n- dimensional OWA operator is a mapping f: $R^n \rightarrow R$ in order to weighting vector $W=(w_1,\ldots,w_n)^T$ defined as:

$$f(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{j} w_j b_j$$

,

where b_j is the jth largest element in the set of the aggregated objects $\{a_1, \ldots, a_n\}$. It is assumed that $w_j \geq 0$ for all j and $\sum_{j=1}^{n} w_j = 1$.

In the fundamental aspect of OWA aggregation is that the weights are assigned to ordered objects. Thus, object a_i is not connected with a particular weight w_j. On the other hand, a weight w_j is connected with a particular rank position within ordered objects.

The OWA operator allows to model different aggregation functions which are many up from the maximum through the mean to the minimum. Thus, it enables to model different mental state of a decision-maker such as being optimistic or pessimistic.

Yager introduced two main characterizing function for the weighted vector W of an OWA. First of them is the orness function defined as $\alpha = \frac{1}{n-1}\sum_{i=1}^{n} w_i(n-i)$ where $\alpha \in [0,1]$. It characterizes the degree to which the aggregation is like a logical *or* operation. Its minimum value 0 corresponds to the vector W where $w_n = 1$ and $w_j = 0$ for all $j \neq n$. This scenario implies that the OWA becomes a minimum operator. Its maximum value 1 corresponds to the vector W where $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$. This scenario implies that the OWA becomes a maximum operator. Another special scenario is the one where $w_j = \frac{1}{n}$ for all j. In this scenario, α equals to 0.5 and the OWA represent arithmetic average operator. The second is to differentiate weight distribution at a given α, called measure of dispersion or entropy is defined as $disp(W) = -\sum_{i=1}^{n} w_i \ln w_i$, where $0 \leq disp(W) \leq \ln(n)$. It provides a degree of information according to the objects are used.

One of the major challenges in using the OWA operators is to generate its weights. A number of approaches have been suggested for obtaining OWA weights since the introduction of OWA operators, such as, quantifier guided aggregation [1,10], exponential smoothing [11], learning form observations [11,12,13], minimal variability weighting vector [14], maximum entropy [15,16] etc.

In this paper, we first reviewed the approach of obtaining OWA weight from observation proposed by Yager. Then, we proposed an approach for characterizing OWA weights based on probability density function. Finally, a computational experiment was given to illustrate proposed method.

II. Learning OWA operators from observations

Yager developed an algorithm for determination of the OWA weights [11]. It can be used to learn the weights from an observations of the performance. Given are a collection m observation each comprised of n objects $(a_{k1}, a_{k2}, \ldots, a_{kn})$ and associated single aggregated value denoted as $d_k, k = 1, \ldots, m$. Assume that a set of m alternatives and n criteria which are used to evaluate each of the alternatives, are presented. Thus, an expert provides an aggregation score for these alternatives one by one. Obtained aggregation scores are denoted d_k.

The goal of this approach is to obtain a weighting vector W which is known as the OWA operator. First of all, objects of the all observations are reordered where b_{kj} is the jth largest element of the objects collection of kth observation. Whilst finding the vector of OWA weights $W = [w_1, \ldots, w_n]^T$, the

approach uses these ordered objects. In order to find a vector OWA weights W, the aggregation operator will be approximated by the instantaneous error e_k where

$$e_k = \frac{1}{2}(\hat{d}_k - d_k)^2,\ k = 1, \dots, m.$$

We denote $\hat{d}_k$ as the estimate of the aggregated value d_k. On the other hand, to avoid the constraints on w_i , the OWA weights are defined as follows:

$$w_i = \frac{e^{\lambda_i}}{\sum_{j=1}^{n} e^{\lambda_j}},\ i = 1, \dots, n$$

Thanks to aforementioned transformation, the weights w_i will be positive and will sum to 1 for any values of the parameters λ_i. In accordance with expressed above, the algorithm based on gradient descent technique was developed by Yager defines as in the Algorithm 1.

Algorithm 1.

Step 1. Initialize λ_i and calculate w_i for all $i = 1, \dots, n$.

Step 2. Calculate $\hat{d}_k = b_{k1}w_1 + \cdots + b_{kn}w_n$ for each observation $k = 1, \dots, m$.

Step 3. If $\sum_{k=1}^{m} e_k = \sum_{k=1}^{m} \frac{1}{2}(\hat{d}_k - d_k)^2 > \epsilon$ where $\epsilon > 0$ is a real number close enough to zero, then stop.

Step 4. Update $\lambda_i = \lambda_i - \beta w_i (b_{ki} - \hat{d}_k)(\hat{d}_k - d_k)$ for each observation where β is a learning rate.

Step 5. Determine $w_i = \frac{e^{\lambda_i}}{\sum_{j=1}^{n} e^{\lambda_j}}$ for each $i = 1, \dots, n$.

Step 6. Calculate $\hat{d}_k = b_{k1}w_1 + \cdots + b_{kn}w_n$ for each $k = 1, \dots, m$. Then go to Step 3.

III. Characterizing OWA weights based on probability density function

The weights distribution of OWA operator can be viewed as probability density function (PDF). Because weights of an OWA have to satisfy basic assumption of probability, such as, $w_j \in [0,1]$ and $\sum_{j=1}^{n} w_j = 1$. In this context, PDF can be used to characterize OWA weights. The PDFs are continues functions that describes the relative likelihood for this random variable to take on a given value. Here, the area under the curve is "1", too. The OWA weights are discrete and similar to probability mass function which refers to PDF, in which the sum of the weights is equals to "1". Therefore, PDFs can be discretized to accommodate OWA weights. The advantages of using PDF for characterizing OWA weights which are obtained from observation include: (1) simplicity, (2) generalization and (3) reusability. In this context, if we view the weights of OWA operator as PDF, then a basic unit-interval function (BUM) can be seen as the empirical distribution function (EDF) which estimates the cumulative distribution function (CDF) of related weights [17].

The idea of the EDF tests is estimated based on the data with (CDF) to see if there is a good agreement between them. Kolmogorv-Smirnov (KS) goodness of fit test which can be applied to obtain a probability of similarity between two distributions, is the most important and widely known of them. KS test is applicable to unbinned distributions that are functions of a single independent variable, that is, to data sets where each data point can be associated with a single number. The idea behind this test is first to obtain CDF or EDF which estimates CDF of the two distributions that need to be compared. KS test statistic is a particularly simple measure: It is defined as the maximum value of the absolute difference between two cumulative distribution functions. Given n ordered data points, the test statistic which is proposed by Kolmogorov [18], defined as follows:

$$T_n = sup_x(F(x) - F_n(x)),$$

where $F(x)$ is the CDF of a given distribution to be tested to fit, whereas $F_n(x)$ is EDF calculated from actual observed data. The table of critical points $T_\alpha(n)$ of the distribution of T_n is presented for various sample sizes n and significance levels. The limiting form for the distribution function of Kolmogorov's T_n is represented as follows;

$$\lim_{n\to\infty} P\left(\sqrt{n}T_n \le x\right) = 1 - 2\sum_{i=1}^{\infty}(-1)^{1-i}e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x}e^{-(2i-1)^2\pi^2/(8x^2)}$$

.

The first representation given by Kolmogorov, the second coming from a standard relation for theta functions and better suited for small x. Whilst, a table of the distribution was published by Smirnov [19]. Thus, the KS test constructed by using the critical values of the Kolmogorov distribution. The null hypothesis is rejected at level α if following inequality is provided:

$$\sqrt{n}T_n > T_\alpha(n)\ ,$$

where $T_\alpha(n)$ is obtained by Kolmogorov distribution.

The KS test statistic is meant for testing:

H_0: The data follow a specified distribution;
H_1: The data do not follow a specified distribution.

If the obtained T_n exceeds the 1-α quantile as given by the table of quantiles for the Kolmogorov test statistic, then H_0 can be rejected at the level of significance α [20].

Thus, using KS statistic and relation between the weights distribution of OWA operator and PDF, we can characterize OWA weights. It is clear that $F_n(0) = 0$, $F_n(1) = 1$ and it is monotonic, obviously continuous and integrable where $F_n(x)$ can be considered as representation of $F(x)$ which is a BUM function. In this context, PDF shall be referred as a stress function. Moreover, PDF indicates where the stress is place in the determination of the OWA weights obtained when $F_n(x)$ is used as a weight generating function [21]. Thus, PDF which characterizes OWA weights can be consider as stress function.

Under the guidance of all things described above, we propose an algorithm which is used for characterizing OWA

weights through predefined PDF templates, is defined as follows:

Algorithm 2.

Step 1. Use the Algorithm 1 in order to learn the weights of OWA operator from observations.
Step 2. Calculate the function $F_n(x)$ according to learned weights $w_i, i = 1, \dots, n$.
Step 3. Determine a significance level α and determine the set of tested stress functions (PDFs) $f^i(x), i = 1, \dots, t$.
Step 4. Initialize $i = 1$.
Step 5. Calculate the CDF $F^i(x)$ according to the PDF $f^i(x)$.
Step 6. Calculate $T_n = sup_x(F^i(x) - F_n(x))$.
Step 7. If $\sqrt{n}T_n \leq T_\alpha(n)$ then return $f^i(x)$ as a stress function of the OWA operator. Go to Step 10.
Step 8. $i = i + 1$ and if $i \leq t$ go to Step 5.
Step 9. There is not any fitting stress function from the set of template functions.
Step 10. Stop.

We shall illustrate characterizing the weights of OWA operator which is obtained from observations, in the following section.

IV. Computational Experiments

The collection of 5 observation and related aggregated value given in Table 1. Each observation has 5 objects.

TABLE 1. The collection of observations of data

Observation	Objects					Aggregated Value
1	0.4	0.1	0.3	0.8	0.7	0.67
2	0.1	0.7	0.4	0.2	0.9	0.71
3	1	0	0.3	0.5	0.6	0.75
4	0.2	0.2	0.1	0.4	0.8	0.56
5	0.6	0.3	0.2	0.1	0.9	0.67

The aggregated values for each observation was calculated by an expert decision. First of all, the OWA weights obtained by the Algorithm 1. The Algorithm 1 was applied on the sorted objects. It was initialized by $\lambda_i = 0$ and $w_i = 0.20$ for all $i = 1, \dots, 5$. β which is a learning rate, was used as 0.35. The estimated values of λ_i after 300 iterations were:
$\lambda_1 = 1.40; \lambda_2 = 0.20; \lambda_3 = -0.33; \lambda_4 = -0.53; \lambda_5 = -0.74$.
By the obtained λ_i , the following OWA weights were approximately determined:
$w_1 = 0.57; w_2 = 0.17; w_3 = 0.10; w_4 = 0.08; w_5 = 0.07$.
When the learning process was finalized, estimated aggregated values $\hat{d}_k$ were calculated as follows:
$\hat{d}_1 = 0.65; \hat{d}_2 = 0.70; \hat{d}_3 = 0.75; \hat{d}_4 = 0.57; \hat{d}_5 = 0.67$
Thus, the sum of squared instantaneous errors which is shown below, were close enough to zero.
$e_1 = 0.000151; e_2 = 0.000035; e_3 = 0.000005;$
$e_4 = 0.000069; e_5 = 0.000008$
and

$$\sum_{k=1}^{m} e_k = 0.000268.$$

After the end of the learning process, characterizing OWA weights phase was applied. In this phase for illustration to the approach, we used exponential distribution templates using the parameter value specified for different means. So, KS test applied the exponential distribution using the parameter value specified for $\mu = 0.01$, initially. Then, μ is increased for each step until KS test fails to reject the null hypothesis at the default 5% significance level. In this context, first-time fitted distribution of the obtained weights was the exponential distribution using μ=0.083. Determined KS test statistic was 0.5589 and critical value was 0.5633. It is clear that we did not obtain the perfect fitting but acceptable. If a better fit is desired, it should be studied observations with more data.

Plot of EDF and the exponential CDF for a visual comparison is given as follows:

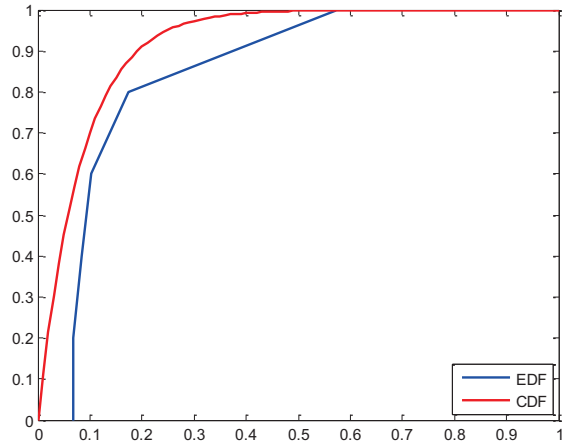


Fig. 1. Plot of EDF and the exponential CDF for a visual comparison

Thus, we obtained the PDF which indicates the stress function of obtained OWA weights, is given as follows:

$$f(x) = \frac{1}{0.083} e^{-\frac{1}{0.083}x}; x > 0.$$

Finally, the mental state of a decision maker was generalized as stress function. But to assure $\sum_{j=1}^{n} w_j$ equals to 1 and get valid set of OWA weights, we can make a normalization of the OWA weights. The normalization is defined as follows:

$$w_j = \frac{f(\frac{j}{n})}{\sum_{j=1}^{n} f(\frac{j}{n})}, j = 1, \dots, n.$$

Thus, these w_j which are obtained from the stress function, are considered as a valid set.

V. Conclusion

Our objective in this paper introduces a new approach for generalization OWA weights from the optimistic/pessimistic mental model of the decision maker. This generalization was realized by introducing a methodology based on the idea of the combination of Yager's learning algorithm and using PDF as a

stress function which is obtained from KS test. By this way, we have a number of useful features. First, since OWA weights generalization is realized by a function, same function is used for any *n*. So, we can obtain consistent OWA weights for set of object of different sizes. Another useful feature of this approach is that it allows to a user very easily characterize the OWA operator by stress function. The stress function provides us a very powerful visual tool. In this way, a user can easily see where the weights will be allocated. As a result, we proposed the new approach to characterize and to generalize OWA weights and investigated some features of this approach. Furthermore, we provided a numerical example of this approach. In future studies, we will try to improve our approach for getting better characterization of the OWA weights.

REFERENCES

[1] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", IEEE Trans. Syst., Man, Cybern., col. 18, pp. 183-190, 1988.

[2] G. Bordogna and G. Pasi, "Controlling retrieval through a user adaptive representation of documents", Int. J. Approx, Reason, vol. 12, pp. 317-339, 1995

[3] G. Bordogna and G. Pasi, "Linguistic aggregation operators in fuzzy information retrieval", Int. J. Intell, Syst,, vol. 10, pp. 233-248, 1995

[4] B. Bouchon-Meunier and M. Rifqi, R. R. Yager and J. Kacprzyk, The Ordered Averaging Operators: Theory and Applications, pp. 29-35, 1997, Kluwer Academic

[5] H. B. Mitchell and D. D. Estrakh, "A modified OWA operator and its use in lossless DPCM image compression", Int. J. Uncertain., Fuzzi. Knowl. Based Syst., vol. 5, pp. 429-436, 1997

[6] W. Pedrycz, "Logic-based neurons: Extensions, uncertainty representation and development of fuzzy controllers", Fuzzy Sets Syst., vol. 66, pp. 251-266, 1994

[7] H. J. Skala, "Concerning ordered weighted averaging aggregation operators", Statist. Papers, vol. 32, pp. 35-44, 1991

[8] R. R. Yager and J. Kacprzyk, The Ordered Weighted Averaging Operators: Theory and Applications, 1997, Kluwer

[9] D. Filev and R. R. Yager, "On the issue of obtaining OWA operator weights", Fuzzy Sets and Systems, vol. 94, pp. 157-169, 1998.

[10] R.R. Yager, Families of OWA operators, Fuzzy Sets and Systems 59 (2) (1993) 125–143.

[11] D. Filev, R.R. Yager, On the issue of obtaining OWA operator weights, Fuzzy Sets and Systems vol. 94 no. 2 pp. 157–169, 1998.

[12] V. Torra, OWA operators in data modeling and reidentification, IEEE Transactions on Fuzzy Systems vol. 12 no. 5 pp. 652–660, 2004.

[13] R.R. Yager, D.P. Filev, Induced ordered weighted averaging operators, IEEE Transactions on Systems, Man and Cybernetics, Part B 29 (2) pp. 141–150, 1999.

[14] R. Fuller and P. Majlender. On obtaining minimal variability OWA operator weights[J], Fuzzy Sets and Systems, vol. 136, no. 2, pp. 203-215, 2003.

[15] D. Filev, R.R. Yager, Analytic properties of maximum entropy OWA operators, Information Sciences vol. 85 no. 1–3, pp. 11–27 1995.

[16] R. Fulle´r, P. Majlender, An analytic approach for obtaining maximal entropy OWA operator weights, Fuzzy Sets and Systems vol. 124 no. 1 pp. 53–57, 2001.

[17] R. Sadiq and S. Tesfamariam, Probability density functions based weights for ordered weighted averaging (OWA) operators: An example of water quality indices, European Journal of Operational Research vol. 182, pp.1350-1368, 2007.

[18] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione". G. Ist. Ital. Attuari 4: 83–91, 1933.

[19] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions". Annals of Mathematical Statistics 19: 279–281, 1948.

[20] N. M. Razali, T. B. Wah, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, Journal of Statistical Modeling and Analytics, vol.2, no.1, pp. 21-33, 2011

[21] R.R. Yager, Using Stress Functions to Obtain OWA Operators, IEEE Transactions on Fuzzy Systems, vol. 15, no. 6, pp. 1122-1129, 2007.

The most common problems of Data Mining

Sabina Fomenko
ANAS Institute of Information Technology
Baku, Azerbaijan
sabina.fomenko@gmail.com

Sayar Abdullayev
ANAS Institute of Information Technology
Baku, Azerbaijan
depart5@iit.ab.az

***Abstract*—The article deals with the concept of Data Mining. Data Mining Describes the appearance, the basic essence of the problems and in some problems the process and methods of solution are discussed, including the application. Big Data and Data Mining are compared.**

***Keywords*—Big Data; Data Mining; classification; clustering; forecasting; visualization.**

I. Introduction

In order to get the constructive and necessary information of prior methods that applied in statics and mathematics required a lot of time. It led to the formation of Data Mining.

Data Mining as a term originated in 1978. Until the mid-1990s, the data were processed and analyzed in the framework of Applied Statistics. At the same time more often the problem of processing small databases was solved. During that period of time the concept of Data Mining began to gain popularity with a modern twist. One of the founders of the Data Mining direction is Gregory Piatetsky-Shapiro [1].

The concept of Data Mining represents extraction (mining) data (data). Synonyms of the Data Mining can be considered of the concept of "knowledge discovery in databases" (Knowledge Discovery in Databases, KDD) and "Data Mining".

Data Mining is used wherever there is any data. But nowadays Data Mining techniques primarily intrigued businesses which deploy projects, the basic information on the data stores.

The main characteristic feature of Data Mining, is the combination of a large mathematical tools (from the traditional statistical research till the most recent cybernetic methods) with the latest developments in the sphere of information technology.

Data Mining is a multi disciplinary sphere, which arose and formed on the basis of such sciences as pattern recognition, database theory, applied statistics, artificial intelligence, etc. (Figure 1)

Every year Data Mining is dedicated to a number of scientific and practical conferences. One of them is the International Conference on Knowledge Discovery Data Mining (International Conferences on Knowledge Discovery and Data Mining).

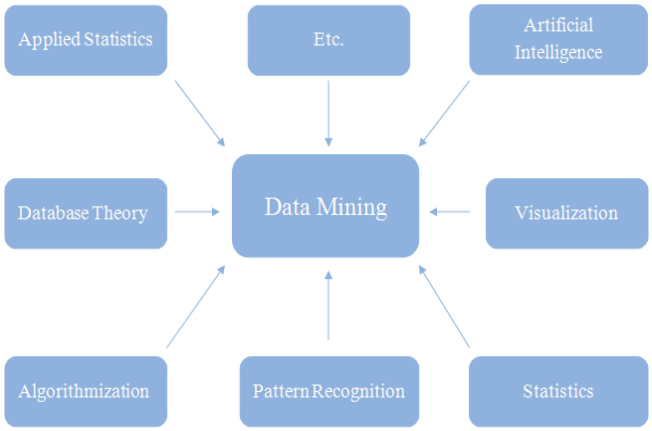


Figure 1. Data Mining as a multidisciplinary field

II. Objectives Data Mining

The template concept consists of Data Mining technology base. The concept reflects in the data the unexpected, unobvious regularity. Such regularity consists of the tacit knowledge. The concept is a pattern. They are characterized by sampling the data expressed in a form intelligible to the man. Data Mining problems are solved with the detection result of the imperceptible patterns. Sometimes Data Mining is called task laws or techniques [2].

There is no consensus on what tasks you can rank as Data Mining. In many publications the following Data Mining tasks are listed:

1. Classification,
2. Clustering,
3. Forecasting,
4. Association,
5. Visualization,
6. Analysis and detection of deviations,
7. Assessment,
8. Analysis of the relationship,
9. Debriefing.

The classification task. The most often solved, as well as the simplest task of Data Mining is the classification.

Classification - is ordered in accordance with the certain principles of the mass of objects having similar classification criteria by which determined the similarities or differences between these objects.

The following rules should be followed in the problem of classification:

- only one base should be applied in each of the division;
- the total amount of categorical concepts of equal volume of generic concepts that share. Dividing is necessarily proportionate;
- the volumes of dividing members can not cross. It is necessary to be excluded by each other mutually;
- the division has to be consistent.

It is possible to distinguish the artificial and the natural classification. Artificial is called classification carried out by their appearance. Auxiliary classification is needed to make the most of the processes and phenomena of the given order. Natural classification is called, which is carried out on the essential features that characterize the internal community of objects and phenomena. Natural classification is a result and an important means of the scientific research.

The problem of classification can be solved by using the following methods:

- nearest neighbor;
- induction of decision trees;
- Bayesian networks;
- neural networks.

The clustering problem. In the clustering the classes of the investigated the data set are uncertain previously.

Clustering is used with the aim of partition a set of objects in similar clusters or classes. Representing the sample data in the form of points in the feature space, the clustering problem can be reduced to the concept of "points of accumulation".

Studying the "data structure" with the help of clustering the investigated analysis can be performed.

Clustering is used in many industries. It was originally used in biology, anthropology and psychology.

Clusters are non-intersecting and overlapping.

Sometimes the result of the use of various methods of clustering the different results are obtained. It depends on the characteristics of an operation algorithm. There are more than hundreds of different clustering algorithms.

An example of a way to solve the clustering problem is education "without a teacher".

The prediction task. In the various sectors of human activity, one can meet the task of forecasting. Development of information technologies and the development of prediction methods are directly related to each other. Since it requires a detailed study of the initial set of appropriate data and methods of analysis, it can be considered one of the most difficult tasks Data Mining. The purpose of forecasting is to predict future events.

Prediction aimed at identifying trends in the dynamics of a particular object or event based on an analysis of its condition in the past and the present. Therefore, the decision of forecasting problems requires some training sample data.

Doing the task of forecasting functional relationship is set between the independent and the dependent variables. In many areas of human activity there are common forecasting and demand problem. As a result, the prediction may be noted decrease in the risk of making incorrect, unreasonable or arbitrary decisions.

Association task. In solving the problem, the association patterns are found among the related events in the data set.

Associations are different from the classification and clustering in that the search takes place between the search patterns of several events occurring simultaneously.

The problem of finding association rules is solved with the help of many algorithms.

The visualization task. Visualization allows seeing the final result of the calculation and organizing computing process management. The result of the use of visualization is a graphical image data. Visualization is presented in the form of histograms, graphs, charts, diagrams, etc. Visualization is used in problems of classification and clustering.

The visualization is practically no need for special training of the user.

The role of imaging is the followings: support of coherent and interactive investigation; assist in the presentation of results; the usage to generate visual images and their understanding.

Analysis of relations. This problem with the help of which are based in the dataset.

Summarizing. The purpose of this task is a description of the specific groups of objects selected from the analyzed data set [3].

III. Data Mining and Big Data

Big Data and Data Mining are two different concepts. Each of the concepts is used in the processing of large data volumes. In addition, these two terms are used for two different directions of this kind of work.

Big Data - This term refers to a set of big data. Big Data is the data that has already degenerated simple databases used in the earliest time, which were more expensive and had less capacity. Examples of it are data sets that are too large to be processed in different systems.

Lately the word Big Data, one can say, has become the main stream. With Big Data are closely related terms such as data analysis, data science, data analysis, data collection, and machine learning.

Unlike Big Data, Data Mining is used to find specific information among the large data set. The synonym for this type of activity can be considered as a statement of "looking for a needle in a haystack". It means that large data sets are automatically collected that can be connatural and similar. Data Mining can include the use of different programs [5].

Data Mining - is the process of purification of large data preparation and subsequent analysis or use in machine learning algorithms. To merge and transform vast amounts of data, Data Mining is necessary to possess exceptional qualities Sensor detection, wonderful intuition and technical skills [6].

Conclusion

Thus, Data Mining technology with each passing day it becomes more and more popular. With the help of Data Mining one can solve many problems and process large amounts of data.

Nowadays there are several points of view on Data Mining. Some people believe that it is a mirage that distracts the attention from the classical data analysis. Others believe that Data Mining is an alternative to the traditional approach to the analysis. As well as in all cases, there is a middle, in which is being considered the possibility of joint use of the latest achievements in the field of Data Mining and classical statistical analysis of data.

In order to use the Data Mining the special skills are required. Not having it is very difficult to work with Data Mining. Qualifications must constantly increase.

REFERENCES

[1] I.A.Chubukova, “Data Minig” The course of lectures online University INTUIT, 2008, pp. 7-16.

[2] V.Dyuk, A.Samoylenko “Data Minig” publishing house «Piter», 2001, pp. 14-20.

[3] http://bug.kpi.ua/stud/work/RGR/DATAMINING/tasksofdm.html

[4] Kdnuggets, “Data Mining, Analytics, Big Data, and Data Science”, [Online]. Available: www.kdnuggets.com

[5] Techopedia, "What is the difference between big data and data mining?", [Online]. Available: https://www.techopedia.com/7/29678/technology-trends/what-is-the-difference-between-big-data-and-data-mining

Syntactic Text Analysis Without a Dictionary

Bessmertny I. A., Professor
ITMO University
Saint Petersburg, Russian Federation
bia@cs.ifmo.ru

Platonov A.V., Postgraduate Student
ITMO University
Saint Petersburg, Russian Federation
avplatonov@corp.ifmo.ru

Poleschuk E.A., Postgraduate Student
ITMO University
Saint Petersburg, Russian Federation
eapoleschuk@corp.ifmo.ru

Ma Pengyu, Postgraduate Student
ITMO University
Saint Petersburg, Russian Federation

***Abstract* — Syntactic text analysis is a very important step of automatic text processing. The key problem is that all existing approaches are dictionary-dependent. It can be impossible to analyze all sentences because of the lack of one word in the dictionary. However even the presence of the dictionary does not resolve the phrases interpretation ambiguity. At the same time fusional languages contain enough information about the parts of speech. Particularly, the endings of nouns, adjectives, verbs and participles cannot fully indicate their case, number and face because of semantic ambiguity. Prepositions also cannot always correctly indicate the cases. This yields a variety phrase interpretation. To solving this problem sharing affixes and prepositions is proposed. However fusional languages have some disadvantages. One of this is the allowance of building phrases with random word order. It leads to increase the number of options. The paper proposes to use sentences templates with analyzed phrases for ambiguity decrease. In addition usage of affixes with prepositions for text analysis provides the basic elements of the semantic analysis without a dictionary.**

***Index Terms* — syntactic analysis, dictionary, fusional languages, affixes, phrase pattern**

I. Introduction

An important stage of automatic text processing is syntactic text analysis. Standard syntactic analysis approach is based on a grammar and vocabulary in which each word of a text is assigned a lexical unit and part of speech. Main element of a syntactic analyzer is a dictionary, independently of sentences model based on the constituents [1] or a dependency grammar [2, 3]. Initially syntactic analyzers were created for formal languages. In particular, they were created for programming languages with restricted dictionary. A word's presence in a text and its absence in a dictionary are to be considered as a mistake. Transfer of technologies primarily used by compilers of programming languages to natural languages has created a many problems. There are problems such as significantly large dictionary size, homonymy and polysemy appear in this transfer. Because of this, natural languages ambiguity began to appear in parsing and words absence in the dictionary makes sentences parsing problem completely impossible to execute. So the syntactic analysis problem based only on a grammar, with minimal dictionary using, is considered to be interesting.

II. Problem Statement

Natural language texts processing can have different aims, in particular, classification, clustering, annotation, machine translation and others. Each of these goals is reached when the text semantics is extracted. Sentence is known as a minimal semantic unit of the text. It consists of words where each of them has one or more meanings. Hence, the text meaning can be understood as knowing the meanings of words in a sentence. The meaning of words is dependent not only on their denotations, but also on their role in a sentence. Each combination of words meaning produce a phrase interpretation. Thus, the phrase semantic analysis has a goal to eliminate multiple interpretations to single.

Syntactic text analysis is the first step which allows reduces the number of interpretations of the sentences. At this step sentences structure hypotheses applied. Existing solutions are based on dictionaries. Interpretations are made from sets of morphemes, which mainly used from dictionaries. If the word is not presented in the dictionary, interpretation will not be formed, and as a result text analysis becomes incomplete. Meanwhile, fusional languages, which include, in particular, the Russian language, contain information to identify the parts of speech in morphemes (prefixes, suffixes and endings). People use this approach when reading a text in a foreign language. The paper [3] demonstrated that people can understand a text with 30% of unknown words because the knowledge of grammar allows us to perform syntactic text analysis successfully.

Thus, we assume that syntactic analysis is possible for fusional languages just on the basis of grammar, namely, syntax and morphemic.

III. Syntactic Text Analysis Problem Solution Based on Grammar

Syntactic analysis problem is to prove such hypothesis: “Does the phrase belongs to a particular syntactic sentences model?”. This model usually represented as a tree. In this

paper we don't solve the problem of sentence model synthesis. Therefore we use fixed set of linear models, the simplest of which is as follows:

```
phrase([(subject, nominative),(predicate),
        (addition, accusative)]),
```

where the `subject` is in the nominative case, `addition` is in the accusative case. Here and below used the Prolog language notation. It should be noted the subject can be represented by a noun or an adjective and a noun, `predicate` can be a verb or an adverb and a verb, and an addition can be a noun or an adjective and a noun, as shown below.

```
subject([noun]).
subject([adjrctive,noun].
predicate([verb]).
predicate([adverb,verb]).
addition([noun]).
addition([adjective,noun]).
```

Noun morphemes have the next definition: ending, ending of the normal form, number, sign of animate / inanimate, gender and case; adjectives have ending, end of the normal form, number, gender, case; verbs are defined by ending, the infinitive ending, number, person, gender, tense. The following example shows the definition of endings of animated feminine singular nouns, corresponding to the words «жена» ("wife"), «сотрудница» ("employee"), etc.

```
    ending(noun,'a',
'a',(singular,animate,female,nominative)).
   ending(noun,'ы',
'a',(singular,animate,female,genitive)).
   ending(noun,'е',
'a',(singular,animate,female,dative)).
   ending(noun,'у',
'a',(singular,animate,female,accusative)).
   ending(noun,'ой',
'a',(singular,animate,female,instrumental)).
   ending(noun,'е',
'a',(singular,animate,female,prepositional)).
```

So, syntactic analysis is a search of morphemes matching for a selected syntactic model. All morphemes should correspond to all the words in the sentence. Also cases should be the same as in the model, number and gender should be the same as in the subject and in the predicate. It should be noted, if the subject is pronoun, its grammatical person should be the same as predicate's person.

Let's consider the Russian language phrase was popularized by Lev Uspensky [7].

"Глокая куздра штеко будланула бокра и курдячит бокрёнка"
(*Glokaya kuzdra shteko budlanula bokra i kurdyachit bokryonka*).

In the phrase, all word stems (glok-, kuzdr-, shtek-, budl-, bokr-, kurd-) are meaningless, but all affixes are used in a grammatically correct way. Consequently, this phrase cannot be parsed with the help of vocabulary.

One of the variant of this sentence's syntactic analysis on Prolog are shown below

```
   [(subject,
(Глокий,adverb,singular,female,nominative),

(куздра,noun,singular,animate,female,nominative
)),
   (predicate,(штеко,adverb),(будлануть,verb,si
ngular,3rd,female,past)),
   (addition,(бокр,noun,singular,animate,male,a
ccusative)),
    (link,и),
    predicate,(кудрячить,verb,singular,3rd,fem
ale,present)),
   (addition,(бокренк,noun,singular,animate,mal
e,accusative))]
```

Obviously, there is only one error (it is marked in bold) in this interpretation. It is the lemma «бокренк» that must be equal to «бокрёнок». This error is caused by incorrect parsing of word «бокрёнка» to «бокренк-а» instead «бокр-енка». Moreover, first parsing variant of this word produce one extra lemma «бокренко». Number of interpretations of sentences is increased by such ambiguities and they up to one hundred variants. Redundant interpretations may be reduced due to complementary passes of syntactical analyzer with minimization of root morphemes. In particular, in first part of example sentence root «бокр» had been found without ambiguities and it may be used in second part of sentence. This operation reduces number of interpretations to 11. After defining morpheme «бокр» it can be found that word «бокр» has an animated form because it has «бокра» in accusative case. If word «бокр» could be animated form it would has «бокр» in accusative case.

Other interpretations are caused by non-usage of grammatical person (first and second) analysis for subject in the sentence. If parsing of pronoun is added then the number of interpretations will decrease to one. It should be noted there is no task of parsing correctness definition. Its correctness is defined by semantic but this sentence is meaningless. Only uniqueness and consistency may be established.

IV. Semantic Definition of Sentence by Syntactic Model

Meanings of sentence parts are defined through syntax in analytic languages like English. It complicates a semantic parsing without context when words have an equal or similar spelling. For example, it may be shown the wordplay: «*What has four wheels and flies*?» Valid answer is - «*A garbage truck*», because «flies» has an insect meaning. In contrast, morphemes in fusional languages carry additional information and allow semantic parsing. In particular prepositions with cases of its objects can define some general meaning. Several

descriptions of prepositions with cases and semantics are presented below.

```
   prep('по', dative,       'moving along').
   prep('в',    accusative,      'movement  from
outside to inside').
   prep('в',  prepositional,'presence inside').
   prep('на', accusative,   'placing the top').
   prep('на', dative,       'precence top').
   prep('от', genitive,     'retreat').
   prep('над',instrumental, 'presence above').
   prep('из',  genitive,        'movement  from
inside to outside').
```

Let's parse phrase «*Мишка косолапый по лесу идет*» (Mishka kosolapyy po lesu idet) by such simple semantic analysis. Preposition «по» with dative case of word shows monosemantic meaning «moving along». There is result of parsing:

```
   [(subject,(косолапый,adverb,singular,male,no
minative),
(Мишка,noun,singular,animate,male,nominative))
,
    (addition, (по,preposition,'moving along'),
   (лес,noun,singular,animate,male,dative)),
    (predicate,
(идти,verb,singular,3rd,male,present))]
```

This is a one of interpretation of example sentence. Other interpretations is appeared by impossibility of definition of object («мишка») and subject («лес») animation, and a controversial extraction of the root morpheme of the word «лесу». However, semantic relation «*мишка* - moving along — *лес*» may be detected in case of multiple sentence interpretation.

Other morphemes also can be carriers of semantic. Despite the variety of prefixes is greater than prepositions their unambiguous interpretation is often impossible. For example prefix «на» unlike preposition «на» does not have independent value in most of cases (*наклон, начинка, напиться, налево*,...). Languages also have conjunctions like «and» («и») or «or» («или») that can define logic of sentence. Unfortunately, in ordinary speech these conjunctions do not mean logical conjunction or logical disjunction. Conjunction «и» («and») often mean disjunction (*Места для пассажиров с детьми и инвалидов*) and «или» («or») mean horn clauses (*Или я веду ее в ЗАГС, или она ведет меня к прокурору*).

V. Conclusion

Our research does not cover all Russian language grammar features it shows that fusional languages phrases contains enough information for syntactic analysis with using of grammar, some non-root morphemes and prepositions and without a complete dictionary. In addition, besides morphological parsing, such approach allows define animation of objects and it can help to define semantic relations in some cases. The main advantage of this method is absence of situation of unknown word while parsing in comparison with other approaches. The absence of unknown words does not restrict syntactic parsing.

Next researches will be dedicated to reducing number of interpretations of each sentence due to complementary syntactical analyzer passes which minimize number of root morphemes.

References

[1] Chomsky N. “Syntactic structures”, The Hague, Mouton, 1957, P. 117.

[2] Partee B., Ter Meulen A., Wall R. “Mathematical methods in linguistics”, Springer, Vol. 30 of Studies in Linguistics and Philosophy, 1990.

[3] Mel’cuk I. A. “Dependency syntax: theory and practice”, SUNY Press, P. 428, 1988.

[4] Kagirov I.A., Leont'eva A.B. “Automatic Russian texts parsing on the basis of grammar constuent”, Izvestiya vysshikh uchebnykh zavedeniy. Priborostroenie. Vol. 51. № 11. pp. 47-51, 2008.

[5] Smirnov I.V., Shelmanov A.O. “Semantic and syntactic analysis of natural language Part 1. Overview of the methods text syntactic and semantic analysis”, Iskusstvennyy intellekt i prinyatie resheniy, № 1. pp. 41-54. 2013

[6] Frumkina R., Dobrovich A. “Kak luzirovat' Simenona, ne zacherivaya v torvar”, Znanie - sila № 1, pp. 36-37., 1971

[7] Uspenskiy L. “Word about words”, Leningrad, Detskaya literatura. 1971

Events Analysis Based on Internet Information Retrieval and Process Mining Tools

Irina Shalyaeva
Business Informatics Department
National Research University Higher School of Economics
Perm, Russian Federation
ishalyaeva@bk.ru

Lyudmila Lyadova
Business Informatics Department
National Research University Higher School of Economics
Perm, Russian Federation
llyadova@hse.ru

Viacheslav Lanin
Business Informatics Department
National Research University Higher School of Economics
Perm, Russian Federation
vlanin@hse.ru

***Abstract* — This paper presents preliminary result of research project, which is aimed to combine ontology information retrieval technology and process mining tools. The ontologies describing both data domains and data sources are used to search news in the Internet and to extract facts. Process Mining tools allows finding regularities, relations between single events or event types to construct formal models of processes which can be used for the next ensuing analysis by experts. An applicability of the approach is studied with example of the environmental technogenic disasters caused with oil spills, and followed events. Ontologies allow adjustment to new domains.**

***Index Terms* — fact extraction, event analysis, text mining, process mining, structure-centered information retrieval.**

I. Introduction

Nowadays information retrieval tools allow finding the publications containing information on events (facts), on the objects and relations, on the time when events occurred and places and so on [9, 10, 11]. The approach allowing to reveal dependences between events (facts), information on which is published in the Internet, is offered in this paper.

The ontologies describing both data domains and data sources are used to search news in the Internet and to extract facts. The domain ontologies allow to establish connection between objects and events, to classify facts, to execute a clustering, etc. on the base of information published in the Internet. The ontologies of data sources describe the information sources in the Internet. The results of information retrieval are structured and stored in data base.

Event logs are formed on base of the stored data in formats used in Process Mining tools. The data is cleared and detailed at preprocessing with ontologies. Process Mining tools allows to find regularities, relations between single events or event types, to construct formal models of processes which can be used for the next ensuing analysis by experts.

At the research prototype development existing tools of information retrieval are used for events finding and logs generating and ProM system is used for process mining.

Process Mining gradually penetrates into the growing number of applications solution. The opportunity, which provides this discipline, to discover, monitor and improve processes, aside from such evident purpose as obtaining data from software systems, are used for different tasks like tracing and analyzing students' learning habits based on MOOC data, ontology-driven data extraction from databases, in Workflow Management Systems for the healthcare industry. This research demonstrates that combining Process Mining with other disciplines and approaches can provide a variety of interesting and nontrivial results; this promotes a lot of interest on the part of the scientific environment.

II. Definitions and Limitations of the Domain

As already mentioned, in our work we suggest to combine web, text and process mining in order to obtain unevident data patterns in a convenient and accessible graphical representation with the possibility of further model analysis. To analyze opportunities and demonstrate the described approach, technogenic accident subject area has been chosen, namely the events related to the oil spill. To understand what kinds of data can be collect from news feeds as part of this theme, a lot of query results were analyzed. Russian web-media and global search engines, like Google, were considered as data providers.

Further, assuming that the user may have two basic types of information needs:

1) gross appearance in the industry, statistics;
2) data on the specific event.

Two corresponding types of requests have been analyzed (f.e. the generalized – "oil spill", and a specific – "oil spill in Sakhalin April 5, 2016"). Obviously, the more general request is executed, the more diverse data we receive. So, by request for oil spill, results can be related to: the elimination of consequences, the sanctions measures, the new methodology to eliminate spills, actually oil disaster and many other types of events.

In this paper, we consider simple referential events:

- N. Samoilenko characterizes referential event as: "by event we understand the result of an action, behavior, occurrence, fact, which has a personal or social significance, something new, a change in the situation, the state of affairs".
- Simple events – internal form consists of the primary elements of event, first of all – action and associated

components of activity.

The minimal set of characteristics that we takes into account for events analysis are: participants of the event, geographical location, event border, internal relations between components of a single event, the relation between events. Such static event attributes as the company, the date format and geographical position will be a part of the domain ontology. Events borders were identified in the analysis of news feeds texts on the subject of oil spills. Since the news reports generally displayed the most important aspects of the case described, the event is considered as described in the ontology instance of the class "event" or, if there was no match in the ontology, any verbal constructions: verb + related words, met in the summary of the article. Communication components inside the event set during the semantic analysis that defines the verb structure to identify events. Dependency tree is constructed for each verb. Further, there is the extraction of causal relations. This task is also facilitated by the analysis of news reports, since the size of the text is usually limited from two to five sentences, and the sequence of events often correspond to the stacking order. For escalating a chain of events, sequences extracted from various news reports will be linked by a given feature. News title does not belong to processable text, because it can significantly disturb the process of causal links identifying and duplicate information from the news. However, the information from the header can be used to retrieve objects and event attributes if the event is duplicated in the news text. Also, an event in the header can be used as a marker for a situation by which a news-related situation will bind during further processing. In other words, by these markers, we can associate the events to traces. Trace in the Process Mining is a sequence of events, united by a common use case or a news message in our work. The event is an instance of activity that represents a well-defined process step. Traces are a display of the processes interaction.

III. Filling the Database to Generate Models

Obviously, on the specific request we get most of the same set of events. If this request is generated for a statistic model and converts to a single trace for the overall process, a problem arises. The problem of information duplication in the traces can be solved by means of Process Mining. However, synonymous in the data must be unified with the use of ontology.

Data retrieved from Internet via ontologies stored in the database in the same way as user queries on which we were looking for news.

If the user has generated a request, which items are not in our ontology, we offer it to expand the ontology and may suggest where he could add the concept. As a result of the work on this step system shows the user settings from the ontology that it can optionally specify (for example, to limit the time period, select a location, district, company). Then the system generates the necessary additional requests for news portals, extracts texts, extracts the information and fills the database.

After all, the user is again asked to select what data parameters for the model he is interested in. For that the original user request, extracted data and the data stored in database are analyzed. Summarizes some common virtual representation of the data relating to all the parameters of the request, on the basis of which the user is prompted to choose the criteria that he wants to be taken into account for the construction of the process model. Based on this creates RDF file for storage and log generation.

IV. Using Page Structure Ontology for Information Extraction

In our approach the information search method based on web documents structure analysis and ontologies implementation is offered. Two-level ontology capturing following description is supposed to be developed:

- website (the web document being analyzed source) structure description – main page types and their interconnections;
- Web page information blocks description and their interconnections.

The example fragment of these two levels is depicted in the Fig. 1. The first level ontology per se keeps the description of pages existing on the web site in question – sitemap, but in more simplified and generic form. While developing a specific website description, ontology nodes will be populated with the addresses of visited and analyzed pages. The second level ontology is aimed at keeping the description of information blocks to be found on a web page, for instance, navigation block, which can contain valuable information, as well as that of forming these blocks like form controls, static or dynamic images, tables, text areas and so on and so forth. In order to develop this second level ontology the most widespread HTML template kinds were examined along with template provision elements. While developing a specific webpage description, ontology nodes are supposed to capture markup places unambiguously identifying exact placements of webpage information blocks for further data extraction from it or from lower-level element constituting the block in question.

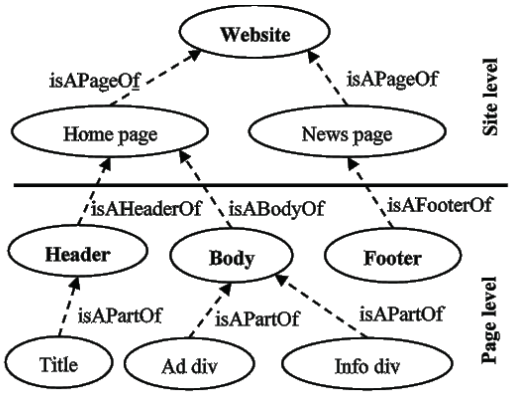


Fig. 1. Example Fragment of Two-level Web Document Ontology

In order to extract information from web documents by means of the proposed two-level structure ontology it is necessary to traverse it and identify the exact information block position for further content query.

Generally, there are several stages in the proposed ontology processing mechanism:

- Loading the ontology from local file or by the URI generated by the ontology editor exploited.
- Traversal algorithm execution and information block identification.
- Saving the placement address of the information division founded for further content query processing.

The main practical application to the offered method is to find more advantageous solution to the information extraction problem by boosting result relevance level paying more attention to the structure and placement of information.

The main advantages to this structure-centered information retrieval approach are as follows:

- web document can be annotated with the structure metadata allowing to take information placement into account;
- structure-centered information retrieval considers and exploits information divisions hierarchical structure interconnections;
- information placement metadata can help identify content duplication and filter it afterwards;
- return result representation is to be enhanced by using structure and placement metadata.

V. The Data Structure / Data Classification / Datatyping for Storing

In processing the results of the above general type of queries, in addition to the described steps for each news item or a trace, it is also necessary to classify the situation described (as shown in Figure 1, the results for "oil spill" are different types of events: social, political and environmental). For each event class should be allocated own attributes that can act as markers depending on the model that the user wants to receive.

Events are classified in terms of the attributes that they may possess. One news post may include events related to different classes. Key attributes are applying to connect them.

The following main types of events and key attributes were identified during the analysis of the news feed of oil disasters:

- Disaster (date, oil company, place) – directly disaster themselves, such as fire, spill, explosion.
- Financial implication (organization) – assessing the financial damage, this includes as a cost for the elimination of consequences as well other economic indicators of enterprises, population and countries.
- Industry news (oil company, publication date) – possible scientific discoveries, achievements, innovations in the field of oil industry, any information related to the operation of companies: enlargement, closure, bankruptcy.
- Sanction (date) – information on the sanctions and penalties.
- Socio-environmental implication (publication date) – the impact on the population, the victims, damage to agriculture, the impact on society and the possible reactions, demonstrations, unrest.
- Socio-political (Date, Place) – influence on government policy, changes in relationships, the impact on foreign trade, sea routes.
- Noise – data from the news that does not belong to a domain but are the results of a query in the Internet search engines (f.e. Define Oil spill at Dictionary.com).

VI. Multiple Advanced Search Queries

As already mentioned, once the data extracted by user request, they are analyzed and, depending on the completeness of the data set the user is able to:

1. Expand the data set (Query for a specific event and query by attributes (geographical position, the company, the time interval)).
2. Build a process models that are available at this stage.

If the user decides to supplement the data with repeated request, the system provides the opportunity to extend the ontology, and using the concept of the updated domain ontology to generate a new request to the network news portals. Advanced data set allows to create more accurate and full covered process models.

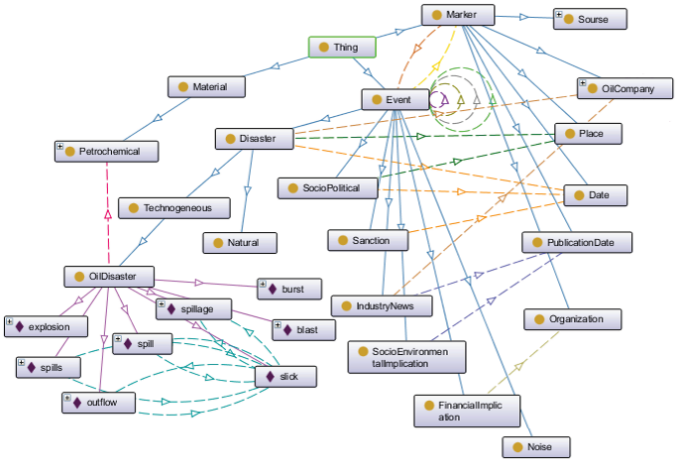


Fig. 2. A fragment of the domain ontology example

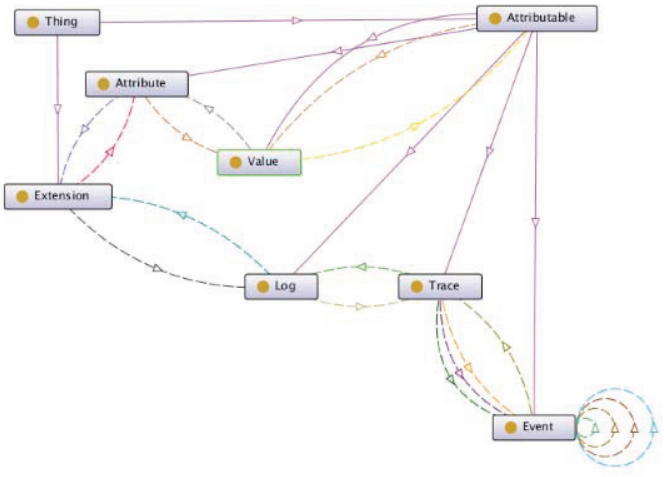


Fig. 3. XES log ontology

Further, the concepts of the domain ontology (Fig. 2) are mapped to the ontology of the log (Fig. 3), and the output document in XES format creates. As a standard set of log extension used to describe data is not enough for our domain, it is necessary to expand. In addition to the below listed standard extensions, we need to include the date of publication, name of the organization that are not directly related to petroleum activities, the location of the event and the source of the news posts.

VII. Using the capabilities of standard XES extensions

Concept extension – store the name of all levels of the log hierarchy elements: case, trace, event. The instance attribute store the identifier of the case so that identical events may differ from each other.

Lifecycle extension – reflects the stage of process lifecycle, described in the transactional model. This model can be arbitrarily set or use one of the standard models – BPAF or Standard lifecycle transition model.

In our case, it does not matter what kind of lifecycle model to choose because we will not have a hard definition of the beginning and end of the process. Each event by itself is a finite process, and depending on the chain of events, which is found, the start and the end of a whole process can be different in each situation. Therefore, we will not have a standard separation like Running or Not Running process (Fig. 4).

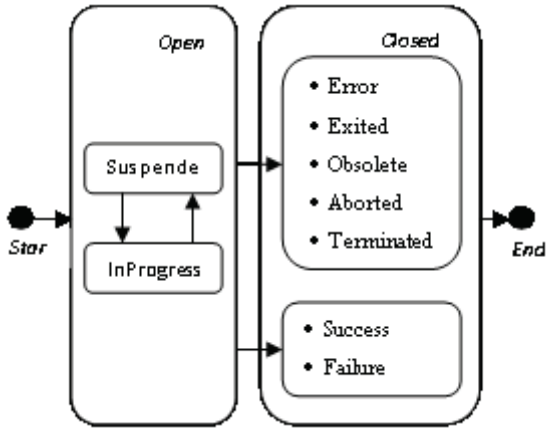


Fig. 4. Lifecycle transactional model

However, it is possible to define some time limit for the determination of completion of various classes of events. For example, if we consider an event that took place 5 years ago, and all related events dated to the same time, the process is considered complete. If we look at recent events, such as the oil spill occurred yesterday, it is obvious that the response to this event may still appear in the news, thus it is necessary to keep tracking the process. And then if the user wants to know how yesterday's oil spill reflect on the company's present, the model can not be constructed as we do not find relevant news and the process receives the status COMPLET = failure. In this case, it is also possible to be in state IN PROGRESS, if passed a certain time period f.e. a year after the start of the trial, and not found any effects events – SUSPENDED.

The beginning and the end for the different types of models as determined individually. If the model is typed, then START and END are already set. When the user builds a new unique model, a model of individual events (a specific disaster), the system will build a model by making the assumption that the event which has no predecessors is a START, and the one which has no followers – END. The user is also given the opportunity to set these parameters of the model if the system mistaken. Organizational extension – identifies three attributes for the events that identify the actor responsible for the event, and its position in the organizational structure. Time extension – date and time when the event happened. Semantic extension defines an attribute that allows to store a number of references to model concepts in a domain ontology in any element of the XES type hierarchy. ID extension – provides unique identifiers for all elements of log hierarchy. Cost extension – in our case carries information about the size of fines, eliminating price effects, damage assessment and other monetary indicators [8].

VIII. Example

According to the data obtained by a general request, the system can build a very simple process models within the same news. Obviously, this is not enough, so the user is prompted to specify a request, choose the information about Shell Spills Oil in the Gulf. Resulting log is given below:

```
<?xml version="1.0" encoding="UTF-8" ?>
<log xes.version="1.0" xes.features="nested-attributes"
openxes.version="1.0RC7" xmlns="http://www.xes-standard.org/">
    <extension name="Lifecycle" prefix="lifecycle"
    uri="http://www.xes-standard.org/lifecycle.xesext"/>
    <extension name="Organizational" prefix="org"
    uri="http://www.xes-standard.org/org.xesext"/>
    <extension name="Time" prefix="time"
    uri="http://www.xes-standard.org/time.xesext"/>
    <extension name="Concept" prefix="concept"
    uri="http://www.xes-standard.org/concept.xesext"/>
    <extension name="Semantic" prefix="semantic"
    uri="http://www.xes-standard.org/semantic.xesext"/>
    <global scope="trace">
        <string key="concept:name" value="__INVALID__"/>
    </global>
    <global scope="event">
        <string key="concept:name" value="__INVALID__"/>
        <string key="lifecycle:transition" value="complete"/>
    </global>
    <classifier name="MXML Legacy Classifier"
    keys="concept:name lifecycle:transition"/>
    <classifier name="Event Name" keys="concept:name"/>
    <classifier name="Resource" keys="org:resource"/>
    <string key="source" value="Rapid Synthesizer"/>
    <string key="concept:name" value="excercise1.mxml"/>
    <string key="lifecycle:model" value="standard"/>
    <trace>
        <string key="concept:name" value="Case3.0"/>
        <event>
          <string key="org:resource" value="Shell"/>
          <date key="time:timestamp" value="2016-05-12"/>
          <string key="concept:name"
           value="subsea flow lines sprung a leak"/>
          <string key="lifecycle:transition"
          value="complete"/>
        </event>
        <event>
          <string key="org:resource" value="Shell"/>
          <date key="time:timestamp" value="2016-05-12"/>
```

```
        <string key="concept:name"
        value="oil spilled into the Gulf of Mexico"/>
        <string key="lifecycle:transition"
        value="complete"/>
      </event>
      <event>
        <string key="org:resource" value="Shell"/>
        <date key="time:timestamp" value="2016-05-19"/>
        <string key="concept:name"
        value="cleanup operation ends"/>
        <string key="lifecycle:transition"
        value="complete"/>
      </event>
  </trace>
</log>
```

Petri net model generated according to this log for request "Shell Spills Oil in the Gulf" is shown below (Fig. 5).

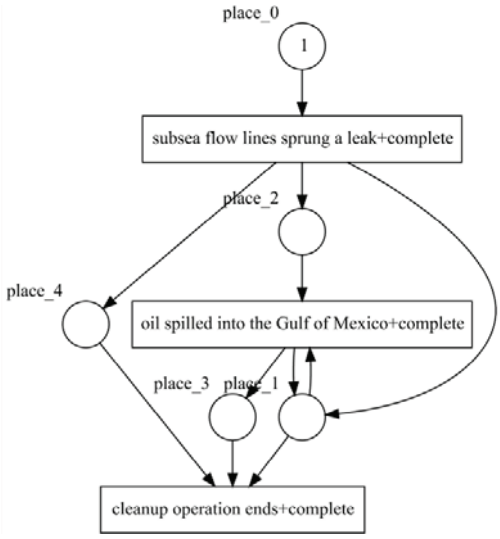


Fig. 5. Petri net model for request "Shell Spills Oil in the Gulf"

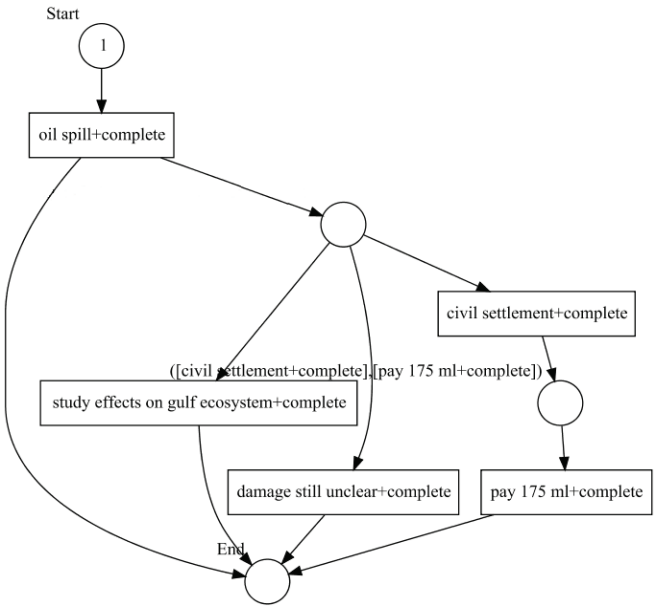


Fig. 6. Model for information request enhanced by parameter Place

Further presented the model of the same process in the context of geographical position, i.e. request was extended by obtained in the previous step attribute Place = Mexico. The results are shown in Fig. 6. In the model appeared some additional information on the socio-economic developments related to the assessment of damage and the imposition of a fine. Thus, clarifying and expanding the information requests the user fills the database with new data and gets more detailed models.

IX. Conclusion

An approbation of the offered approach for the analysis of the processes associated with the environmental technogenic catastrophes caused with oil spills showed prospects of the developed means. Ontologies allow performing flexible tuning for various domains, for example to investigate relations between events in the field of economy, policy and etc.

References

[1] M. Leemans; W.M.P. van der Aalst. Process Mining in Software Systems: Discovering Real-Life Business Transactions and Process Models from Distributed Systems // In: Proceedings of 18th International Conference on Model Driven Engineering Languages and Systems (MODELS), 2015. Pp.44-53.

[2] P. Mukala, J. Buijs, M. Leemans, W. van der Aalst. Learning Analytics on Coursera Event Data: A Process Mining Approach. In: Proceedings 5th International Symposium on Data-driven Process Discovery and Analysis 2015. Pp.18-32.

[3] D. Calvanese, M. Montali, A. Syamsiyah, W.M.P. van der Aalst. Ontology-Driven Extraction of Event Logs from Relational Databases. In: Business Process Management Workshops 2015.

[4] R.S. Mans. Workflow Support for the Healthcare Domain. PhD Thesis. Technische Universiteit Eindhoven, Eindhoven, 2011.

[5] N.A. Samojlenko. Semantika sobytijnosti i sposoby ee vyrazheniya: avtoref. dis. kand. filol. nauk. Alma-Ata, 1991.

[6] V.E. Goldin. Imena rechevuh sobytiy, postupkov i zhanry russkoi rechi // Zhanry Rechi. – Saratov, 1977. – Pp.23-34.

[7] P.P. Maslov. obnaruzhenie i izvlechenie prichinno-sledstvennyh zakonomernostej iz teksta na estestvennom yazyke. In: Proceedings of Conference "Znaniya-Ontologii-Teorii". 2009.

[8] Draft Standard for XES - eXtensible Event Stream - for achieving interoperability in event logs and event streams. The Institute of Electrical and Electronics Engineers, 2016.

[9] V. Peña-Araya. Galean: Visualization of Geolocated News Events from Social Media / V. Peña-Araya, M. Quezada, B. Poblete // Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15). ACM New York. 2015. Pp. 1041-1042.

[10] M. Schuhmacher. Finding Relevant Relations in Relevant Documents. In: Advances in Information Retrieval: Proceedings of 38th European Conference on IR Research, ECIR. Padua, Italy, March 20-23, 2016. Pp. 654-660.

[11] A. Vokhmintsev, A. Melnikov. The Knowledge on the Basis of Fact Analysis in Business Intelligence. In: Digital Product- and Process Development Systems.– 2013.– Pp.134-141.

Approach to Defining the Maximum Amount of Flow in Dynamic Fuzzy Network with the Given Vitality Degree

Bozhenyuk Alexander, Gerasimenko Evgeniya
Southern Federal University
Taganrog, Russia
Avb002@yandex.ru, egerasimenko@sfedu.ru

Igor Rozenberg
Public corporation "Research and development institute of railway engineers"
Moscow, Russia
i.yarosh@vniias.ru

***Abstract*— Present paper is devoted to defining the maximum amount of flow in dynamic graph with bounds of flow not equal to zero with the given vitality degree. The approach takes into account fuzziness of the network's parameters, values and measurements. The considered approach allows taking into account profitability of transportation via setting the bounds of flow not equal to zero. Peculiarity of present task is in considering the reliability of transportation, expressed by the vitality parameter and considering arcs with the required vitality. Considered graph is described as fuzzy lower and upper-capacitated network with vitality degrees and crisp values of the time-parameters. All parameters of the considered network can vary over time. These methods can be applied in the real railways, roads and air roads solving the task of the optimal cargo transportation.**

***Index Terms*— Fuzzy network, fuzzy lower flow bound, fuzzy vitality degree.**

I. Introduction

Nowadays the flow problems and appropriate problem statements [1] are very important, as they allow state and solve various flow problems from searching the maximum flow to finding the flow of the minimum cost and minimum cost routes of flow transmission in the different types of networks.

Important sphere of researches is dynamic networks [2-4], that consider time parameters set to the arcs. Another significant tool is considering dependence of bounds of flow on flow departure time [5] and operating with fully dynamic networks instead of stationary-dynamic ones [6], using the notions of the time-expanded graphs [7-8].

Flow problems are connected with uncertainty of some kind, because changes in environment, measurement errors influence bounds of flow and vitality parameters. Therefore, we propose to examine these problems in terms of fuzziness, therefore, we turn to the fuzzy graphs for solving such problems.

Vitality parameter [9-10] peculiar to arcs of the network usually isn't taken into account while studying networks. Its conventional definition was introduced by the authors H. Frank and I. Frisch in [11] as sensitivity of the network to damages. However, vitality applied to the networks is ability of its objects and links among them to be resistant to weather conditions, traffic accidents and its combinations, and save and restore (fully or partially) objects themselves and their connections, arc capacities of the network's sections in case of damage. Nowadays, vitality of the network isn't taken into account, while railways and roads include the complex objects, such as stations, distillation ways, culverts, wagon, passenger and cargo managements. Sometimes network's parameters can be set qualitatively. Thus, one can set the notion "vitality degree" considering the roads and railways. In this case "vitality degree" is considered as probability of trouble-free operation of the road section and some subjective value, such as importance and reliability, etc.

Other words this paper presents method of determining the maximum amount of flow in dynamic network under fuzzy conditions with given vitality degree.

Present paper is structured in such a way. Main concepts and notions are presented in the Section 2. Section 3 presents method of finding the flow under fuzzy conditions with vitality degree. Section 4 provides numerical example illustrating the main steps of the method. Section 5 observes conclusion and future work.

II. Basic concepts and notions

Let us present basic rules and definitions of this method.

Rule 1 of turning from the expanded network to fuzzy graph without bounds of flow [12].

Go to fuzzy graph $\tilde{G}_p^* = (X_p^*, \tilde{A}_p^*)$ from $\tilde{G}_p = (X_p, \tilde{A}_p)$ adding the artificial vertices s^* and t^* and arcs going from the node $(t, \forall\theta \in \mathrm{T})$ to $(s, \forall\theta \in \mathrm{T})$ with $\tilde{u}^*(t, s,\ \forall\theta \in T,\ \forall\theta \in T) = \infty$, $\tilde{l}^*(t,\ s,\ \forall\theta \in T,\ \forall\theta \in T) = \tilde{0}$, $\tilde{v}^*(t, s, \forall\theta \in T, \forall\theta \in T) = \tilde{1}$ in the graph $\tilde{G}_p$. For arcs with $\tilde{l}(x_i, x_j, \theta, \vartheta) \neq \tilde{0}$: 1) reduce $\tilde{u}(x_i, x_j, \theta, \vartheta)$ to $\tilde{u}^*(x_i, x_j, \theta, \vartheta) = \tilde{u}(x_i, x_j, \theta, \vartheta) - \tilde{l}(x_i, x_j, \theta, \vartheta)$, $\tilde{l}(x_i, x_j, \theta, \vartheta)$ to $\tilde{0}$, $v^*(x_i, x_j, \theta, \vartheta) = v(x_i, x_j, \theta, \vartheta)$. 2) Enter the arcs linking s^* and (x_j, ϑ), arcs going from t^* to (x_i, θ)

that have $\tilde{u}^*(s^*, x_j, \forall, \vartheta) = \tilde{u}^*(x_i, t, \theta, \forall) = \tilde{l}(x_i, x_j, \theta, \vartheta)$, $\tilde{l}^*(s^*, x_j, \forall, \vartheta) = \tilde{l}^*(x_i, t, \theta, \forall) = \tilde{0}$, $\tilde{v}^*(x_i, x_j, \theta, \vartheta) = \tilde{v}(x_i, x_j, \theta, \vartheta)$.

Definition 1

Fuzzy residual network $\tilde{G}_p^{*\mu} = (X_p^{*\mu}, \tilde{A}_p^{*\mu})$ is the network without bounds of flow $\tilde{G}_p^* = (X_p^*, \tilde{A}_p^*)$, which is constructed according to the following rules: if

$$\begin{cases} \tilde{\xi}^*(x_i, x_j, \theta, \vartheta) < \tilde{u}^*(x_i, x_j, \theta, \vartheta), \\ \tilde{v}^*(x_i, x_j, \theta, \vartheta) \geq \tilde{v}_{req}, \end{cases}$$

then include corresponding arc from $(x_i^{*\mu}, \theta)$ to $(x_j^{*\mu}, \vartheta)$ in $\tilde{G}_p^{*\mu}$ with $\tilde{u}^{*\mu}(x_i, x_j, \theta, \vartheta) = \tilde{u}^*(x_i, x_j, \theta, \vartheta) - \tilde{\xi}^*(x_i, x_j, \theta, \vartheta)$, $\tau^{*\mu}(x_i, x_j, \theta, \vartheta) = \tau^*(x_i, x_j, \theta, \vartheta)$.

If

$$\begin{cases} \tilde{\xi}^*(x_i, x_j, \vartheta, \theta) > \tilde{0}, \\ \tilde{v}(x_i, x_j, \theta, \vartheta) \geq \tilde{v}_{req}. \end{cases}$$

Then include the corresponding arc from $(x_j^{*\mu}, \vartheta)$ to $(x_i^{*\mu}, \theta)$ in $\tilde{G}_p^{*\mu}$ with $\tilde{u}^{*\mu}(x_j, x_i, \vartheta, \theta) = \tilde{\xi}^*(x_i, x_j, \theta, \vartheta)$ and $\tau^{*\mu}(x_j, x_i, \vartheta, \theta) = -\tau^*(x_i, x_j, \theta, \vartheta)$.

Rule 2 of obtaining the admissible flow

Go to the network $\tilde{G}_p$ from $\tilde{G}_p^*$ in such a way: delete all nodes and arcs that are artificial. The feasible flow $\tilde{\xi} = (\tilde{\xi}(x_i, x_j, \theta, \vartheta))$ of the value $\tilde{\sigma}$ is defined as: $\tilde{\xi}(x_i, x_j, \theta, \vartheta) = \tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{l}(x_i, x_j, \theta, \vartheta)$, where $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta)$ – the flows, passing the arcs of the graph without bounds of flow.

Due to proposed notions, go to the developed algorithm.

III. Proposed method

Examine the task of determining the maximum amount of flow in dynamic graph in terms of fuzziness and vagueness and given vitality degree, presented by the model (1)-(6).

$$Maximize\ \tilde{v}(p) \quad (1)$$

$$\sum_{\theta=0}^{p}\left(\sum_{x_j \in \Gamma(x_i)} \tilde{\xi}_{ij}(\theta) - \sum_{x_j \in \Gamma^{-1}(x_i)} \tilde{\xi}_{ji}(\theta - \tau_{ji}(\theta))\right) = \tilde{v}(p), x_i = s, \quad (2)$$

$$\sum_{\theta=0}^{p}\left(\sum_{x_j \in \Gamma(x_i)} \tilde{\xi}_{ij}(\theta) - \sum_{x_j \in \Gamma^{-1}(x_i)} \tilde{\xi}_{ji}(\theta - \tau_{ji}(\theta))\right) = \tilde{0}, x_i \neq s, t; \theta \in T, \quad (3)$$

$$\sum_{\theta=0}^{p}\left(\sum_{x_j \in \Gamma(x_i)} \tilde{\xi}_{ij}(\theta) - \sum_{x_j \in \Gamma^{-1}(x_i)} \tilde{\xi}_{ji}(\theta - \tau_{ji}(\theta))\right) = -\tilde{v}(p), x_i = t, \quad (4)$$

$$\tilde{l}_{ij}(\theta) \leq \tilde{\xi}_{ij}(\theta) \leq \tilde{u}_{ij}(\theta), \theta + \tau_{ij}(\theta) \leq p, \theta \in T, \quad (5)$$

$$\tilde{v}_{ij}(\theta) \geq \tilde{v}_{req}, s(\theta) + \tau_{st}(\theta) \leq p, \theta \in T. \quad (6)$$

Step 1. Turn to the graph $\tilde{G}_p$, which is presented as fuzzy static one and expanded in the time dimension from the initial dynamic graph $\tilde{G}$.

Step 2. Build graph $\tilde{G}_p^* = (X_p^*, \tilde{A}_p^*)$ due to the *rule 1.*

Step 3. Obtain graph $\tilde{G}_p^{*\mu}$ in the form of the residual network due to the definition 1.

Step 4. Find a path $\tilde{P}_p^{*\mu}$ from the source to the sink in the network, obtained in the step3 due to the breadth-first-search.

4.1 If such a path $\tilde{P}_p^{*\mu}$ is obtained, go to the step 5.

4.2 If we find amount of flow value $\tilde{\phi}^* < \sum_{\tilde{l}(x_j, x_i, \vartheta, \theta) \neq \tilde{0}} \tilde{l}(x_i, x_j, \theta, \vartheta)$, there is no augmenting path. Exit.

Step 5. Pass the flow value $\tilde{\delta}_p^{*\mu} = \min[\tilde{u}(\tilde{P}_p^{*\mu})]$, $\tilde{u}(\tilde{P}_p^{*\mu}) = \min[\tilde{u}^{*\mu}(x_i, x_j, \theta, \vartheta)$, $(x_i, \theta), (x_j, \vartheta) \in \tilde{P}_p^{*\mu}$ along this path $\tilde{P}_p^{*\mu}$.

Step 6. Replace the values of flow in $\tilde{G}_p^*$: value $\tilde{\xi}^*(x_j, x_i, \theta, \vartheta)$ from (x_j^*, θ) to (x_i^*, ϑ) by $\tilde{\xi}^*(x_j, x_i, \theta, \vartheta) - \tilde{\delta}_p^{*\mu}$ for arcs from $(x_i^{*\mu}, \vartheta)$ to $(x_j^{*\mu}, \theta)$ in $\tilde{G}_p^{*\mu}$, such as $((x_i^{*\mu}, \vartheta), (x_j^{*\mu}, \theta)) \notin \tilde{A}_p^*$, $((x_i^{*\mu}, \vartheta), (x_j^{*\mu}, \theta)) \in \tilde{A}_p^{*\mu}$ and value $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta)$ from (x_i^*, θ) to (x_j^*, ϑ) by $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{*\mu}$ for arcs from $(x_i^{*\mu}, \theta)$ to $(x_j^{*\mu}, \vartheta)$ in $\tilde{G}_p^{*\mu}$, such as $((x_i^{*\mu}, \theta), (x_j^{*\mu}, \vartheta)) \in \tilde{A}_p^*$, $((x_i^{*\mu}, \theta), (x_j^{*\mu}, \vartheta)) \in \tilde{A}_p^{*\mu}$. Replace $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta)$ by $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{*\mu} \tilde{P}_p^{*\mu}$.

Step 7. Compare flow value $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{*\mu} \tilde{P}_p^{*\mu}$ and $\sum_{\tilde{l}(x_j, x_i, \vartheta, \theta) \neq \tilde{0}} \tilde{l}(x_i, x_j, \theta, \vartheta)$:

7.1. If $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{*\mu} \tilde{P}_p^{*\mu}$ (the value of flow) is less than $\sum_{\tilde{l}(x_j, x_i, \vartheta, \theta) \neq \tilde{0}} \tilde{l}(x_i, x_j, \theta, \vartheta)$, turn to the 3 step.

7.2. If the found value of flow $\tilde{\xi}^*(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{*\mu} \tilde{P}_p^{*\mu}$ is equal to $\sum_{\tilde{l}(x_j, x_i, \vartheta, \theta) \neq \tilde{0}} \tilde{l}(x_i, x_j, \theta, \vartheta)$, go to the network $\tilde{G}_p$ from the network $\tilde{G}_p^*$, as stated in *rule 2*. Go to step 8.

Step 8. Build residual graph $\tilde{G}_p^{\mu}(\tilde{\xi})$.

Step 9. Find the path $\tilde{P}_p^{\mu}$ in $\tilde{G}_p^{\mu}(\tilde{\xi})$, shortest in terms of arcs quantity.

9.1. If $\tilde{P}_p^{\mu}$ is obtained, turn to the step 10.

9.2. If an augmenting path in the graph $\tilde{G}_p(\tilde{\xi})$ doesn't exist, the maximum amount of flow $\tilde{\xi}(x_i, x_j, \theta, \vartheta) + \tilde{\delta}_p^{\mu} \tilde{P}_p^{\mu} = \tilde{v}(p)$ is obtained. Turn to step 12.

Step 10. Transfer $\tilde{\delta}_p^{\mu} = \min[\tilde{u}(\tilde{P}_p^{\mu})]$, $\tilde{u}(\tilde{P}_p^{\mu}) = \min[\tilde{u}^{\mu}(x_i, x_j, \theta, \vartheta)\ (x_i, \theta), (x_j, \vartheta) \in \tilde{P}_p^{\mu}$ on the shortest path.

Step 11. Recalculate amount of flow in the network $\tilde{G}_p(\tilde{\xi})$.

Step 12. Go to the original dynamic network $\tilde{G}$.

IV. Numerical example

Let us represent proposed algorithm. Let the initial fuzzy dynamic network is presented in Fig. 1. A problem is in finding the maximum value of flow in the initial dynamic network with the given vitality degree no less than $0,\tilde{7}$ and present final value as triangular number.

Values of upper bounds of flow $\tilde{u}_{ij}$, which are fuzzy and can vary over time θ, are given in the Table I. Values of lower bounds of flow $\tilde{l}_{ij}$, which are fuzzy and can vary over time θ, are given in the Table II. Time parameters τ_{ij}, which can change their values in time θ, are presented in the Table III. Fuzzy vitality parameters $\tilde{v}_{ij}$, which can change their values in time θ, are indicated in the Table IV.

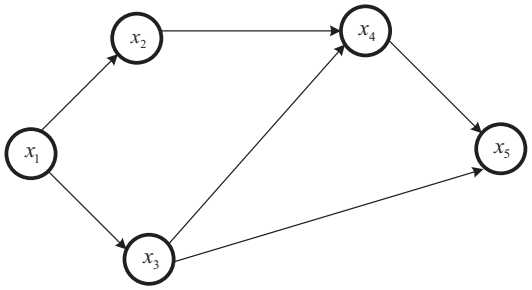


Fig. 1. Initial dynamic graph $\tilde{G}$

Construct time-expanded network according to Fig. 2.

Go to the network consists of the arcs with whole upper flow bounds and find the augmenting paths in the graph in Fig. 3: $\tilde{P}_1^{*\mu} = s^*, (x_5, 2), (x_1, 0), t^*$ with $\tilde{7}$ flow units, $\tilde{P}_2^{*\mu} = s^*, (x_2, 1), (x_3, 2), (x_5, 3), (x_1, 0), t^*$ with $\tilde{3}$ flow units, $\tilde{P}_3^{*\mu} = s^*, (x_2, 1), (x_3, 2), (x_5, 3), (x_1, 0), (x_4, 1), t^*$ with $\tilde{7}$ flow units.

We obtain graph with maximum value of flow, as presented in Fig. 4. Therefore, the task has a solution and we turn to initial graph, as presented in Fig. 5. Finding the augmenting ways and passing flows along them, we obtain graph in Fig. 6.

The maximum value of flow in the original dynamic network with the value of vitality degree no less than $0,\tilde{7}$ is $2\tilde{5} + 1\tilde{0} = 3\tilde{5}$ flow units.

TABLE I. Upper Bounds of Flow $\tilde{u}_{ij}$ in the Fuzzy Form, Vary in Time θ

Arcs	Values of upper bounds of flow $\tilde{u}_{ij}$ in the fuzzy form at time θ, units.			
	0	*1*	*2*	*3*
(x_1, x_2)	$2\tilde{5}$	$2\tilde{0}$	$2\tilde{5}$	$4\tilde{0}$
(x_1, x_4)	$1\tilde{0}$	$2\tilde{0}$	$2\tilde{5}$	$2\tilde{5}$
(x_1, x_5)	$1\tilde{8}$	$1\tilde{8}$	$3\tilde{0}$	$3\tilde{5}$
(x_2, x_3)	$3\tilde{5}$	$3\tilde{0}$	$3\tilde{5}$	$1\tilde{8}$
(x_3, x_4)	$1\tilde{5}$	$2\tilde{7}$	$3\tilde{3}$	$2\tilde{5}$
(x_3, x_5)	$5\tilde{5}$	$4\tilde{5}$	$4\tilde{0}$	$5\tilde{5}$
(x_4, x_5)	$2\tilde{0}$	$2\tilde{0}$	$1\tilde{8}$	$2\tilde{8}$

TABLE II. Lower Bounds of Flow $\tilde{l}_{ij}$ in the Fuzzy Form, Vary in Time θ.

Arcs	Values of lower bounds of flow $\tilde{u}_{ij}$ in the fuzzy form at time θ, units.			
	0	*1*	*2*	*3*
(x_1, x_2)	$1\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$
(x_1, x_4)	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$
(x_1, x_5)	$\tilde{0}$	$\tilde{0}$	$\tilde{0}$	$2\tilde{0}$
(x_2, x_3)	$\tilde{6}$	$\tilde{0}$	$1\tilde{5}$	$\tilde{0}$
(x_3, x_4)	$\tilde{0}$	$\tilde{8}$	$\tilde{0}$	$\tilde{0}$
(x_3, x_5)	$2\tilde{5}$	$1\tilde{5}$	$\tilde{0}$	$\tilde{0}$
(x_4, x_5)	$\tilde{0}$	$\tilde{5}$	$\tilde{0}$	$1\tilde{0}$

TABLE III. Parameters of Time τ_{ij}, Vary in Time θ

Arcs	Parameters of time τ_{ij} at time θ, units of time.			
	0	*1*	*2*	*3*
(x_1, x_2)	1	1	1	2
(x_1, x_4)	1	3	2	2
(x_1, x_5)	4	4	1	1
(x_2, x_3)	4	1	1	1
(x_3, x_4)	1	1	2	2
(x_3, x_5)	2	2	1	1
(x_4, x_5)	5	4	1	3

Calculate borders of deviation of the found fuzzy number "near $3\tilde{5}$".

Since the calculations with fuzzy numbers are cumbersome, we suggest operating fuzzy numbers according to the method, described in [8]. In this case we will operate the central values of numbers given in fuzzy form, blurring the found value at the last step and presenting it as a triangular the number.

Therefore, borders of deviation of the calculated fuzzy amount of flow "near $3\tilde{5}$" corresponded to the maximum

amount of flow in the network $\tilde{G}$ are calculated according to the basic arc capacities' values in Fig. 7.

TABLE IV. FUZZY VITALITY PARAMETERS v_{ij}, VARY IN TIME θ

Arcs	Fuzzy vitality parameters $\tilde{v}_{ij}$ at time θ, vitality units			
	0	*1*	*2*	*3*
(x_1,x_2)	$0,\tilde{8}$	$0,\tilde{4}$	$0,\tilde{6}$	$0,\tilde{5}$
(x_1,x_4)	$0,\tilde{7}$	$0,\tilde{2}$	$0,\tilde{8}$	$0,\tilde{9}$
(x_1,x_5)	$0,\tilde{4}$	$0,\tilde{8}$	$0,\tilde{6}$	$0,\tilde{3}$
(x_2,x_3)	$0,\tilde{7}$	$0,\tilde{8}$	$0,\tilde{7}$	$0,\tilde{4}$
(x_3,x_4)	$0,\tilde{7}$	$0,\tilde{9}$	$0,\tilde{7}$	$0,\tilde{6}$
(x_3,x_5)	$0,\tilde{3}$	$0,\tilde{4}$	$0,\tilde{7}$	$0,\tilde{4}$
(x_4,x_5)	$0,\tilde{8}$	$0,\tilde{3}$	$0,\tilde{3}$	$0,\tilde{4}$

The detected value is among adjacent values of the arc capacities: $3\tilde{1}$ that has the deviation to the left equals $l_1^L = 8$, right deviation – $l_1^R = 7$ and $4\tilde{4}$ that has deviation to the left $l_2^L = 9$, deviation to the right $-l_2^R = 10$. We obtain deviations: $l_1^L \approx 8$, $l_1^R \approx 7$.

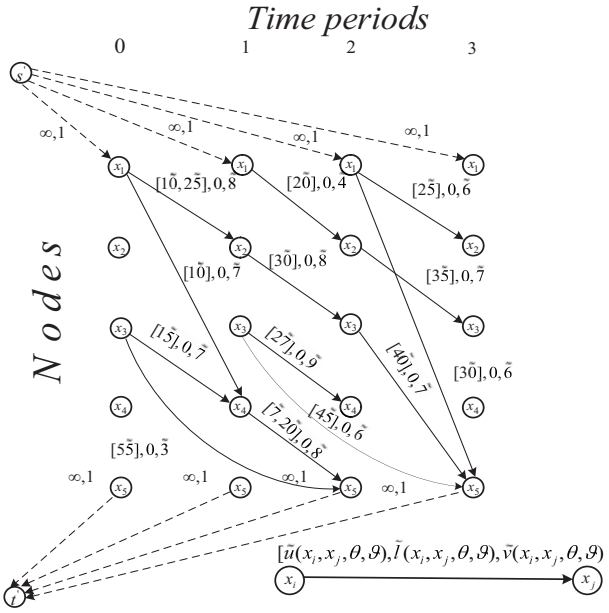


Fig. 2. Time-expanded graph $\tilde{G}_p$

Therefore, the maximum value of flow in dynamic network with the given vitality degree no less than $0,\tilde{7}$ can be represented by fuzzy triangular number (27, 35,42) units.

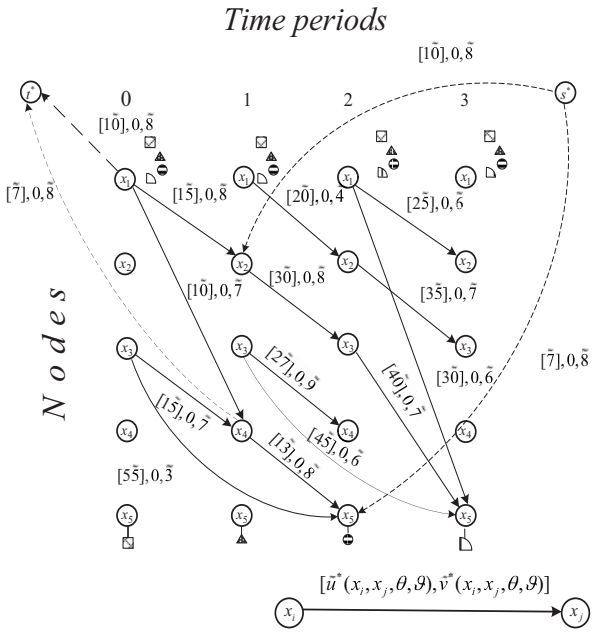


Fig. 3. $\tilde{G}_p^*$ – Time-expanded network

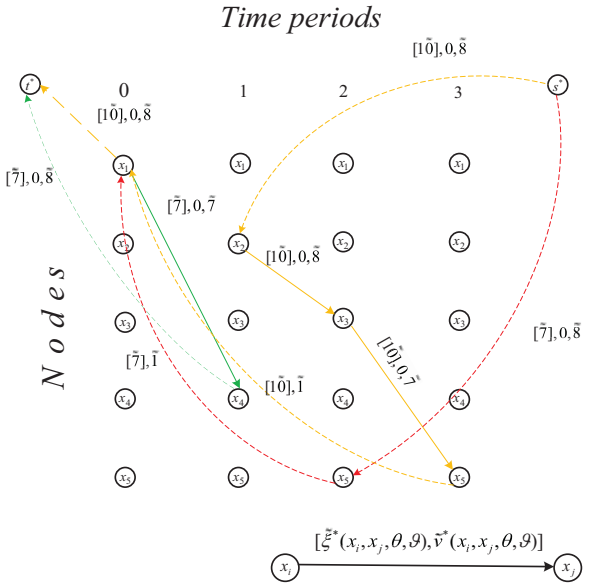


Fig. 4. Network $\tilde{G}_p^*$ consists of the arcs with the maximum flow

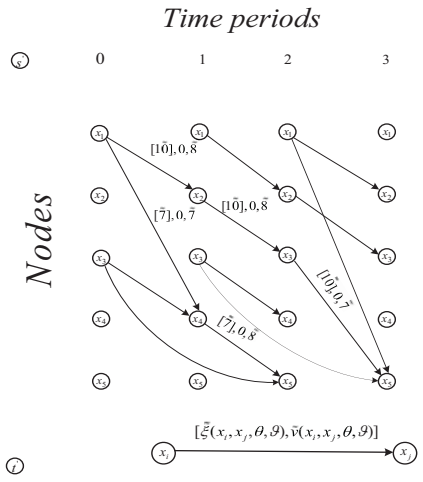


Fig. 5. Graph $\tilde{G}_p$ with the feasible flow

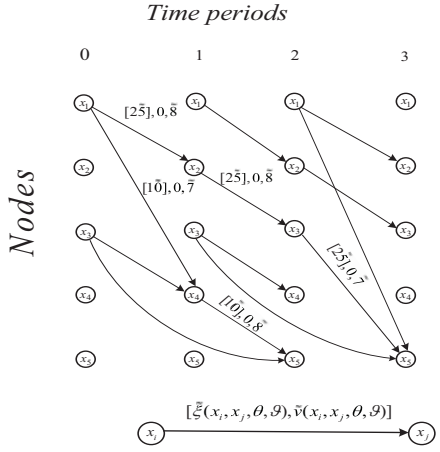


Fig. 6. Graph $\tilde{G}_p$ with the maximum flow

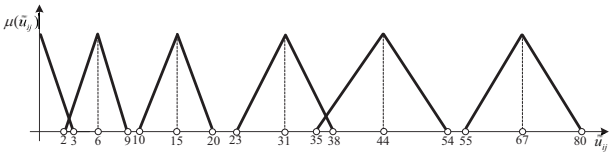


Fig. 7. Functions of membership of arc capacities' basic values in $\tilde{G}$

V. Conclusion and Future Work

Present paper observes a task of determining the maximum value of flow in a dynamic graph, presented as a network in fuzzy conditions, which arcs have flow bounds that are nonzero, with given vitality degree. Parameters of the network depend on the departure time of flow and can vary over time and described as fuzzy ones The approach to the task of obtaining the maximum amount of flow in the graph that has bounds of flow that are nonzero, based on the introduced definitions and rules. The solution of the problem has important practical value while planning transportation, as allows to consider profitability of transmission and fuzzy - dynamic nature of its parameters. In the future works we are going to propose methods of increasing the vitality degree in fuzzy dynamic networks.

Acknowledgment

Present work has been supported by the Russian Foundation for Basic Research, Project № 16-01-00090 a, and the Ministry of Education and Science of the Russian Federation under Project № 213.01-11/2014-48 (Base part, State task 2014/174)

References

[1] Ahuja T.L. Magnanti, J. B. Orlin, Network Flows: Theory, Algorithms, and Applications. New Jersey: Prentice-Hall, Inc., 1993, 846 p.

[2] S. Lovetskii, I. Melamed, "Dynamic flows in networks", Automation and Remote Control, Vol. 48(11), 1987, pp. 1417–1434; Translated from Avtomatika i Telemekhanika, vol. 11, pp. 7-29.

[3] Halpern, J. "A generalized dynamic flows problem". Networks, vol. 9, 1979, pp. 133-167.

[4] B. Kotnyek, "An annotated overview of dynamic network flows", Technical report, INRIA, Paris, № 4936, 2003. http://hal.inria.fr/inria-00071643/.

[5] L. Chabini, "Discrete dynamic shortest path problems in transportation applications: Complexity and algorithms with optimal run time", Transport. Res. Record, vol. 1645, 1998, pp. 170-175.

[6] A. Bozhenyuk, E. Gerasimenko, I. Rozenberg, and I. Perfilieva, "Method for the minimum cost maximum flow determining in fuzzy dynamic network with nonzero flow bounds," in proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology, 30th June-3rd July, Gijón, Asturias (Spain), vol. 89, José M. Alonso, H. Bustince, M. Reformat, Eds: Atlantic Press, 2015, pp. 385–392.

[7] M. R. Silver and O. L. de Weck, "Time-expanded decision networks: A framework for designing evolvable complex systems," Systems Engineering, vol. 10, no. 2, 2007, pp. 167–188.

[8] A. Bozhenyuk, E. Gerasimenko, and I. Rozenberg, "Determining the minimum cost flow in fuzzy dynamic network with GIS «ObjectLand»," in Proceedings of the 2015 9th International Conference on Application of Information and Communication Technologies (AICT), 14-16 October. Rostov-on-Don, Russia, 2015, pp. 294–298.

[9] A. Bozhenyuk, I. Rozenberg, and T. Starostina, "A method of increase of vitality degree of fuzzy graph," in Proceedings of 13th Zittau Fuzzy Colloquium, September 13-15. Zittau: Hochschule Zittau/Goerlitz, 2006, pp. 235–240.

[10] V. Bozheniuk, A. Bozhenyuk, and S. Belyakov, "Optimum allocation of centers in fuzzy transportation networks with the largest vitality degree," in: Proceedings of the 2015 Conference of the International Fuzzy System Association and the European Society for Fuzzy Logic and Technology: Atlantis Press, 2015, pp. 1006–1011.

[11] H. Frank and I. T Frisch, Communication, Transmission, and Transportation Networks: Addison Wesley, 1971.

[12] A. Bozhenyuk and E. Gerasimenko, "Flows finding in networks in fuzzy conditions," in Supply Chain Management Under Fuzzines, Studies in Fuzziness and Soft Computing, vol. 313, part III, Cengiz Kahraman and Basar Öztaysi, Eds. Berlin, Heidelberg: Springer-Verlag, 2014, pp. 269–291.

VLSI Elements Placement Based on Simulation of Bats Behavior in Nature

Elmar Kuliev, Andrey Lezhebokov, Yuri Kravchenko
Department of CAD
Southern Federal University
Rostov-on-Don, Russia
elmar_2005@mail.ru, legebokov@gmail.com, krav-jura@yandex.ru

***Abstract* — The paper deals with NP-hard optimization problems with the use of behavioral foundations of subject in nature. Swarm intelligent methods and algorithms are the most promising ones. Suggested algorithm is based on the simulation of bats' behavior. The authors concern a behavioral model of a bat – it moves within search space for the purpose of finding quasi-optimal solutions. Also there are represented approaches to adjust control parameters of the algorithm (frequency, amplitude, direction of movement). On the basis of bat's behavior a schematic diagram is developed. Initial search space are formed by sequential, random and iterative algorithms. To confirm theoretical estimations of time complexity and effectiveness of the developed algorithm computational experiments are carried out.**

***Index Terms* — placement, adaptive procedures, evolutionary simulation, swarm algorithm, bats' behavior, software.**

I. Introduction

Nature has always been a source of inspiration for researchers. On the basis of natural principles and mechanisms there was developed a set of algorithm for solving different optimization problems. These algorithms belong to metaheuristic class and involve evolutionary and swarm intelligent techniques [1].

When dealing with practical problems, it is often necessary to choose the best solution according with some criteria. Mathematically, the choice is formalized as a optimization problem.

Well-known mathematical programming methods focused on optimization problems are deterministic iteration procedures for incremental improvement of a current solution. Effectiveness of such methods is based on objective function features in respect of optimization problem [1-3]. For many practical problems the use of deterministic methods are unreasonable. In this case there are used stochastic methods such as a genetic algorithm (GA), a particle swarm optimization algorithm (PSO), an ant colony optimization algorithm (ACO) etc. that can process a lot of current solutions simultaneously.

Multi-agent population algorithms deal with a set of alternative solutions [2]. Each solution is improved and estimated, so it has an effect on improvement of other alternative solutions. The most of population methods borrow this concept from biology: a process of the best solution search copies a natural search process or behavior of some species taking into account it specific features. These algorithms are named as bioinspired algorithms – a reach source of unusual numerical methods that allow to solve complex problems in the context of lack of information about optimized function [2-4, 5, 7-11].

The most swarm algorithms solve unconstrained optimization problems with real variables and have some parameters needed for one or another problem (for example, a size of alternative solutions population) [3].

The main aim of our research is investigation of effectiveness of swarm intelligent algorithm – a bat algorithm (BA).

II. Principles of Bat Behavior

The bat algorithm is a metaheuristic algorithm, created by X. S. Yang in 2010, inspired by echolocation behavior of bats [2-3].

The most of bats' species have echolocation ability for detecting food and obstacles. Parameters of sound impulse are changes in wide range according to different behavioral structure. Bats use short frequency-modulated signals within limits of one octave. But some species of bats does not use a frequency modulation of it sound impulse.

The BA conforms to the following rules:

1. all bats use echolocation to analyze a distance and distinguish food and natural obstacles;

2. Bats flies by a random way with velocity v_i at position x_i with a fixed frequency f_{min}, variable wavelength λ and loudness A_0. They may regulate a wavelength (of frequency) of impulse and impulse velocity $r[0,1]$.

3. Loudness is change from highest A_0 to lowest A_{min} value.

III. Model of Bats Algorithm

Fig. 1 illustrates a model of an optimization problem solution on the basis of bats swarm behavior. The placement problem is performed on a set of solutions X. Part of solutions may be infeasible under the initial conditions and restrictions (on fig.1 such areas specify as "obstacle"). Each search agent (a bat) has following characteristics: current position (x_i), velocity (v_i), loudness (A_0) and frequency (w_i). Loudness and

frequency of sound impulse are changed towards the some purpose (optimum or obstacle).

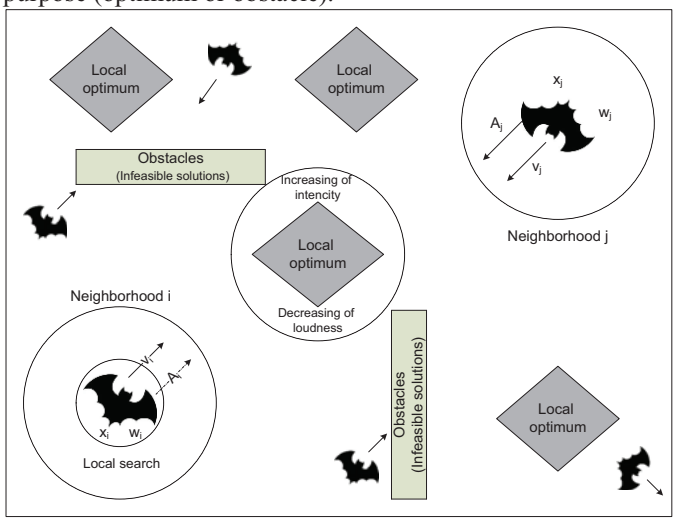


Fig. 1. Model of swarm algorithm based on bats' behavior

As the agent is coming closer to local (or global) optimum signal loudness decreases but its intensity is increased. After each iteration bats define it new position in local search area. To increase algorithm effectiveness in case of multidimensional search (encoding/decoding operation) we have an opportunity to choose random coordinates. All modifications in the algorithm deal with a change of control parameters (velocity, loudness, intensity) [12-15].

A number of bats is assigned by an operator and initial position and values of control parameters are assigned by generation of initial position. This procedure is implemented by combining the swarm algorithm and methods of random or iterative search [13, 16, 18].

IV. THE BAT ALGORITHM

The scheme of the developed swarm algorithm is shown on Fig. 2. As previously noted, the developed algorithm based on simulation of bats' behavior. The algorithm combines with classical methods of random and iterative search for generation of initial position of all search agents (Block #2) and generation of neighborhoods within search area.

For each bat generation of random initial values of velocity, loudness, intensity of signal is implemented in Block #3. The algorithm works iteratively and an iteration counter is verified in Block #12. In the Block #4 there are checked whether all bats were examined for the application of evolutionary rules.

In Blocks #5, 6, 7 it is performed an evolution of each bat according to its position and objective function value.

Intensity: $w_i = w_{min} + (w_{max} - w_{min}) * \beta$.

Velocity: $v_i^{(t+1)} = v_i^{(t)} + (x_i^{(t)} - x_i) * w_i$.

Position: $x_i^{(t+1)} = x_i^t + v_i^{(t+1)}$.

Each agent search its new position by random moving within current neighborhood:

$$x_i^{new} = x_i^{old} + e * A(t). \quad (1)$$

An evolution of parameters that characterize sound impulse is calculated as follows:

$$A_i^{(t+1)} = \alpha * A_i^{(t)} \quad (2)$$

Values of weighting coefficients are defined by decision making person.

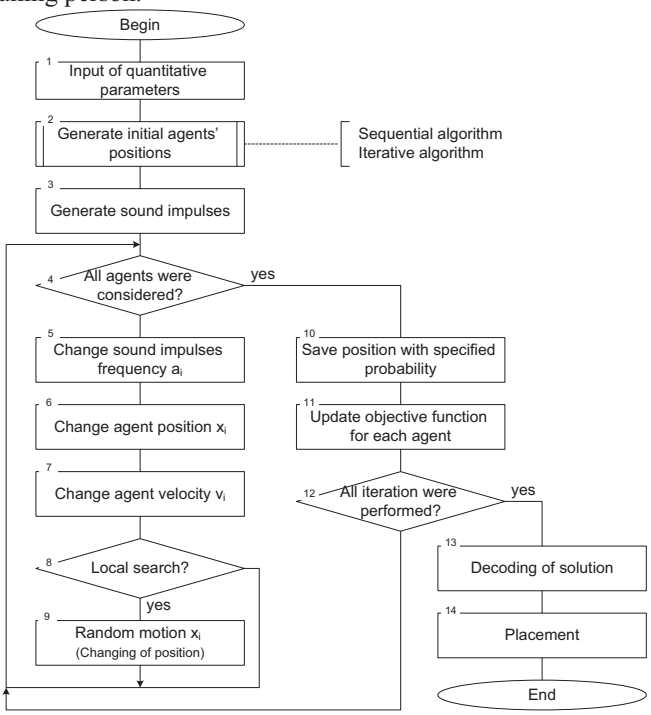


Fig. 2. Placement based on bat algorithm

In Block #8 it is generated a random probability and, in case of successful verification, it is transfer to Block #9 in which generated local solutions (it quantity are set by decision making person) within current neighborhood.

When all agents have been considered it is calculated it new position (on the basis of local search results) with given probability (Block #10). After that, for those agents which change its position there are defined objective function values. (Block #11).

As soon as the algorithm performs given numbers of iteration (Block #12) obtained solution are decoded (from numerical sequence to coordinates of discrete workspace) in Block #13. In Block #14 the best solution will be applied for the placement problem (Fig. 3).

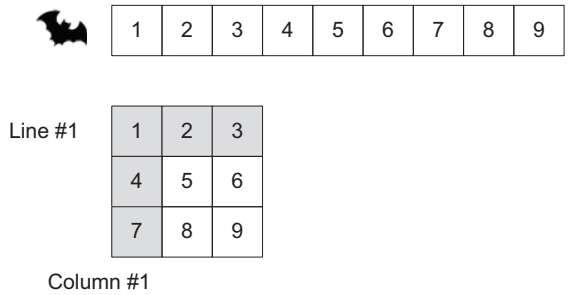


Fig. 3.Solution decoding for placement problem

The developed algorithm has several advantages in comparison with other swarm algorithms [17-19]. Firstly, it is the ability to take into account illegal solutions. Secondly, the algorithm controls an intensive of search by each agent which allow to obtain optimal solutions in appropriate time.

V. Experiments

Analyze and investigations of the developed algorithm were carried out on the basis of circuits with different numbers of elements. The main goal of experiments is calculation of algorithm time complexity [20].

To calculate time complexity there was carried out a set of experiments with changeable number of elements: from 100000 to 1000000 in increments 100000.

Table 1 and fig. 4 show the time complexity of the developed algorithm.

TABLE I. Comparison of time complexity of algorithms

	Number of elements								
	100 000	200 000	300 000	400 000	500 000	600 000	700 000	800 000	900 000
GA (s)	11,05	21,57	33,78	54,01	72,15	94,01	125,6	169,5	185,7
ACO (s)	21,4	52,12	71,82	110,1	180,5	205,9	298,8	359,6	404,9
BA (s)	35,58	84,45	157,56	248,78	351,7	409,9	558,9	694,8	857,7

The diagram of dependence of time complexity on number of elements confirms an assumption about time complexity of the developed algorithm which represented as $O(\alpha*n^2)$.

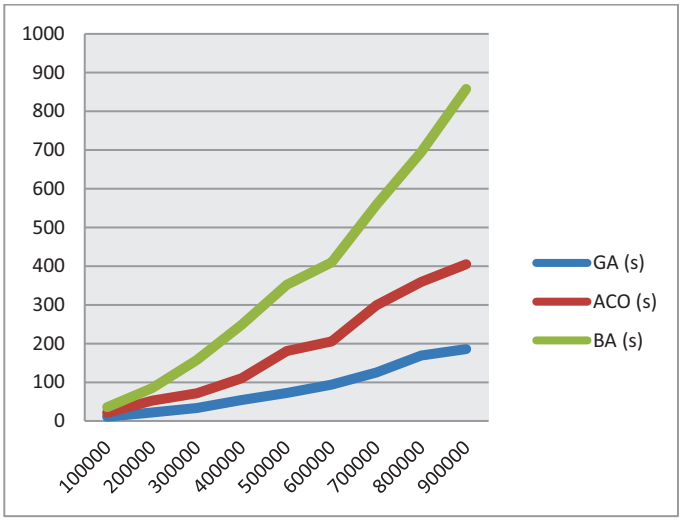


Fig. 4. The diagram of dependence of time complexity on number of elements

To obtain an effectiveness of the developed algorithm there were carried out investigations of solutions quality on the basis of test examples. The algorithm effectiveness will refer to quality of solutions obtained by this algorithm (Table 2). Fig. 5 shows a bar graph of obtained solutions.

TABLE II. Comparison of algorithm effectivness

	Number of elements				
	100000	250000	500000	750000	1000000
GA	87,7	128,4	142	171,4	215,7
ACO	75,1	110,1	133	154,9	203,8
BA	66,5	98,3	101,7	139,1	177,4

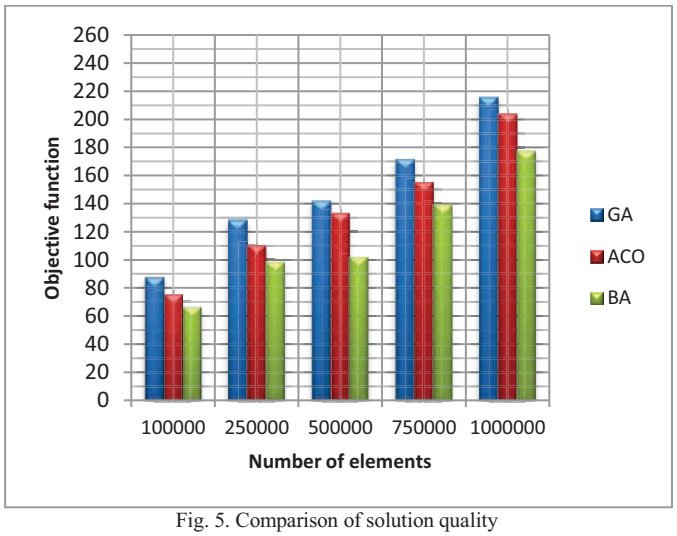


Fig. 5. Comparison of solution quality

Experimental results show that the most effective algorithm is the algorithm based on the behavior of bats. It is more efficient on average 15% than the GA and 10% than ACO.

VI. Conclusion

In the paper the authors suggested the new swarm algorithm based on the bats' behavior within restriction area to solve design optimization problems. The model of bat's behavior is considered. An agent is moving within search area to find local optimal solutions. New approaches to for changing control parameters of the algorithm (frequency, motion direction. amplitude) are developed. The scheme of bats' algorithm is made. Initial search space is formed by sequential, random and iterative algorithms. Conducted sets of experiments allow to specify theoretically estimations of the algorithm time complexity. In the best case time complexity of the developed algorithm is represented as O(nlogn), in the worst case - $O(n^2)$.

Acknowledgment

This research is supported by grants of the Ministry of Education and Science of the Russian Federation, the project #8.823.2014.

REFERENCES

[1] I.P. Norenkov and N.M. Arutyunyan, “Evolutionary methods in solving problems of selection of project”, Science and education: scientific publication Bauman MSTU #9, pp. 1-12, 2007.

[2] A.P. Karpenko, “Modern search optimization algorithms. Algorithms inspired by nature”, Moscow, p. 448, 2014.

[3] Sh.A. Ahmedova, “The effectiveness of "schooling" optimization algorithm”, Proceedings of the XIII Regional scientific student conference on mathematics and computer science, Krasnoyarsk, p. 9-12, 2010.

[4] V.V. Kureichik and D.Y. Zaporozhets, “Modern problems when placing elements of VLSI”, Proceedings of Southern Federal University. Technical science, vol. 7 (120), pp. 68-73, 2011.

[5] E.V. Kuliev and A.A. Lezhebokov, “Hybrid placement algorithm of VLSI components”, vol. 11 (136), pp. 188-192, 2012.

[6] V.V. Kureichik, V.V. Bova and V.V. Kureichik Jr., “Combined search for design problems”, Educational resources and technology, vol. 2(5), pp. 90-94, 2014.

[7] V.V. Kureichik and V.V. Kureichik Jr., “Bionic search in the design and management”, Proceedings of Southern Federal University. Technical science, vol. 11 (136), pp. 178-183, 2012.

[8] D.U. Zaporozhets, D.V. Zaruba and V.V. Kureichik, “Representation of solutions in genetic VLSI placement algorithms”, Proceedings of IEEE East-West Design and Test Symposium, EWDTS 2014.

[9] D.Y. Zaporozhets, D.V. Zaruba and V.V. Kureichik, “Hybrid bionic algorithms for solving problems of parametric optimization”, World Applied Sciences Journal, 23 (8), pp. 1032-1036, 2016

[10] E.V. Kuliev, A.N. Dukkardt, V.V. Kureychik and A.A. Legebokov “Neighborhood research approach in swarm intelligence for solving the optimization problems”, Proceedings of IEEE East-West Design and Test Symposium, EWDTS 2014

[11] V.V. Kureichik and D.Y. Zaporozhets, “Swarm algorithm for optimization problems”, Proceedings of Southern Federal University. Technical science, vol. 7 (108), pp. 28-32, 2010.

[12] V.V. Bova, A.A. Lezhebokov and L.A Gladkov,. “Problem-oriented algorithms of solutions search based on the methods of swarm intelligence”, World Applied Sciences Journal, 27 (9), pp. 1201-1205, 2013

[13] D. Zaruba, D. Zaporozhets and V. Kureichik, “VLSI placement problem based on ant colony optimization algorithm”, Advances in Intelligent Systems and Computing, 464, pp. 127-133, 2016

[14] V. Kureichik, V. Kureichik Jr. and V. Bova, “Placement of VLSI fragments based on a multilayered approach”, Advances in Intelligent Systems and Computing, 464, pp. 181-190, 2016

[15] V.V. Kureichik and D.V. Zaruba, “The bioinspired algorithm of electronic computing equipment schemes elements placement”, Advances in Intelligent Systems and Computing, 347, pp. 51-58, 2015.

[16] D.Y. Zaporozhets, D.V. Zaruba and V.V. Kureichik, “Hierarchical approach for VLSI components placement”, Advances in Intelligent Systems and Computing, 347, pp. 79-87, 2015

[17] D.U. Zaporozhets, D.V. Zaruba and V.V. Kureichik, “Representation of solutions in genetic VLSI placement algorithms”, Proceedings of IEEE East-West Design and Test Symposium, EWDTS 2014.

[18] E.V. Kuliev, A.A. Lezhebokov and A.N. Dukkardt, “The approach to the study of neighborhoods in the swarm algorithms for solving optimization problems”, Proceedings of Southern Federal University. Technical science, vol. 7 (156), pp. 15-26, 2014.

[19] E.V. Kuliev, D.Y. Zaporozhets and A.M. Ksalov, “Bioinspired search to solve the problem of accommodation of VLSI components”, Proceedings of the Kabardino-Balkar Scientific Centre of Russian Academy of Sciences, vol. 6(62), pp. 58-65, 2014.

[20] ,E.V. Kuliev and A.A. Lezhebokov, “Investigation of the characteristics of the hybrid placement algorithm”, Proceedings of Southern Federal University. Technical science, vol. 3, pp. 255-261, 2013.

Fuzzy Method of Revealing Implicit Causality within a Control System

Dmitry M. Nazarov, Svetlana V. Begicheva, Dmitry A. Azarov
Business Information Science Dept.
Ural State University of Economics
Yekaterinburg, Russia
slup2005@mail.ru

***Abstract*—The article addresses the problem of identifying subtle causal relations between sets of data defining a control system. The key pros and cons of using a fuzzy method are later examined. It explains the primary aspects of the fuzzy model for revealing implicit causality within a control system based on a combination of a fuzzy set approach and data mining techniques. The tool for carrying out evaluation in natural language resulting from utilizing an automated learning algorithm is advanced afterwards. The paper lastly proves its applicability to modeling mutual effects of a number of economic objects on each other.**

***Index Terms*—Causality, control system, data mining, fuzzy model, implicit factor.**

I. Introduction

Discovering control systems' patterns of behavior becomes one of the main priorities of latest studies, as it enables to anticipate their expected features and proprieties, and to predict likely future states. The information, collected in such a way and concerning aspects of the system interaction with the external environment, is vitally crucial to accomplishing realistic tasks, connected to gaining control of the system performance, notably to eliminating major constraints and achieving the results set. Nevertheless, in a range of examples, describing the system prospective conduct based exclusively on the basic features is complicated by its multifunction nature, significant number of elements, proprieties, connections and relations, existing implicit influences, complex structure etc. Moreover, when producing relatively highly accurate findings the need for processing big data [1] unavoidably arises resulting from specifying the condition of the system internal and ambient environments, consequently imposing the requirement for Business Intelligence [2] type solutions and leading to extra sophistication. With this background test object simulation is considered as a promising way of resolving the difficulties and determining causality among the system objects depending on relations between the exerted effect and the reaction assessed via data mining techniques. We suggest employing the fuzzy method for indicating implicit causality between the control system objects.

Throughout this paper implicit influences mean non-evident impacts within the system caused by implicit factors and capable of producing a synergistic effect. Implicit factors are considered as non-evident factors, having a significant effect on the system performance based on subtle information, previously not taken into account, which is practically useful and available for generating knowledge and making decisions [3].

II. Advantages and Disadvantages of Using Fuzzy Techniques

Despite the existence of a significant range of methodological approaches to establishing causal relations between the control system objects (see studies by C.W.J. Granger [4], R.E. Bellman [5], I.Yu. Shpolyanskaya [6], V.A. Kharitonov, A.O. Alekseyev [7], A.M. Malyshenko [8] etc.), the authors believe the developed tools do not completely recognize the uncertainty factor. This feature may be critical in conditions of the implicitness. Therefore, we should focus on application of fuzzy set means, which allow us to overcome the abovementioned disadvantages in the analysis.

Application of fuzzy systems has the following advantages: possibility of studying the parameters beyond the conventional formalization methods; development of solutions in conditions nonsufficient to perform another data analysis (e.g. statistical analysis); application of crude data in calculations; versatility [9–10] etc. [11–12]. Moreover, the fuzzy representation of the model indices improves the study results, allowing us to evaluate not only the index values, but the verisimilitude of each value and its degree of confidence as well. The fuzzy logic enables the reliability evaluation of fuzzy ranking of variations by defined, most credible values of indices, characterizing these variations, increasing the output validity as a result. In addition, application of fuzzy set means allows us to study stability of the model endogenous indices with respect to variation of its exogenous indices. It leads to the possibility of the quantitative evaluation of consequences of higher or lower volatility of various input variables for the input parameter stability. The main disadvantages of this method are the absence of principles of selection of a proper membership function specified by the fuzzy set theory, biased formulation of an initial set of rules for fuzzy input etc. [11–12]. However, the aforesaid approach disadvantages are collateral during processing of fuzzy data and do not reduce the relevance of this

approach, which is confirmed by a high scientific appeal to study such fuzzy systems.

III. Fuzzy Method of Revealing and Assessing Implicit Causality

The basic idea of applying a fuzzy approach consists in the fact that a certain index is assumed to be interval and to be determined (fuzzified) by a span instead of a number. This point aims to reflect an actual situation of more or less exactly specified threshold values, within which a parameter can vary. As a consequence, we need to formalize the understanding of the concerned index probable values along with indication of the set of probable values and degrees of uncertainty of their adoption. Then, after calculation of the probability distribution of the overall index, we should pass on defuzzification and interpretation stages based on the system of rules and using developed output tools.

The model-building task is to find and to quantify causal relations between implicit factors which are the control system objects, on directly related parameters and through them on certain key indices. The important aspect is finding the very implicit factors. The basic model-building tools are fuzzy binary relations, their composition, and data mining algorithms. The model-building logic is illustrated below.

IV. Logic behind a Fuzzy Model for Discovering Implicit Causality within a Control System

Suppose we are given a set of $A=\{a_1,a_2,\ldots,a_n\}$. With the specified degree of probability we can find two elements among the elements of the set, which have a relatively small mutual effect, but there is an element distinctive from the above two elements, with introduction of which the effect becomes significant.

Some definitions are introduced below. Suppose U is any set, U^2 is Cartesian square of this set ($U^2=U\times U=\{(a;b):a,b\in U\}$). The fuzzy binary relation on the U set is the U^2 fuzzy subset. Conventional notation format of the fuzzy binary relation for the discrete and continuous sets are represented in equations 1 and 2, respectively.

$$\Gamma=\frac{\sum_{U^2}\mu_\Gamma(u_i,u_j)}{(u_i,u_j)}, \tag{1}$$

$$\Gamma=\int_{U^2}\frac{\mu_\Gamma(x,y)}{(x,y)}. \tag{2}$$

Matrices whose elements are values of the membership function of $\mu_\Gamma(x,y)$ fuzzy binary relation are denoted as J_Γ.

The composition of Γ_1 and Γ_2 fussy binary relations is specified by such a fuzzy binary relation of $\Gamma=\Gamma_1\circ\Gamma_2$, that the equation 3 is valid:

$$\frac{\mu_{\Gamma_1\circ\Gamma_2}(x,y)}{(x,y)}=\\ =\bigcup_{z\in U}\left(\left(\mu_{\Gamma_1}(x,z)/(x,z)\right)\bigcap(\mu_{\Gamma_2}(z,y)/(z,y)\right). \tag{3}$$

Given that the intersection of $\mu_{\Gamma_1}(x,z)/(x,z)$ and $\mu_{\Gamma_2}(z,y)/(z,y)$ single-point fuzzy sets is generally performed by the logical T –norm, and its union is performed by the logical T –conorm: $a\bigcap b=\min(a,b), a\bigcup b=\max(a,b)$, equation 3 takes the form of the equation 4.

$$\frac{\mu_{\Gamma_1\circ\Gamma_2}(x,y)}{(x,y)}=\\ =\frac{\max\limits_{z\in U}\left(\min(\mu_{\Gamma_1}(x,z),\mu_{\Gamma_2}(z,y)\right)}{(x,y)}. \tag{4}$$

Equations 5 and 6 are the relation composition graph for discrete and continuous sets, respectively.

$$\Gamma_1\circ\Gamma_2=\sum_{U^2}\frac{\mu_{\Gamma_1\circ\Gamma_2}(x,y)}{x,y}=\\ =\sum_{U^2}(\max_{z\in U}(\min\mu_{\Gamma_1}(x,z),\mu_{\Gamma_2}(z,y)))/(x,y), \tag{5}$$

if U is a finite set;

$$\Gamma_1\circ\Gamma_2=\int_{U^2}\mu_{\Gamma_1\circ\Gamma_2}(x,y)/(x,y)=\\ =\int_{U^2}\frac{\max\limits_{z\in U}\left(\min\mu_{\Gamma_1}(x,z),\mu_{\Gamma_2}(z,y)\right)}{(x,y)}, \tag{6}$$

if U is part of the number axis or the entire number axis.

Therefore, from equation 4 for a U finite set we can obtain:

$$J_{\Gamma_1\circ\Gamma_2}=J_{\Gamma_1}\circ J_{\Gamma_2}=\\ =(\max_k(\min\left(\mu_{\Gamma_1}(u_i,u_k),\mu_{\Gamma_2}(u_k,u_j)\right)_{nxn}=\\ =\left(\mu_{\Gamma_1\circ\Gamma_2}(u_i,u_j)\right)_{nxn},$$

where n is a number of U set elements.

Build-up the J_Γ matrix for the A set.

$$J_\Gamma = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22} & \dots & s_{2m} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nm} \end{pmatrix},$$

where s_{ij} ($0 \le s_{ij} \le 1; i = 1,2,\dots,n; j = 1,2,\dots,m$) is an extent of the a_i index effect on the a_j index.

Then, to find mediate effects we can calculate values of J_{Γ^2} matrix using the equation 4:

$$J_\Gamma = J_\Gamma \cdot J_\Gamma = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \dots & \dots & \dots & \dots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{pmatrix}.$$

The reflexive selection of mediated factors [13] performed as part of data mining assumes existence of such a pair of s_{ij} and f_{ij}, that $s_{ij} << f_{ij}$, and indicates the presence of an implicit influence, which becomes apparent due to an intermediate factor within the system [3].

V. Model Adaptation to Handling Fuzzy Binary Correspondence

When necessary, the model may be adapted to work with fuzzy binary correspondences. The conceptual similarities and differences of this adaptation are described below.

Binary correspondence in the case of $A \times B$ set means a Γ subset of the Cartesian product of sets A and B: $\Gamma \subseteq A \times B$. The Cartesian product $A \times B$ describes the product of all the sets that feature A set element in the first position, and B set element in the second position. In the case when $A = B$, binary correspondences amount to the mode of binary relations $\Gamma \subseteq A^2$.

Compositions of binary correspondences and binary relations are defined similarly. Compositions of fuzzy binary correspondences $\Gamma_1 \subseteq A \times B$ and $\Gamma_2 \subseteq B \times C$ are defined by the equation 7, provided that equation 8 is correct for the membership function.

$$\Gamma = \Gamma_1 \circ \Gamma_2 \subseteq A \times C, \tag{7}$$

$$\begin{gathered} \mu_{\Gamma_1 \circ \Gamma_2}(x,y)/(x,y) = \\ = \bigcup_{z \in B} \left(\mu_{\Gamma_1}(x,z)/(x,z) \cap \mu_{\Gamma_2}(z,y)/(z,y) \right) \\ (x \in A, y \in C). \end{gathered} \tag{8}$$

The intersection and the union of single-point fuzzy sets in the case of binary correspondences are made according to the abovementioned rules. In this case, the equation 8 goes over to the equation 9:

$$\begin{gathered} \mu_{\Gamma_1 \circ \Gamma_2}(x,y)/(x,y) = \\ = \left(\max_{z \in B} \left(\min(\mu_{\Gamma_1}(x,z), \mu_{\Gamma_2}(z,y)) \right) \right)/(x,y) \\ (x \in A, y \in C). \end{gathered} \tag{9}$$

The graph of finite set correspondences' composition follows the equation 10; the equation 11 defines the graph for the sets that represent an interval of the number axis or the entire number axis.

$$\begin{gathered} \Gamma_1 \circ \Gamma_2 = \sum_{A \times C} \mu_{\Gamma_1 \circ \Gamma_2}(x,y)/(x,y) = \\ = \sum_{A \times C} \left(\max_{z \in B} \left(\min(\mu_{\Gamma_1}(x,z), \mu_{\Gamma_2}(z,y)) \right) \right)/(x,y), \end{gathered} \tag{10}$$

where A, B, and C are finite sets;

$$\begin{gathered} \Gamma_1 \circ \Gamma_2 = \int_{A \times C} \mu_{\Gamma_1 \circ \Gamma_2}(x,y)/(x,y) = \\ = \int_{A \times C} \max_{z \in B} (\min(\mu_{\Gamma_1}(x,z), \mu_{\Gamma_2}(z,y)))/(x,y), \end{gathered} \tag{11}$$

if A, B, and C are an interval of the number axis or the entire number axis.

From the equation 10 it follows that the matrix of the composition of $J_{\Gamma_1 \circ \Gamma_2}$ relations, when A, B, and C are finite sets, is nothing but a maximin matrix product of J_{Γ_1} and J_{Γ_2}:

$$\begin{gathered} J_{\Gamma_1 \circ \Gamma_2} = J_{\Gamma_1} \cdot J_{\Gamma_2} = \\ = \left(\max_{k=1,2,\dots,p} (\min(\mu_{\Gamma_1}(x_i,z_k), \mu_{\Gamma_2}(z_k,y_j)) \right)_{m \times n} = \\ = \left(\mu_{\Gamma_1 \circ \Gamma_2}(x_i,y_j) \right)_{m \times n}, \end{gathered}$$

where p is the amount of B set elements; m is the amount of A set elements; n is the amount of C set elements.

Building a model for evaluation of implicit causality in the control system applying fuzzy binary correspondences falls into two steps:

- Building sub-models A including the set of implicit factors, B incorporating the set of indirect indices, and C involving the set of key parameters.
- Integration of sub-models into a general model, its analysis and solution of the problem set.

The operation sequence of the first step may include the following procedures:

- Primary specification of a number index set for each sub-model.
- Making lists of number index sets.

The operation sequence of the second step includes the following:

- Evaluation of interdependence between the indices in pairs: (A,B), (A,C), (B,C).

- Detecting of indirect effects that indices of the sub-model A have on indices of the sub-model C.
- Explanation of the obtained results.

Dependences are defined using J_{AB}, J_{BC} and J_{AC} matrices for the set of A, B, and C indices:

$$A = \{a_1, a_2, ..., a_n\},$$
$$B = \{b_1, b_2, ..., b_m\},$$
$$C = \{c_1, c_2, ..., c_k\},$$

$$J_{AB} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22} & \dots & s_{2m} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nm} \end{pmatrix},$$

$$J_{AC} = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1k} \\ z_{21} & z_{22} & \dots & z_{2k} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nk} \end{pmatrix},$$

$$J_{BC} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1k} \\ u_{21} & u_{22} & \dots & u_{2k} \\ \dots & \dots & \dots & \dots \\ u_{m1} & u_{m2} & \dots & u_{mk} \end{pmatrix},$$

where $s_{ij}\ (0 \le s_{ij} \le 1; \quad i = 1,2,...,n; \quad j = 1,2,...,m)$ is an extent of the a_i index effect on the b_j index, $z_{ij}\ (0 \le z_{ij} \le 1; \quad i = 1,2,...,n; \quad j = 1,2,...,k)$ is an extent of the a_i index effect on the c_j index, $u_{ij}\ (0 \le u_{ij} \le 1; \quad i = 1,2,...,m; \quad j = 1,2,...,k)$ is an extent of the b_i index effect on the c_j index.

The extent of the direct effect a_i makes on c_1 is determined by the z_{i1} element of the J_{AC} matrix. Similarly, the extent of the direct effect a_i makes on $c_2,...,c_k$ is determined by $z_{i2},...,z_{ik}$ numbers. In addition to the direct effect, the a_i index affects $c_1, c_2, ..., c_k$ through the intermediate factor b_j, which is an index of the sub-model B. The extent of the indirect effect a_i has on $c_1, c_2, ..., c_k$ through b_j are taken as values $z_{i1}^*, z_{i2}^*, ..., z_{ik}^*$, which are minimums of s_{ij} and corresponding $u_{j1}, u_{j2}, ..., u_{jk}$:

$$z_{i1}^* = \min(s_{ij}, u_{j1}), z_{i2}^* = \\ = \min(s_{ij}, u_{j2}), ..., z_{ik}^* = \min(s_{ij}, u_{jk}).$$

The equation 12 specifies combined indirect effect that a_i element produces on c_j:

$$z_{ij}^* = \\ = \max(\min(s_{i1}, u_{1j}), \min(s_{i2}, u_{2j}), ..., \min(s_{im}, u_{mj})). \quad (12)$$

The matrix J_{AB} and J_{BC} product (equation 13) specifies the indirect effect A set elements cause on C set elements through B:

$$J_{AC}^* = J_{AB} \cdot J_{BC} = \begin{pmatrix} z_{11}^* & z_{12}^* & \dots & z_{1k}^* \\ z_{21}^* & z_{22}^* & \dots & z_{2k}^* \\ \dots & \dots & \dots & \dots \\ z_{n1}^* & z_{n2}^* & \dots & z_{nk}^* \end{pmatrix}, \quad (13)$$

where z_{ij}^* is calculated using the equation 12.

If the extent of direct effect A makes on C, determined by following the steps of hierarchy analysis, exceeds the indirect effect, then it is not worth being taken into account. If the inequality $z_{ij}^* - z_{ij} > 0$ is valid, then we showed the indirect and previously ignored effect that i-th implicit factor makes on j-th resulting index. Moreover, evaluation of the extent of such effect may be considered as $z_{ij}^* - z_{ij}$ difference.

Consequently, using compositions of binary correspondences implicit, indirect relations and cross-effects between the elements of A and C sets may be found out, when correspondences for $A \times B$ and $B \times C$ sets are given; intermediate factors may be specified as well.

VI. Model Application and Testing

The algorithm mentioned above is automated and presented as a web-service developed on Joomla! 1.5 platform, written in PHP 3.1 language, and posted on the website http://bi.usue.ru/nauka/imin/. The most recent version 1.0.3 allows to insert the data in the natural language with the help of "fuzzy controller", to check the correctness of the input values, as well as to get the result indicating the implicit factors and the causal effect strength.

Critical capabilities of the model when operated on a certain control system ensure its applicability for the analysis of a wide range of data. The model using Goguen's fuzzy implication [14] and defuzzification according to the method of the centre of gravity [15–16] was officially approved in the research of the implicit causality within the corporate culture [17], the military-industrial complex of the Russian Federation [18], and demonstrated a high reliability (measure of inaccuracy of significant subtle effect finding was below 3% regardless of the subject of research). Moreover, as one of the most promising methods of finding and evaluation of implicit influences, the aforesaid algorithm in the course of implementation of the system of machine learning may be

applied for the purpose of handling the problems of adaptable control with application of artificial neural networks [12,18].

VII. Conclusion

Thus, the problem of revealing implicit causality that exists among the control system objects, under toughening requirements to description of the current and estimated condition of the systems, remains topical. The system simulation based on the study of the cause and effect linkage between its objects represents one of the auspicious research methods to overcome control problems with the aim of ensuring the required performance. Particular attention in the course of designing a model under the state of uncertainty should be paid to fuzzy techniques offering a number of critical benefits compared to conventional methodological approaches that ensure a widespread application of a fuzzy method due to its flexibility without diminishing the reliability of the acquired results.

The fuzzy model described in detail in the paper is intended to work with a wide range of control systems and makes it feasible, based on the composition of fuzzy binary relations, fuzzy binary correspondences, and data mining, to spot the implicit causal effects within. The above algorithm gives the opportunity to detect the unobvious elements of a control system, as well as to calculate their effects. The obtained results contribute towards enhancing the appropriateness of the research subject behavior assessment.

References

[1] Marz N., Warren J. Big Data: Principles and best practices of scalable realtime data systems. Greenwich, CT: Manning Publications Co., 2015. 455 pp.

[2] Baars H. et al. Shaping the Next Incarnation of Business Intelligence //Business & Information Systems Engineering. 2014. Vol. 6. N 1. Pp. 11-16.

[3] Nazarov D. M. The fundamental basis of implied in the system of economic development organizations // Izvestia of St. Petersburg State University of Economics. 2015. N 3 (93). Pp. 7-14.

[4] Granger C. W. J. Causality, cointegration, and control //Journal of Economic Dynamics and Control. 1988. Vol. 12. N 2. Pp. 551-559.

[5] Bellman R. E. Adaptive control processes: a guided tour. Princeton university press, 2015. 276 pp.

[6] Shpolyanskaya I.Yu. Information systems consumer appeal for small businesses // Applied Informatics. 2006. N 2. Pp.58-65.

[7] Kharitonov V. A., Alekseyev A. O. The conception of causality at the problem of socio-economic systems control //Management of economic systems: Scientific electronic journal. 2012. N.46 (10). P. 7.

[8] Malyshenko A.M. Dynamic systems causality indices and their use in schematic design and in functional reproducibility assessment of automatic control systems // Bulletin of the Tomsk Polytechnic University. 2013. N 5. Pp. 37-44.

[9] Cybenko G. Approximation by superpositions of a sigmoidal function //Mathematics of control, signals and systems. 1989. Vol. 2. N 4. Pp. 303-314.

[10] Kosko B. Fuzzy systems as universal approximators //Computers, IEEE Transactions on Computors. 1994. Vol. 43. N 11. Pp. 1329-1333.

[11] De Reus N. M. Assessment of benefits and drawbacks of using fuzzy logic, especially in fire control systems. TNO-report N FEL-93-A158. The Hague: TNO physics and economics laboratory, 1994. 39 pp.

[12] Kruglov V. V., Dli M. I., Golunov R. Yu. Fuzzy logic and artificial neural networks: textbook. Moscow: Publishing house of physical and mathematical literature, 2001. 224 pp.

[13] Lefebvre V. A., Smolyan G. L. Algebra of Conflict. Moscow: Znanie, 1968. 64 pp.

[14] Goguen J. A. The logic of inexact concepts //Synthese. 1969. Vol. 19. N 3. Pp. 325-373.

[15] Mamdani E. H. Applications of fuzzy algorithms for simple dynamic plants // IEE. 1974. Vol. 121. Pp. 1585-1588.

[16] Mamdani E. H., Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller //International journal of man-machine studies. 1975. Vol. 7. N 1. Pp. 1-13.

[17] Nazarov D. M. The model for assessing implicit factor influence on organization performance indicators // BI technologies in optimization of business processes: International workshop (Yekaterinburg, 2 December 2015). Yekaterinburg: Ural State University of Economics Press, 2015. Pp. 29-34.

[18] Azarov D. A. Revisiting modeling Russia's military-industrial complex economic influence on national economy at the present stage // BI technologies in optimization of business processes: International workshop (Yekaterinburg, 2 December 2015). Yekaterinburg: Ural State University of Economics Press, 2015. Pp. 35-38.

Opinion Mining and Sentiment Analysis for Contextual online-Advertisement

Assoc.Prof. Abzetdin Z. Adamov
Applied Research Center for Data Analytics and Web Insights (CeDAWI), Qafqaz University
Baku, Azerbaijan
aadamov@qu.edu.az

Prof. Esref Adalı
Faculty of Computer Eng. and Informatics
Istanbul Technical University
Istanbul, Turkey
adali@itu.edu.tr

Abstract – With rapid expansion of the Internet and increasing amount of time users spend online, the Internet evolves from entertainment environment towards highly dynamic and flexible business medium. Online advertisement has become one of the most successful business model for Internet environment. There are two major types of online advertisement: sponsored search and contextual display advertisement. This paper dedicated on contextual display advertisement. Generally, contextual advertisement implementations based on topical or keyword-based relevance approach. This study addresses the mechanism of advanced contextual advertisement based on opinion about specific topic within content of webpage. Use of Natural Language Processing and Sentiment Analysis aims to determine the writer's attitude towards particular topic as: positive, negative, or neutral. This approach helps to develop an advertisement system that is more content-sensitive and consequently has higher ROI of marketing.

Keywords – Opinion mining, Sentiment analysis, NLP, Contextual advertisement

I. Introduction

Rapid growth of Internet economy has a great impact to transformation of advertisement industry. According to eMarketer, the global advertisement market value will increase 6% over 2014 to $592 billion in 2015. TV which has been unaltered leading media of advertising industry for many years, tends to lose its value and impact in front of aggressive expansion of the Internet Advertisement. As the Internet Advertisement continues a rapid growth, it is expected to become largest advertising category by 2019 [1].

Modern digital advertisement systems evolved from static banner and pop-ups into sophisticated combination of rich media, video, interactive applications and banners. Complex application (script) embedded into web page analyze hundreds of parameters including website's URL and ranking, visitor's geo-location, visitor's default language, browser software, Internet speed, content of web page, etc. as a result chooses one specific banner of thousands available in ads repository to display in the banner area.

Even with such an advancements in the area of online advertisement, the effectiveness and ROI (Return of Investment) is still a challenging problem that attract researchers over the past few years in addressing questions such as: How to define relevance between visitor's interest and specific advertisement? How to build correlation between opinion in textual content of web page and specific advertisement? How to prevent visitors blocking ads?

Sentiment analysis is using machine learning algorithms and natural language processing techniques in order to identify whether a given phrase or sentence express positive, negative, or neutral meaning [2, 3].

Providing an efficient way to analyze user's sentiment has become major concern for both academic research and commercial companies. Machine learning approaches such as Opinion Mining or Sentiment Analysis can largely answer these questions. Sentiment Analysis or Sentiment Classification is a set of Machine Learning Algorithms and Natural Language Processing techniques used to identify whether a given phrase or sentence express positive, negative, or neutral meaning, thus consequently analyze people's opinions, attitudes, emotion, filings toward entities, individuals, events, products, and their attributes [4].

When it comes to online advertisement, Sentiment Analysis is aimed to identify the attitude of author of a specific website content to the particular topic. After this matching advertisement system can chose specific advancements campaign that is most relevant to the main topic of web page and sentiment (positive or negative) of the author.

For example, it will not be appropriate to put mobile phone advertisement on top of the web page with content devoted to the health risks of mobile phone radiation. At the same time, it may be appropriate to advertise the mobile phone that have specific attribute of low radiation level (SAR - Specific Absorption Rate).

II. Architecture for Scalable Contextual ads System

A.Sentiment Analysis

As it was mentioned earlier, keyword-based contextual approach in online advertisement is not enough efficient, because it doesn't take into account the sentiment or writer's opinion to the topic of written text. It would not be

appropriate to put into web-page ad of product or service about which author promotes negative opinion. To eliminate this lack of correspondence between content and advertisement, the sentiment analysis and classification was used.

There are many approaches and algorithms for sentiment classification, but generally all of them serve the same abstract purpose. "Sentiment analysis traditionally emphasizes on classification of web comments into positive, neutral, and negative categories". [5]

There are two main approaches in sentiment classification regarding specifying the scope of sentiment: binary sentiment and multi-class sentiment. In case of binary sentiment classification, positive and negative terms were identified by querying a collection of appropriate dictionaries and sentence datasets. The second method is more complex approach and emulates real human emotions, which has more complicated and diverse spectrum than just positive and negative. In this paper the first approach has been used. [6, 7]

B. Topic Extractor and Ad Campaign Matcher

After extracting sentiment bearing sentences, the next step is extracting all noun phrases or topical words those are related with sentiment words. Since, most of the advertisement campaigns specify topical or advertising keywords very precisely, the intersections of topical words between ads and extracted nouns are very low. So, considering the advertising keywords from advertisement campaign and selected topical words from web-page is not enough for effective topical words and ad keywords matching. In order to solve this problem, it is proposed to extend topical words using several synonyms for each of noun. POS (Part-of-Speech) tagging is used to mark all nouns, which are considered as candidate topical words. [8]

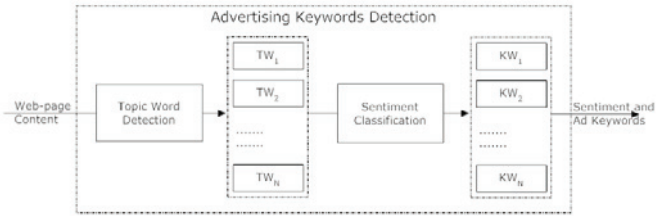


Figure 1. Data Analytics Engine Architecture

The algorithm of determining the advertising keywords is shown in Figure 1., that is based on approach proposed in [7]. Even in proposed approach the first step is about extracting topical words, and second about sentiment classification, they could be swapped. In this case we can eliminate the number topical words to deal with, dealing with just those located within sentiment bearing sentences.

C. Architecture

Two architectures are proposed for Contextual Online-Advertisement System: Stand-alone (basic) Architecture and Scalable Repository-based Architecture.

Stand-alone (basic) Architecture works according to the following instructions (see Figure 2.):

1. A Web-page with banner, embedded Ads-script and some HTML content is loaded into a browser
2. The Ads-script communicates with remote Online-Ad-System requesting most relevant ad for specific URL
3. Crawl the content of the web-page from specified URL
4. Parse web-page and apply pre-processing functions
5. Determine Topical Words and Sentiment polarity on these words using Text Classification approach
6. Identify most relevant Ad-campaign based on Topical Words and Sentiment
7. Send selected ad to web-page

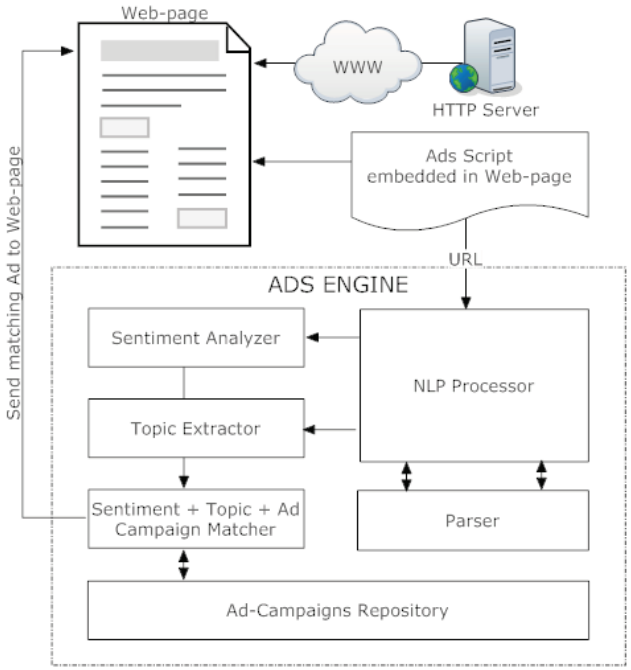


Figure 2. Stand-alone (basic) Architecture of the Contextual Online-Advertisement System

Let see now the Scalable Repository-based Architecture and how it operates. The most significant difference between two architectures is that, second architecture uses web-pages' repository, which consists of the web-pages those already crawled and stored in Hadoop-HDFS platform (more details are available at [9]). This helps to increase the efficiency of the system through one of most time-consuming crawling function. To do so, the system checks if the web-page with specific URL already exists in repository and calls the crawling function just in case if not. As it can be seen from Figure 3., the first two steps are same as was described for Stand-alone (basic) Architecture:

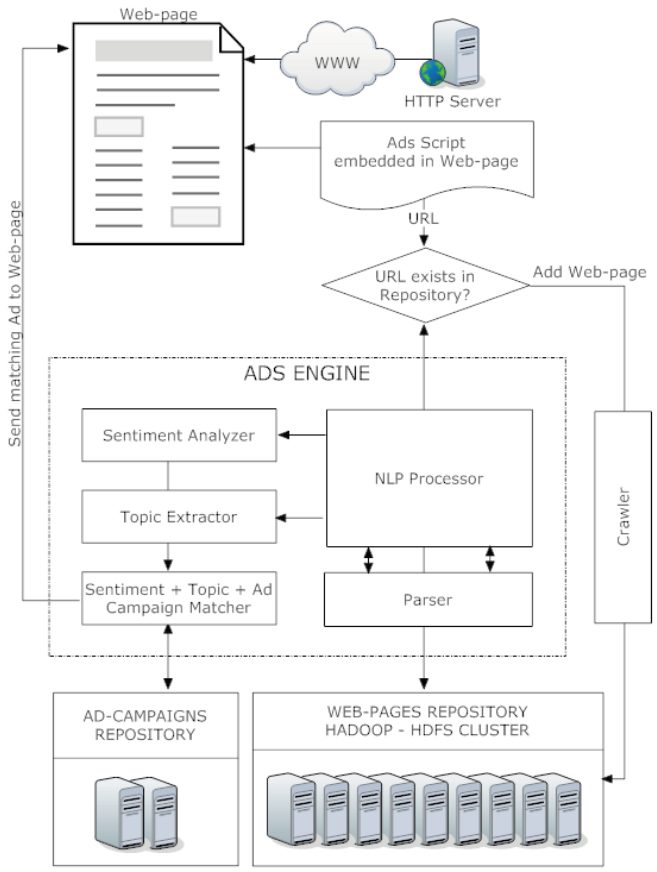


Figure 3. Scalable Repository-based Architecture of the Contextual online-Advertisement System

1. A Web-page with banner, embedded Ads-script and some HTML content is loaded into a browser
2. The Ads-script communicates with remote Online-Ad-System requesting most relevant ad for specific URL
3. Checks if the web-page with specific URL already exists in repository
   a) If not exist, crawl URL and store into a Web-repository
   b) If exists, retrieve web-page content from a Web-repository
4. Parse web-page and apply pre-processing functions
5. Determine Topical Words and Sentiment polarity on these words using Text Classification approach
6. Identify most relevant Ad-campaign based on Topical Words and Sentiment
7. Send selected ad to web-page

CONCLUSION

One of the most important findings to emerge from this study is that, the contextual advertisement based on sentiment analysis gives much better results than the keyword-based approach. This is true not just for ads, but for any other area where an opinion mining is important, including risk management, topic or event extraction, predictive analytics, etc. Also, two architectures are proposed for contextual online-advertisement system: Stand-alone (basic) Architecture and Scalable Repository-based Architecture.

In order to achieve higher accuracy rate on the machine learning algorithms, further work needs to be done extending positive and negative word dictionaries as well as positive and negative sentence datasets.

REFERENCES

[1] eMarketer, "Advertisers Will Spend Nearly $600 Billion Worldwide in 2015", December 10, 2014, URL:http://emarketer.com/

[2] Jeonghee Yi; Nasukawa, T.; Bunescu, R.; Niblack, W., "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques," in Data Mining, 2003. ICDM 2003. Third IEEE International Conference on , vol., no., pp.427-434, 19-22 Nov. 2003

[3] Cambria, E.; Schuller, B.; Yunqing Xia; Havasi, C., "New Avenues in Opinion Mining and Sentiment Analysis," in Intelligent Systems, IEEE , vol.28, no.2, pp.15-21, March-April 2013

[4] Bing Liu, “Web Data mining. Exploring Hyperlinks”, Content, and Usage Data, Springer, 2011

[5] Keke Cai; Spangler, S.; Ying Chen; Li Zhang, "Leveraging Sentiment Analysis for Topic Detection," in Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on, vol.1, no., pp.265-271, 9-12 Dec. 2008

[6] Alistair Kennedy and Diana Inkpen, "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. In Computational Intelligence", 2006

[7] Andr´es Cassinelli, Chih-Wei Chen, "Boost up! Sentiment Categorization with Machine Learning Techniques", June 5 2009

[8] Guang Qiu, Xiaofei He, Feng Zhang, “DASA: Dissatisfaction-oriented Advertising based on Sentiment Analysis”, Journal of Expert Systems with Applications, 2010

[9] Abzetdin Z. Adamov, "Text Analysis Case Study: Determining Word Frequency based on Azerbaijan top 500 websites", in Application of Information and Communication Technologies (AICT), 2015 9th International Conference, 14-16 Oct. 2015

Database Schema Method for Automatic Semantic Errors Resolving During Information Systems Integration

Andrey V. Grigoryev, Alexander A. Kropotin, Alexander G. Ivashko
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
a.v.grigorev@utmn.ru, a.a.kropotin@utmn.ru, a.g.ivashko@utmn.ru

***Abstract*—at present time there are many enterprises that prefer to integrate out-of-the-box systems (with integration application) instead of developing new enterprise system special for them. Integration applications provide functionality to synchronize data between databases of out-of-the-box systems. However, there are problems during integration and we propose a method for these problems automatically solving. For example, one of the problem is to determine same concepts and attributes in the integrated databases. In this article, we present a method based on the semantic web approaches: OWL and its mathematical base - Description Logics. We present mathematical model of the database as set of the OWL concepts, roles and axioms, which determines rules for database integration consistency. Each concrete database has its own mathematical representation and transformation rules are presented following in the article. Set of such rules is presented as set of functions from database elements to ontology elements.**

***Index Terms*— Database, schema, ER-model, OWL, ontology, description logics.**

I. INTRODUCTION

A. Necessity of Data Integration

At the present, in the information technology world there is a tendency of transition from new information systems creating to software integration. Integration is a synchronization of data between software databases, i.e. interchanging records between tables presenting same concepts. Many scientific schools around the world are taking part in researches related to information systems integration: Department of Informatics – Database Technology, University of Zurich [1], Wright State University [2], Dublin City University School of Computing [3]. In addition, many conferences on data integration are conducted every year: Smart Systems Integration, International Conference on Enterprise Information Systems, International Conference on Data Integration in the Life Sciences, etc. All these facts prove actuality of the research presented in this article.

There are issues and challenges in database integration problem. In this article we present mathematical model of relational database metamodel based on description logics (DL) formalism. The model with DL reasoning algorithm (called tableau algorithm) will solve problem related to checking mergability of databases and determining sameness of tables (and attributes) in databases.

Main issue is to check whether databases can be integrated.

B. Database integration issues

There are some problems during database integration, which cannot be solved without human. It is presented in different sources by different authors. We emphasize main of them:

1. First of all (semantic integration is impossible without it) is the problem of finding matches between concepts in the databases (synonymy determining). To synchronize data automatically it is needed to find out same concepts in the databases. Human can solve this problem manually, but there are some automatically approaches [4, 5, 6]. We consider developing an alternative method in further researches.

2. Equivalence among constructs of the model. In conceptual models, several combinations of constructs can model the same application domain equivalently. Consequently, "richer" models give rise to a larger variety of possibilities to model the same situation [7, 8].

3. Incompatible design specifications. Erroneous choices regarding names, types, integrity constraints, etc. may result in erroneous inputs to the schema integration process [7, 8].

4. Common concepts. Owing to the causes for schema diversity described above, it may very well happen that the same concept of the application domain can be represented by different representations in different schemas and several types of semantic relationships can exist between

such representations. They may be identical, equivalent, compatible, or incompatible [7, 8].

5. Concepts related by some semantic property. Regarding the concepts in component schemas that are not the same but are related we need to discover all the interschema properties that relate them [7, 8].

Main interest is the second task. Example of this issue is presented on the figures 1a and 1b. At the figure 1a schema for employers-organization relationships is presented. At the figure 1b another schema for same domain is presented. Schemas are different because they model categorization of the employer-driver-manager relationships. At schema 1a it is modelled as one concept, at schema 1b it is modelled as three different concepts.

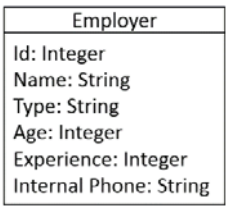


Fig 1a. Staff Model As One Concept

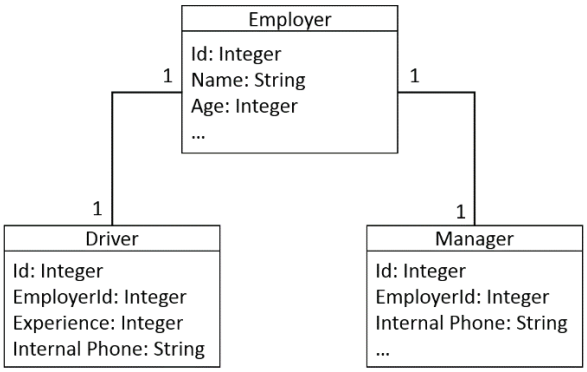


Fig 1b. Staff Model As Three Concept

Presented task can be solved automatically using intelligence approach based on semantic web methods. In the next section, we present mathematical basis of semantic web methods called description logics. We use the most expressive logics called DL SROIQ and OWL 2 [10] language for describing it in computer memory.

Our approach based on database schema mathematical model creating and apply reasoning algorithm that answers on question if two database schemas can be integrated. Answer will be provided as "Yes/No" for two databases.

Rest part of the article has next structure. In the section 2 we represent description logics SROIQ formalism used as base of the mathematical model. Section 3 contains description of the developed mathematical model. Consistency of the presented mathematical model is presented in section 4. Further researches plan is presented in the section 5.

II. Preliminaries

A. Ontologies and OWL

In this paper, we use artificial intelligence methods based on semantic web methods to check inconsistences during software integration. Ontologies and ontology description language are widely used for solving medicine and biology tasks: many knowledge bases included in the medicine decisions making support systems were developed. These systems allows to advice on the treatment appointment for patients. In the biology, there are many knowledge bases, which allow classifying living creatures by some dedicated feature.

The main reason of ontologies wide usage is existence of a method that can automatically check if a knowledge base is inconsistence, i.e. it can check if there are some facts, which possibly can explicitly exists in an ontology and contrary to each other. Existence of such method allows checking inconsistence of some given fact in correspondence of the ontology. This in its turn allows performing logical reasoning in ontology to check subsumption of some concept to another and perform classification of concepts.

OWL allow creating description of subject areas as knowledge bases named ontologies. Ontology is a set of subject area concepts and subject area relation between these concepts. In this section, we briefly describe mathematical base of the OWL named description logics. OWL is a just XML syntax for description logics axioms so avoid cluttering with hardly understandable XML tags we will use description logics syntax.

B. Description Logics

Description logics is a powerful instrument for knowledge bases description, but it is not as powerful as first order logics. Nevertheless, in the our work we use description logics because of it is a decidable fragment of the first order logics and this fact allows to develop such knowledge bases, that allows to perform reasoning, i.e. it allows to determine implicitly defined relations between concepts in the subject area which are logical consequences of the given axioms.

In the description logics the concept is a class of individuals (objects of subject area), which satisfy to certain property in description logics. For instance, "Human" concept defines a set of all objects that are humans. Concepts are not only names of certain properties. In the table 1 are presented all constructors, which define an individual's classes of SROIQ description logics (mathematical basis of the OWL) of the web ontology language (OWL).

TABLE I. SROIQ Description Logics Concepts Constructors

Name	Syntax	Description
Conjunction	$C_1 \sqcap C_2$	Defines individuals having the properties C_1 and C_2
Disjunction	$C_1 \sqcup C_2$	Defines individuals having the

Name	Syntax	Description
		properties C_1 or C_2
Existence quantifier	∃R.C	Defines individuals having the relation *R* with the one or more another individuals, which have the property *C*
Universal quantifier	∀R.C	Defines individuals having the relation *R* with the only those individuals, which have the property *C*
<= cardinality constructor	n R <= C	Defines individuals having the relations *R* with more or equal than *n* individuals, which have a property *C*
>= cardinality constructor	n R >= C	Defines individuals having the relations *R* with less or equal than *n* individuals, which have a property *C*

Knowledge base is defined by sets of equivalence axioms (**C≡D**) and subsumption axioms (**C⊑D**). These axioms are described in TBox axioms set containing general rules about subject area. Concrete assertions about individuals of subject area are contained in the ABox axioms set. Such assertions has the form of the class assertions axioms (**i:C**, where "i" is the individual name and "C" is a concept name) or object property assertion axioms (**(a, b): R**, where "**a**" and "**b**" are the individuals names and "**R**" is the object property name).

Important thing in description logics is an interpretation that is a base of a knowledge base semantics. To understand an interpretation consider following example: let there is an axiom **C ⊑ B** mean all objects **C** are **B** also, i.e. in set theory we can to say, if there is some object c with **C** property (or simple **c** in **C**), then object **c** has property **B** (**c** in **B**).

Kernel of the description logics is the tableau method for a model consistency checking of the concept defined by description logics rules. Model of the concept is a couple of individuals set **X**, set of relations between individuals **R** and sets **q(x)** for all **x∈X**, that contains subject area concepts containing individual **x**. If set **q(x)** contains concept ⊥ or pair of the counter liters **C** and **¬C**, or set of relations **R** contains pair **(x, y)** in both two counter relations $\mathbf{r_i}$ and $\neg\mathbf{r_i}$ ($\mathbf{r_i} \in \mathbf{R}$), then there is a clash in model else it is a right model.

Tableau algorithm has an exponential complexity due to a non-deterministic choice on performing satisfiability checking on **C⊔D** or **n R >=** concepts:

- During concept processing ***C⊔D***, a status of interpretation should be saved and one of the concepts ***C*** or ***D*** should be added to the set of concepts ***q(x)***, by the table algorithm. Thus, if on any step there was a contradiction, it should be return to the next remembered status, in which other choice case is possible.
- During concept processing a of ***n R*** >= individual, which belongs to this concept, shouldn't have more than ***n*** of individuals, which are in the relation of ***R*** with the considered individual and if it is some such individuals, that, probably, some couple of individuals represents one individual and for this purpose their combining is required.

To realize such choice, which will allow to find quickly the correct concept, during processing a disjunction or the correct configuration of combining of individuals, during processing n R> = rule expanding, the new method of determination of coherence of conjunction of two concepts was offered in this paper.

In [9] authors introduced a method that can determine satisfiability of two concepts conjunction in certain cases. We propose an extension of such method and use it to determine consistency of merged individuals.

III. METHOD DESCRIPTION

Mathematical model is presented as a function transforming database concepts to ontology concepts. Following we describe these functions.

A. The ontological model of a conceptual objects domain representation

Before introducing developed method of automatic database mergeability checking we provide small example of inconsistence and how it can be solved with reasoning in description logic knowledge bases. Consider following scenario. Let there is an information system worked in some enterprise. It includes an ERP system with database including two tables: Supplier and City. Supplier can supply gifts to an only one city. The small part of this database schema is presented on fig. 2a. The head of the IT department wants to embed a CRM system where Supplier has a City where he works. It is showed on fig 2b. Can these two schemas be integrated?

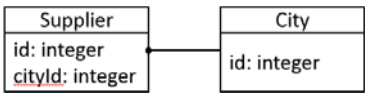


Fig 2a. ERP System Schema

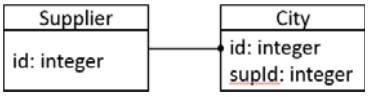


Fig 2b. CRM System Schema

Looking ahead, we will consider every entity of the schema (table in database) as a DL concept, and every attribute as a DL role. So two these schemas will be presented as following ontology:

Supplier ⊑ ⊤
City ⊑ ⊤
Supplier ⊑ ∃cityId.City
City ⊑ ∃supId.Supplier

Reasoning on this knowledge base answers that there are no issues, it is just needed to add both relations into the resulting database. Also it is needed to check A compatibility of the attributes and DL can answer this question too.

B. Formal representation of the method

Mathematical model is presented as a function transforming database concepts into ontology concepts. Following we describe these functions.

Description logics atomic concepts are proposed to put in line with entities sets (database tables) and values sets (data types) of the entity-relationship model as in eq. 1. Then in accordance with description logic rules each atomic concept should be defined in terminological axiom terms of a general concept inclusion in the universal concept.

$$\{A_n \sqsubseteq \top\}_{n=\overline{1,N}}, \tag{1}$$

Where A_n is the atomic concept; N is a count of all tables A_n; ⊤ is the universal concept.

There are no record belonging to different tables in the database because of every two table in the database are disjoint. Define axioms to represent this fact in eq. 2. Corresponding constraints are:

a) each entity set distinct from the others entities sets, in case where an atomic concept are putted in line with an entity set of entity-relationship model;

b) occurrences of the same type values, in case where an atomic concept are putted in line with a value set of the entity-relationship model.

$$\{A_n \sqsubseteq \prod_{n<m\leq N} \neg A_m\}_{n=\overline{1,N}}, \tag{2}$$

Where A_n, A_m are atomic concepts; N is a count of all A_n concepts (symbol ∏ means description logics conjunction of many concepts).

Description logics individuals are proposed to put in line with tables records and concrete values of the data type. In accordance with description logics rules each individual should be defined in assertional axiom terms of a concept assertion as presented in eq. 3.

$$\{\{a_j^n : A_n\}_{j=\overline{1,J^n}}\}_{n=\overline{1,N}}, \tag{3}$$

Where a^n_j is the individual (record) of a concept (table) A_n; J^n is a count of all individuals of the concept A_n; A_n is an atomic concept; N is a count of all concepts A_n.

Further we will use a "concept" as a DL representation of the entity (table) concept, "individual" as a record, role as an attribute.

Because of all records are not the same, individuals difference should be defined as individuals constraints assertions axiom. This axiom is presented in eq. 4. That corresponds to unique constraints of entities sets.

$$\{\{a_j^n \not\approx \{a_k^n\}_{j<k\leq J^n}\}_{j=\overline{1,J^n}}\}_{n=\overline{1,N}}, \tag{4}$$

Where a^n_j, a^n_k are individuals of a concept A_n; J^n is a count of all individuals of a concept A_n; N is a count of all concepts A_n.

In case if all records of the table or data type was identified on the concepts it is needed to define constrains of the belonging individuals in terminological axiom terms of a general concept inclusion as in eq. 5:

$$\{A_n \sqsubseteq \bigsqcup_{1\leq j\leq J^n}\{a_j^n\}\}_{n=\overline{1,N}}, \tag{5}$$

Where A_n is the atomic concept; N is a count of all concepts A_n; a^n_j is the record of the table A_n; J^n is a count of all individuals of the concept A_n.

Every attribute of the database (not a table) is presented as a role in description logics. Atomic roles are proposed in eq. 6:

a) to put in line with relationships sets and attributes of the entity-relationship model;

b) to implement the determining association between atomic concepts and concrete domains values in case where an atomic concepts are putted in line with values sets of the entity-relationship model.

To define attributes it is needed to add a terminological axiom of a general role inclusion with the universal role.

$$\{\{R_i^n \sqsubseteq U\}_{i=\overline{1,I^n}}\}_{n=\overline{1,N}}, \tag{6}$$

Where R^n_i is an atomic role defined on the concept A_n; U is the universal role; I_n is a count of all roles defined on the concept A_n; N is a count of all concepts A_n.

To define attribute metadata (domain and range) to the table it is needed to present a general concept inclusion that define:

a) the role functional property is presented in eq. 7;
b) the role domain is presented in eq. 8;
c) the role range is presented in eq. 9.

$$\{\{\top \sqsubseteq \leq 1 R_i^n\}_{i=\overline{1,I^n}}\}_{n=\overline{1,N}}, \tag{7}$$

$$\{\{\exists R_i^n . \top \sqsubseteq A_n\}_{i=\overline{1,I^n}}\}_{n=\overline{1,N}}, \tag{8}$$

$$\left\{\left\{\begin{matrix}\top \sqsubseteq \forall S_i^n . C^i, & (a)\\ \top \sqsubseteq \forall [T_i^n] . P_{D^i}, & (b)\end{matrix}\right\}_{i=\overline{1,I^n}}\right\}_{n=\overline{1,N}}, \tag{9}$$

Where R^n_i, T^n_i, S^n_i are atomic roles defined on the concept A_n. Besides, T^n_i is a concrete role, S^n_i is an abstract role, R^n_i is an or concrete either abstract role: $R_i^n = T_i^n \vee S_i^n$; I^n is a count of all roles defined on the concept A_n; N is a count of all concepts A_n; A_n is the atomic concept; C^i is the concept; D^i is

the value set; $\boldsymbol{P} \subseteq \{=, <, >, \neq, \geq, \leq\}$ is a predicate set defined on the value set $\boldsymbol{D^I}$.

Primary keys must be presented as attributes with specific DL constraint:

$$\{\{HasKey(R_i^n)\}_{i=\overline{1,I^n}}\}_{n=\overline{1,N}}, \quad (10)$$

Where $\boldsymbol{R^n_i}$ is the atomic role defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{I^n}$ is a count of all roles defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$.

All attributes of the table are presented as terminological axioms of a general concept inclusion with table concept in left of the axiom and conjunction of concepts on the right. Eq. 11a is for the foreign keys and eq. 11b is for not foreign keys attributes:

$$\{\begin{cases} A_n \sqsubseteq \prod_{1\leq i\leq I^n} \exists S_i^n.\top, & (a) \\ A_n \sqsubseteq \prod_{1\leq i\leq I^n} \exists [T_i^n].P_{D^i}, & (b) \end{cases}\}_{n=\overline{1,N}}, \quad (11)$$

Where $\boldsymbol{A_n}$ is the atomic concept; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$; $\boldsymbol{T^n_i}$, $\boldsymbol{S^n_i}$ are atomic roles defined on the concept $\boldsymbol{A_n}$. Besides $\boldsymbol{T^n_i}$ is a concrete role, $\boldsymbol{S^n_i}$ is an abstract role, $\boldsymbol{R^n_i}$ is or concrete role either abstract role: $R_i^n = T_i^n \vee S_i^n$; $\boldsymbol{I^n}$ is a count of all roles defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{A_n}$ is the atomic concept; $\boldsymbol{C^i}$ is the concept; $\boldsymbol{D^i}$ is the value set; $\boldsymbol{P} \subseteq \{=, <, >, \neq, \geq, \leq\}$ is a predicate set defined on the value set $\boldsymbol{D^i}$.

In case if abstract roles are reflexive, i.e. domain and range of role is the same atomic concept that corresponds to unary relationship of the entity-relationship model, in addition to the terminological axioms in eq. 11b terminological axioms of a general concept inclusion eq. 12 should be defined.

$$\{A_n \sqsubseteq \prod_{1\leq i\leq I^n} \forall S_i^n.\neg Self\}_{n=\overline{1,N}}, \quad (12)$$

Where $\boldsymbol{A_n}$ is the atomic concept; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$. Besides, $\boldsymbol{S^n_i}$ is an abstract role defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{I^n}$ is the count of all roles defined on the concept $\boldsymbol{A_n}$, which are putted in line with unary relationships of the entity-relationship model; ***Self*** is the logical symbol, i.e. it have an interpretation which fixed and independent from a knowledge base, indicating the reflection, the concept link with itself.

By assertional axioms (14) of role assertion is proposed to determine:

a) entity tuples of the entity-relationship model. Then in this case it is possible to define existence eq. 14.1.a or absence eq. 14.2.b of relationship between entities;

b) relationship tuples of the entity-relationship model. Then in this case it is possible to define not null or null values of attributes. Axioms for these cases are presented in eq. 14.

$$\{\{\begin{cases} \{(a_j^n, a_k^m): S_i^n, & (1) \\ \{(a_j^n, a_k^m): \neg S_i^n\}_{k=\overline{1,J^m}}, & (2) \end{cases}, \quad (a) \\ \left(a_j^n, =_{v_l^i}\right): [T_i^n], \quad (b) \}_{i=\overline{1,I^n}}\}_{j=\overline{1,J^n}}\}_{n=\overline{1,N}}, \quad (13)$$

Where $\boldsymbol{a^n_j}$ $\boldsymbol{a^m_k}$ are individuals of the concepts $\boldsymbol{A_n}$, $\boldsymbol{A_m}$ respectively; $\boldsymbol{J^n}$ is the count of all individuals of the concept $\boldsymbol{A_n}$; $\boldsymbol{n}$, $\boldsymbol{m}$ are not necessarily distinct; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$; $\boldsymbol{T^n_i}$, $\boldsymbol{S^n_i}$ are atomic roles defined on the concept $\boldsymbol{A_n}$, besides $\boldsymbol{T^n_i}$ is the concrete role, $\boldsymbol{S^n_i}$ is the abstract role; $\boldsymbol{I^n}$ is a count of all roles defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{v^i_l}$ is a value from set $\boldsymbol{D_i}$.

Thus, set of terminological axioms in eq.1 – eq. 2, eq. 5 – eq. 12 and the assertional axioms eq. 3 – eq. 4, eq. 13 described in description logics terms are proposed to term the conceptual domain objects ontology (15).

$$\mathcal{K} = \mathcal{T} \cup \mathcal{A}, \quad (14)$$

Where $\mathcal{K}$ is the conceptual domain objects ontology; $\mathcal{T}$ is a terminological axioms set eq. 1 – eq. 2, eq. 5 – eq. 12; $\mathcal{A}$ is a assertional axioms set eq. 3 – eq. 4, eq. 13.

The extending model is necessary because the open world assumption is implemented in description logics reasoning, i.e. if the some logical assertion is not true it is not means that this assertion is false. Following extensions of a model are proposed:

a) transitive roles extension;
b) atomic concepts classes extension.

Description logics transitive roles (16) are proposed to put in line with relationships sets composition through $\boldsymbol{n \geq 3}$ entities sets of the entity-relationship model. Then in accordance with description logics rules each atomic role should be defined in terminological axiom terms of a general role inclusion in the universal role.

$$R_I \sqsubseteq U, \quad (15)$$

Where $\boldsymbol{R_I}$ is the atomic role; $\boldsymbol{U}$ is the universal role; $\boldsymbol{I}$ is a count of all roles: $I \doteq \sum_{1\leq n\leq N} |I^n| + 1$; $\boldsymbol{I^n}$ is the count of all roles defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$.

To define transitive roles in the knowledge base it is needed to add axioms for:

a) role transitivity property as in eq. 16;
b) define a transitive role as a super role for all other roles as in eq. 17.

$$Tr(R_I), \quad (16)$$

$$\{\{S_i^n \sqsubseteq R_I\}_{i=\overline{1,I^n}}\}_{n=\overline{1,N}}, \quad (17)$$

Where $\boldsymbol{S^n_i}$, $\boldsymbol{R_I}$ is the atomic roles, besides $\boldsymbol{S^n_i}$ is the abstract role, $\boldsymbol{R_I}$ transitive atomic role; $\boldsymbol{I}$ is the count of all roles: $I \doteq \sum_{1\leq n\leq N} |I^n| + 1$; $\boldsymbol{I^n}$ – is a count of all roles defined on the concept $\boldsymbol{A_n}$; $\boldsymbol{N}$ is a count of all concepts $\boldsymbol{A_n}$.

Description logic derivative concepts eq. 19 – eq. 20 are proposed to put in line with such elements types of the entity-relationship model as:

a) entity set of the entity-relationship model (19);
b) value set of the entity-relationship model (20).

$$\left\{ValueSet \equiv \top \sqcap \prod_{1\le i\le I^n} \exists [T_i^n].P_{D^i}\right\}_{n=\overline{1,N}}, \quad (18)$$

$$EntitySet \equiv \top \sqcap \neg ValueSet, \quad (19)$$

Where ***ValueSet***, ***EnitySet*** are derivative concepts which are putted in line with such elements types of the entity-relationship model as entity set and value set respectively; T_i^n is the concrete role; I^n is a count of all roles defined on the concept A_n; D^i is the value set; $P \subseteq \{=, <, >, \neq, \geq, \leq\}$ is a predicate set defined on the value set D^i; N is a count of all concepts A_n which are putted in line with values sets.

The merging ontologies domain conceptual objects (presented in eq. 15) is proposed to term ontology of set of domain conceptual object different representation (presented in eq. 20).

$$\mathcal{K} = \bigcup_{1\le l\le L} \mathcal{K}_l, \quad (20)$$

where $\mathcal{K}$ is the ontology of set of domain conceptual object different representation; L is the count of different ontology union $\mathcal{K}_l$: $L \geq 2$; $\mathcal{K}_l$ is the domain conceptual objects ontology.

Then in this case in integrated terminological axiom set $\mathcal{T} \subseteq \mathcal{K}$ should be defined equal atomic concepts (1) of different ontologies in terminological axiom terms of a concepts equivalence (21).

$$\left\{\left\{\left\{R_i^l \equiv R_i^h\right\}_{i=\overline{1,I^l}}\right\}_{l<h\le L}\right\}_{l=\overline{1,L}}, \quad (21)$$

Where R^l_i, R^h_i are atomic roles defined on the ontology $\mathcal{K}_l$ and $\mathcal{K}_h$ respectively; I^l is a count of all roles defined on the ontology $\mathcal{K}_l$; L is a count of all ontologies $\mathcal{K}_l$.

In addition, in integrated terminological axiom set $\mathcal{T} \subseteq \mathcal{K}$ should be defined equal atomic roles (6) of different ontologies in terminological axiom terms of a concepts equivalence (23), (a) if count of ontology $\mathcal{K}_l$ roles more than count of ontology $\mathcal{K}_h$ or (b) otherwise.

$$\begin{cases} \left\{\left\{\left\{A_n^l \equiv A_n^h\right\}_{n=\overline{1,N^l}}\right\}_{l<h\le L}\right\}_{l=\overline{1,L}} & (a) \\ \left\{\left\{\left\{A_n^l \sqcap \prod_{1\le i\le I} \exists S^l.C^i \equiv A_n^h\right\}_{n=\overline{1,N^l}}\right\}_{l<h\le L}\right\}_{l=\overline{1,L}} & (b)' \end{cases} \quad (22)$$

Where A^l_n, A^h_n are atomic concepts of the ontologies $\mathcal{K}_l$ и $\mathcal{K}_h$ respectively; S^l is the abstract transitive role of the ontology $\mathcal{K}_l$; I is a count of all roles of the concept A^l_n are putted in line with roles of the concept A^h_n which have not equal roles from the ontology $\mathcal{K}_l$; N^l is a count of all concepts of the ontology $\mathcal{K}_l$; L is a count of the all ontologies $\mathcal{K}_l$.

In case of integrated assertional axioms set $\mathcal{A} \subseteq \mathcal{K}$ contains at least two assertional axioms set $\mathcal{A}_l \subseteq \mathcal{A}$: $\mathcal{A}_l \neq \emptyset$, in which defined individuals, in these assertional axioms set should be defined the individuals equivalence of different ontologies $\mathcal{K}_l$ in assertional axioms terms (24).

$$\left\{\left\{\left\{a_j^l \approx a_j^h\right\}_{j=\overline{1,J^l}}\right\}_{l<h\le L}\right\}_{l=\overline{1,L}}, \quad (23)$$

where a^l_j, a^h_j are individuals of the ontologies $\mathcal{K}_l$ и $\mathcal{K}_h$ respectively; J^l is a count of all individuals of the ontology $\mathcal{K}_l$; L is a count of all ontologies $\mathcal{K}_l$.

IV. Evaluation and Discussions

Software implementing presented method was developed. It get database descriptions from definition scripts and transforms it to knowledge bases. Software was developed using Java language, because of there is a library called OWL API that allows creating knowledge bases in OWL language. Besides, there are many open source reasoning software allowing solving task of logical reasoning. To check same names of the tables we use greedy algorithm with similarity checking. In the future researches we consider to use a method based on the Kun algorithm for max weighted matching in bipartite graph.

To the moment, software supports MySQL dialect scripts. HermiT reasoner [11] is using as reasoning tool, because of HermiT shows great results on past OWL Reasoning Evaluation Workshop.

Unfortunately, we test our software only during development on test databases with no more than 10 tables and tests shows 100% regular results.

In the future, we will extend our researches, we plan to develop, and present ready software tested on enterprise databases. Moreover, there are few systems with opened database creating scripts, and to build knowledge bases we will use data description using data definition in web services. For example, WSDL-definition with a XSD schemas can be considered as a source of the database schema.

Acknowledgment

We would like to thank our colleagues who have helped in the development of independent parts of the software, but who wish to remain outside the scope of the co-authors of the article: Michael Samoilov, Nikita Durynin, Alexander Stupnikov and Dusso Bruno Jean-Marie.

References

[1] Patrick Ziegler, Klaus R. Dittrich, “Data Integration — Problems, Approaches, and Perspectives.” In Conceptual Modelling in Information Systems Engineering, Springer, 2007. (Conference or Workshop Paper). p. 39-58.

[2] Adila Alfa Krisnadhi. “Ontology Pattern-Based Data Integration”, PhD Thesis, Wright State University. December 18, 2015.

[3] Pahl, Claus and Zhu, Yaoling. “Data integration in mediated service compositions” In Computing and Informatics, Volume 31 Issue 6, 2001, pp. 1129-1149.

[4] Natalya F. Noy. “Semantic integration: a survey of ontology-based approaches.” In Newsletter ACM SIGMOD Record Homepage archive, Volume 33 Issue 4, December 2004. pp. 65-70.

[5] AnHai Doan, Alon Y. Halevy. “Semantic-integration research in the database community.” In Journal AI Magazine - Special issue on semantic integration archive, Volume 26 Issue 1, March 2005 pp. 83-94.

[6] Ramon Lawrence, Ken Barker. “Integrating relational database schemas using a standardized dictionary.” In Proceeding SAC '01 Proceedings of the 2001 ACM symposium on Applied computing. ACM New York, NY, USA, 2001, pp. 225-230.

[7] Stephen Mc Kearney, “Schema Integration.”, 2002. Available at URL: http://www.smckearney.com/adb/notes/lecture.schema.integration.pdf

[8] Batini C., Lenzerini M., Navathe S. B. “A comparative analysis of methodologies for database schema integration.” In Journal ACM Computing Surveys, Volume 18 Issue 4, Dec. 1986, pp. 323-364.

[9] Ivashko A.G., Grigoryev A.V., Grigoryev M.V. “Tableau algorithm modification based on complex concepts disjointness checking.” In Bulletin of the Tyumen State University. - 2012, №4. - P. 143-150.

[10] Motik B., Patel-Schneider P.F., Grau B.C. “OWL 2 Web Ontology Language Direct Semantics (Second Edition).” W3C Official Site. 2012. URL: http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211

[11] Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, Zhe Wang. “HermiT: An OWL 2 Reasoner” In Journal of Automated Reasoning October 2014, Volume 53, Issue 3, pp 245-269.

Identification of the Parallel Documents from Multilingual News Websites

Bagdat Myrzakhmetov, Aitolkyn Sultangazina, and Aibek Makazhanov
National Laboratory Astana
53 Kabanbay batyr ave., Astana, Kazakhstan
E-mail: {bagdat.myrzakhmetov, aitolkyn.sultangazina, aibek.makazhanov}@nu.edu.kz

***Abstract*—We present the initial results of our experiments on document alignment for the online news domain. Specifically, as apposed to cross-site comparable news alignment, we focus on the identification of parallel documents from within the same multilingual websites. In such a setting parallel news stories oftentimes turn out to be direct translations of each other with a tendency of sharing common media and displaying proximity in publication date. We leverage this domain-specific property of the data and propose a straightforward yet competitive heuristic that performs on par with a machine learning-based method in terms of precision, and outperforms a widely used bitext extraction system on a range of metrics. Moreover, this heuristic has allowed us to identify comparable documents overlooked by a human annotator.**

Although both rule- and learning-based methods that we present are language independent, we specifically focus on the Russian-Kazakh language pair as the present study is one of the initial steps towards a greater objective of building a corresponding parallel corpus and a machine translation system.

***Index Terms*—Document alignment, parallel corpus, machine translation**

I. Introduction

Parallel documents are segments of text written in different languages that constitute direct translation of each other. Identification of parallel documents, a.k.a. document alignment, is one of the initial steps in building parallel corpora [1], an absolutely vital resource for any statistical machine translation system. In the past various approaches to the problem has been proposed. Most of them can be classified into two groups, as methods that utilize document structure, size and meta information [2]–[4], as well as methods that attempt direct comparison of the content [5]–[7]. Hybrid approaches are common as well [2], [4]. Potential sources of parallel documents have also been explored in depth. Typically multilingual resources such as legal documents [5], movie subtitles [8], and various kinds of multilingual websites [2], [3], [9] are used to extract parallel documents.

In this paper we propose to extract parallel documents from multilingual news websites, where some or most of the news appear in multiple languages on the same website. In such a setting parallel news stories oftentimes turn out to be direct translations of each other with a tendency of sharing common media and displaying proximity in publication date. We leverage this domain-specific property of the data and propose a rule- and learning-based method that both take into account meta information, namely the date of publication and the headline picture used by a given pair of documents. The learning-based method is a hybrid approach that also considers various content similarity features. Our results suggest that although a rule-based heuristic achieves lower recall than a rule-based method, it has a marginal advantage in precision, without being burdened with the need of training data.

The rest of the paper is organized as follows. In the following section we provide a detailed description of rule- and learning-based approaches that we propose. Then in Section III we describe the experimental data and setup, discuss the results of the experiments, report on additional findings regarding comparable documents identification, and present some statistics on extracting parallel documents from the Web. Following that, in Section IV, we discuss the related work. Finally, we draw conclusions and discuss the future work in Section V.

II. Identification of Parallel Documents

In this section we describe in detail the rule- and learning-based methods that we have developed for the task of aligning news articles retrieved from the same website. Given a candidate pair of documents, both methods assume access to the following meta data of each document: (i) a URL link to an image used as the headline picture, and (ii) publication date. Whenever present this information can be easily obtained at the data collection step or afterwards, provided that it is preserved during retrieval.

A. Image and Date-based Heuristic

This method is based on the observation that a good portion of parallel news articles tend to use the same images[1] (usually a headline picture) and get published on the same day. Indeed in our data set of 586 pairs of parallel documents, which we will describe in greater detail in the following section, 475 pairs (81%) shared the same image and 573 (≈98%) were published on the same date.

Given this property of the data, we use the following simple rule: if given source and target articles have been published on the same day and they share the same image as the headline picture, consider them parallel, *provided that the shared image is not used by any other article (source and target alike)*

[1]Obviously, we mean images in the main content section, not the common elements of design.

TABLE I
FEATURES USED FOR THE IDENTIFICATION OF PARALLEL DOCUMENTS

#	Dom.	Description
1	ST	distance in days between publication dates
2	ST	same image (binary feature)
3	ST	same image and publication date (binary feature)
4,5	S,T	length in characters
6	ST	minmax ratio between F4 and F5
7,8	S,T	length in tokens
9	ST	minmax ratio between F7 and F8
10,11	S,T	number of paragraphs
12	ST	minmax ratio between F10 and F11
13,14	S,T	number of symbols
15	ST	minmax ratio between F13 and F14
16	ST	cosine similarity between F13 and F14 BoW vectors
17,18	S,T	number of numerals
19	ST	minmax ratio between F17 and F18
20	ST	cosine similarity between F17 and F18 BoW vectors
21,22	S,T	number of digits
23	ST	minmax ratio between F21 and F22
24	ST	cosine similarity between F21 and F22 BoW vectors
25,26	S,T	number of foreign alphanumerics
27	ST	minmax ratio between F25 and F26
28	ST	cosine similarity between F25 and F26 BoW vectors
29,30	S,T	number of alphanumerics
31	ST	minmax ratio between F29 and F30
32	ST	cosine similarity between F29 and F30 BoW vectors
33,34	S,T	number of tokens in quotes
35	ST	minmax ratio between F33 and F34
36	ST	cosine similarity between F33 and F34 BoW vectors
37,38	S,T	number of tokens in parenthesis
39	ST	minmax ratio between F37 and F38
40	ST	cosine similarity between F37 and F38 BoW vectors

published on that day. The italicized part of the definition ensures that the resulting list of candidate alignments has no two or more source documents aligned to a single target document and vice versa. More formally the application of the rule can be expressed by the following algorithm:

1) get a list D of source document publication dates;
2) $\forall d \in D$ perform steps (3)-(5);
3) for a set of source documents S_d published on day d and the corresponding collection of images build an image-to-document mapping I_d, removing one-to-many assignments if any;
4) repeat step (3) for the target documents T_d published on day d to get a mapping J_d;
5) add the document pairs obtained from the intersection $I_d \cap J_d$ to the collection of candidate alignments C.

In practice there is no need to deal with actual image files, as they can be represented by the corresponding URL links that web pages use to display them. Of course parallel or comparable news stories may use identical copies of the same image and access those through different URLs. In that case images can be downloaded and represented by their md5-sums. However, in our experience of mining Kazakhstani multilingual news sites, such usage of identical images is uncommon. This, however, maybe a behavior idiosyncratic to the investigated Internet segment.

B. Learning-based Method

Apart from the rule-based method we experiment with a machine learning approach that incorporates the image and date-based heuristic and additionally uses various content-based clues such as various length-, punctuation-, numeral-based metrics that have been found useful in the past [3], [6], [9]. In our feature design scheme we use a domain-based approach similar to the one used in [10]–[12], where for a given pair of entities (source and target documents in our case) asymmetric (per-entity) and symmetric (per-pair) features are computed. Thus, a single metric, e.g. length in characters, is computed for source (S-domain) and target (T-domain) documents, and the minimum-to-maximum ratio of the two is used as the feature for the pair (ST-domain). To emulate the rule-based method we use three features that, for a given candidate pair, measure the proximity of publication dates and set binary switches to 1 if the pair shares an image and the documents were published on the same date.

In total we use 40 features, which are listed in Table I together with their domain of application, i.e. S, T or ST. We use abbreviations of the type F<NUM> to refer to features with corresponding ordinal numbers, e.g. F17 refers to the feature #17. Passages such as *F<NUM> BoW vector* should be read as bag of "words" vector for entities which the corresponding feature takes into account, e.g. feature #16 (cosine similarity between F17 and F18 BoW vectors) calculates cosine similarity between bag of symbols (anything non alphanumeric) vectors of source and target documents. Lastly, by *foreign alphanumeric* (F25-26) we mean tokens written in an alphabet that is not used by neither source or target language. In our case it is the Latin alphabet (Kazakh is written in modified Cyrillic, which, in addition to all the letters of standard Russian Cyrillic, includes nine letters for language-specific sounds).

III. EXPERIMENTS AND EVALUATION

In this section we describe the experimental data and setup, discuss the results of the experiments, report on additional findings regarding comparable documents identification, and present some statistics on extracting parallel documents from the Web.

A. Data Set Description

In order to be able to evaluate document alignment methods against the ground truth, we have prepared a hand-labeled data set. First, from a major multilingual news website (`www.kazinform.kz`) we have downloaded news articles in Russian and Kazakh that were published in the last 10 days (as of May, 2016). This way we have obtained 1060 Russian and 1010 Kazakh articles. Then a human annotator

has checked the entire collection and labeled 586 documents as parallel.

B. Experimental Setup

We compare a total of four methods two of which are our rule- and learning-based methods described in the previous section. We also evaluate the combination of these two methods, as well as Bitextor [3], a system for automatic extraction of parallel corpora from the Web. The latter we use in a document alignment mode, in which the input is a directory that contains source and target documents in two separate subdirectories with all the HTML-code preserved, and the output is a list of document pairs identified as parallel by the tool. We also provide Bitextor with a Russian-Kazakh dictionary[2] that contains about 116K entries.

For learning-based method we extract the features described in Section II-B for a number of selected candidate pairs. A number of all possible candidates totals to ≈1M pairs, which is not extremely large and certainly manageable by most of today's hardware. But the problem is that in such a set of training instances the share of negative examples would be extremely large (≈0.9995), which may cause over-fitting of learning algorithms. To account for this we employ a simple candidate selection strategy, which is, again, motivated by the domain-specific properties of the data, namely publication date proximity of the parallel news articles. We extract features only for those document pairs whose target documents appeared no later than on the next day after the source. This modification has allowed us to reduce the number of candidates nearly five fold to ≈227K pairs. As for the actual learning algorithm, we have experimented with several classifiers offered by Weka machine learning toolkit [13], and chose the best performing decision tree-based Random Forrest classifier.

The application of the rule-based method is straightforward: it boils down to the implementation of the algorithm described in Section II-A. Finally, the combination of the rule- and learning-based methods is based on the idea of using the output of the former to train the latter. This is achieved as follows. We build a training set by employing the aforementioned candidate selection strategy, but this time we discard candidate pairs whose source documents do not match the ones retrieved by the rule based-method, even if they satisfy publication date constrains. This allows us to collect a total of ≈90K training instances. We then proceed to use the classifier to identify parallel documents from the remaining set of ≈137K candidate pairs (which were not covered by the rule-based method). The pairs that were labeled parallel by the classifier are then joined with those rendered parallel by the rule-based method, and the resulting set is evaluated.

All of the methods are evaluated in terms of precision (portion of the retrieved set covered by the correctly identified pairs), recall (portion of the golden set covered by the correctly identified pairs), and F-measure, a weighted harmonic mean of the two. The classifier is used in a 10-fold cross-validation setting, where we report results per positive class averaged over folds.

[2] www.mtdi.kz/til-bilimi/sozdikter/oryssha-kazaksha

TABLE II
RESULTS

Method	Pairs retrieved	Precision	Recall	F-measure
RBM	415	**0.93**	0.66	0.77
LBM	495	0.92	0.78	**0.85**
RBM+LBM	769	0.62	**0.81**	0.70
Bitextor	796	0.49	0.67	0.56

C. Results

Table II contains the results of the experiments. As it can be seen, the rule-based method (RBM) achieves the highest precision of 0.93, but conversely the lowest recall of 0.66. As a result, in terms of F-measure the image and date-based heuristic is second only to the learning-based method (LBM), which scores 0.85 on this metric and loses a negligible 0.01 in precision to LBM. The combination of these methods, as one would expect, outperforms both of them in terms of recall, achieving the highest score of 0.81, but displays the second lowest precision of 0.62 unacceptable for practical use. Perhaps most surprisingly, Bitextor, the only tool, which was granted the access to the language-specific information (a bilingual lexicon), has achieved the lowest scores in terms of all metrics (if we were to ignore a negligible 0.01 gain in precision over RBM).

Although learning-based approach outperforms the heuristic in terms of recall and F-measure, from a practical point of view, it has several drawbacks which the heuristic does not have. First, it requires training data, which is not easy to obtain. Second, it doesn't scale well with the size of the input. Lastly, it gives away a smidgen of advantage in precision, which is often traded over recall. Had the gap in precision been wider, we would prefer the heuristic to LBM. In fact, the error analysis that we present in the following subsection suggests that the said gap may, indeed, be wider, as errors made by the heuristic may not be as absolute as we thought.

D. Identification of Comparable Documents

As our experiment has shown, the image and date-based heuristic predicted 415 pairs as parallel (cf. Table II). Of those 387 were present in the golden set, i.e. were correct, which means that precision of the method was $387/415 \approx 0.93$. We have analyzed the remaining 28 pairs which were not found in the golden set, and concluded that 20 of those were in fact pairs of *comparable documents*, which convey the same meaning expressed in two different languages, but do not constitute a direct translation. It was shown [14], [15] that such documents can be useful in various machine translation related tasks. Thus, we speculate that the heuristic may also be useful in identification of comparable documents and in building corresponding corpora.

Lastly, if we were to include those 20 pairs of comparable documents to the golden set, the heuristic would achieve

TABLE III
PARALLEL AND COMPARABLE DOCUMENTS EXTRACTION

Website	Docs, Source	Docs, Target	Retrieved Pairs
www.kazinform.kz	48116	52634	20720
www.expo2017astana.com	352	393	279
www.kaztag.kz	1292	1966	1124
www.mfa.gov.kz	1208	2158	970
total	**50968**	**57151**	**23093**

$(387+20)/415 \approx 0.98$ in precision, which makes this method very appealing to use in practice. In the next subsection we present the the results of our first attempt to do so.

E. Extracting Parallel and Comparable Documents

We have crawled all of the accessible news articles, from four multilingual news websites listed in Table III, and retrieved a total of 50968 source (Kazakh) and 57151 target (Russian) documents[3]. We then fed this data as input to the image and date-based heuristic and retrieved a total of 23093 document pairs as candidates for being parallel and possibly comparable.

In the future we plan to extract parallel sentences from this data. The success of this endeavor depends heavily on the quality of the obtained document alignment. Thus, even though the precision of the heuristic is fairly high, we plan to filter out document pairs that are unlikely to be parallel by applying the STRAND algorithm [2] or our learning-based method.

IV. RELATED WORK

Approaches to the document alignment problem can be broadly divided into three types: (i) structure-based approaches [2]–[4] that utilize non-contextual clues, such as document structure, mark-up and URL similarity; (ii) content-based approaches [4]–[7] that given a pair of documents attempt direct comparison of their content as a whole [7] or as a collection of specific features, such as punctuation, numerals, or infrequent words [4]–[6]; and (iii) the combinations of the two [2], [4].

One of the earliest structure-based approaches uses the STRAND (Structural Translation Recognition Acquiring Natural Data) algorithm developed by Resnik and Smith [2]. This approach attempts to generate candidate pairs of parallel documents based on the text lengths and HTML markup structure. Additionally four features are used to evaluate the quality of document alignment: the percentage of differences in documents, the number of aligned non-markup chunks of unequal length, the length correlation of aligned non-markup chunks, significance level of that correlation.

Ma and Liberman [4] develop a document-aligner tool BITS (Bilingual Internet Text Search), which uses URL matching techniques. Namely, the authors compute Levenshtein distance [16] between the URLs of candidate pairs, assuming that the parallel documents will have the similar URL names. Another approach [17] uses a similar but simpler strategy which is to extract candidate pairs that only differ in language sub-part of the URL structure, such as en, .e., etc. Espla-Gomis [3] developed the tool called Bitextor which uses various heuristics to identify parallel texts, such as file name extension comparison, file size ratio, total length difference. After applying these heuristics, the tool performs fingerprint comparison: for tags and text blocks it applies insertion, deletion and substitution cost functions to identify parallel documents.

The content-based approaches utilize contextual clues that can be used heuristically or in a learning-based fashion. Patry and Langlais [6] use features such as named entities, punctuation, numbers as vectors to identify parallel documents. Cosine similarity of vectors used to measure the quality of parallel documents. Another method which was used by the authors relies on computation of edit distances between candidate pairs, i.e. how many insertions, deletions and substitutions might be needed to convert one document to its candidate document. Ma and Liberman [4] propose to use lexical similarity of candidate pairs and utilize bilingual dictionaries.

Some authors used different features to identify parallel documents effectively and to optimize feature spaces. Enright and Kondrak [5] uses cognate-based approach: they used only *hapax* words as features, i.e. words that appear only once in a document (additional condition was for these words to be at least four characters long). This simple method surprisingly showed good results on aligning of 11 European languages. A drawback of this method is the assumption that for a given document there is always a parallel counterpart. Thus, the method may not work properly on a collection where some documents are not parallel.

An alternative method described by Uszkoreit et al. [7] utilizes the cross-language information. Initially all of the document pairs are translated into a single language by means of a baseline translation system. To detect duplicated documents the authors use standard information retrieval techniques: they identified parallel documents with n-gram matches and matching of rare words. The last one sometimes might be problematic, as a baseline translation system does not always translate rare words. To make document pair-wise comparison effectively, they use index structure.

V. CONCLUSIONS AND FUTURE WORK

In this work we have reported on our initial experiments on parallel documents identification task. We have developed and tested rule- and learning-based approaches that both take into account domain-specific properties of news articles obtained from the same multilingual news websites. Our results suggest that a common practice (employed by such websites) of publishing translated news stories in close temporal proximity and sharing media between them can be exploited to identify parallel and comparable documents with high precision.

In the future we plan to work on improving the recall of our approach. To this end we plan to use the combination of

[3]The choice of languages of source and target documents here is arbitrary. The actual translation direction may differ.

our rule- and learning- based approaches in a more efficient manner. We also plan to experiment with linguistic features to utilize content-based clues more effectively. To this end we plan to move in a language-specific direction and explore possibilities of enriching our approach with morphological [18]–[23] and syntactic [24], [25] parsing techniques developed for Kazakh language.

ACKNOWLEDGMENT

This work has been funded by the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan under the targeted program O.0743 (0115PK02473), and by the Nazarbayev University under the research grant 064-2016/013-2016.

REFERENCES

[1] J. Tiedemann, "Bitext alignment," *Synthesis Lectures on Human Language Technologies*, vol. 4, no. 2, pp. 1–165, 2011.

[2] P. Resnik and N. A. Smith, "The web as a parallel corpus," *Comput. Linguist.*, vol. 29, no. 3, pp. 349–380, Sep. 2003.

[3] M. Esplà-Gomis, "Bitextor: a Free/Open-source Software to Harvest Translation Memories from Multilingual Websites," in *Proceedings of MT Summit XII*. Ottawa, Canada: Association for Machine Translation in the Americas, 2009.

[4] X. Ma and M. Y. Liberman, "Bits: A method for bilingual text search over the web," in *In Proceedings of the Machine Translation Summit VII*, 1999.

[5] J. Enright and G. Kondrak, "A fast method for parallel document identification," in *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA*, 2007, pp. 29–32.

[6] A. Patry and P. Langlais, "Automatic identification of parallel documents with light or without linguistic resources," in *18th Annual Conference on Artificial Intelligence (AI)*, Victoria, British-Columbia, Canada, May 2005, pp. 354–365.

[7] J. Uszkoreit, J. M. Ponte, A. C. Popat, and M. Dubiner, "Large scale parallel document mining for machine translation," in *Proceedings of the 23rd International Conference on Computational Linguistics*, ser. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 1101–1109.

[8] J. Tiedemann, "Building a multilingual parallel subtitle corpus," *Proc. CLIN*, p. 14, 2007.

[9] A. Patry and P. Langlais, "Identifying parallel documents from a large bilingual collection of texts: Application to parallel article extraction in wikipedia," in *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web*. Association for Computational Linguistics, 2011, pp. 87–95.

[10] H. R. Sepehri, A. Makazhanov, D. Rafiei, and D. Barbosa, "Leveraging editor collaboration patterns in wikipedia," in *Proceedings of the HT Conference*, 2012, pp. 13–22.

[11] A. Makazhanov and D. Rafiei, "Predicting political preference of twitter users," in *Proceedings of the 2013 ASONAM Conference*, 2013, pp. 298–305.

[12] A. Makazhanov, D. Rafiei, and M. Waqar, "Predicting political preference of twitter users," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–15, 2014.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," 2009.

[14] D. S. Munteanu and D. Marcu, "Improving machine translation performance by exploiting non-parallel corpora," *Computational Linguistics*, vol. 31, no. 4, pp. 477–504, 2005.

[15] S. Sharoff, R. Rapp, and P. Zweigenbaum, "Overviewing important aspects of the last twenty years of research in comparable corpora," in *BUCC: Building and Using Comparable Corpora*, S. Sharoff, R. Rapp, P. Zweigenbaum, and P. Fung, Eds. Springer, 2013, pp. 1–17.

[16] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," in *Soviet physics doklady*, vol. 10, 1966, p. 707.

[17] J. Chen and J.-Y. Nie, "Automatic construction of parallel english-chinese corpus for cross-language information retrieval," in *Proceedings of the Sixth Conference on Applied Natural Language Processing*, ser. ANLC '00. Stroudsburg, PA, USA: Association for Computational Linguistics, 2000, pp. 21–28.

[18] A. Sharipbayev, G. Bekmanova, B. Ergesh, A. Buribayeva, and M. K. Karabalayeva, "Intellectual morphological analyzer based on semantic networks," in *Proceedings of the OSTIS-2012*, 2012, pp. 397–400.

[19] A. Makazhanov, Z. Yessenbayev, I. Sabyrgaliyev, A. Sharafudinov, and O. Makhambetov, "On certain aspects of kazakh part-of-speech tagging," in *Application of Information and Communication Technologies (AICT), 2014 IEEE 8th International Conference on*. IEEE, 2014, pp. 1–4.

[20] G. Kessikbayeva and I. Cicekli, "Rule based morphological analyzer of kazakh language," in *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 46–54. [Online]. Available: http://www.aclweb.org/anthology/W/W14/W14-2806

[21] A. Makazhanov, O. Makhambetov, I. Sabyrgaliyev, and Z. Yessenbayev, "Spelling correction for kazakh," in *Proceedings of the 2014 Computational Linguistics and Intelligent Text Processing*. Kathmandu, Nepal: Springer Berlin Heidelberg, 2014, pp. 533–541.

[22] J. Washington, I. Salimzyanov, and F. Tyers, "Finite-state morphological transducers for three kypchak languages," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.

[23] O. Makhambetov, A. Makazhanov, I. Sabyrgaliyev, and Z. Yessenbayev, "Data-driven morphological analysis and disambiguation for kazakh," in *Computational Linguistics and Intelligent Text Processing - 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part I*, 2015, pp. 151–163.

[24] F. M. Tyers and J. N. Washington, "Towards a free/open-source universal-dependency treebank for kazakh," in *3rd International Conference on Turkic Languages Processing, (TurkLang 2015)*, 2015, pp. 276–289.

[25] A. Makazhanov, A. Sultangazina, O. Makhambetov, and Z. Yessenbayev, "Syntactic annotation of kazakh: Following the universal dependencies guidelines. a report," in *3rd International Conference on Turkic Languages Processing (TurkLang 2015)*, Kazan, Tatarstan, 2015, pp. 338–350.

Maximum Relevance and Class Separability for Hyperspectral Feature Selection and Classification

Saeed Jahanshahi
Department of Electrical Engineering
Tabriz Branch, Islamic Azad University
Tabriz, Iran
s.jahanshahi1392@gmail.com

***Abstract*—Regarding a growing interest into exploiting hyperspectral images in the plethora of applications such as chemical material identification, agricultural crop mapping, military target detection and etc., myriad approaches have been introducing to interpret and analyze such data. In this paper, I am going to propose a novel method using the combination of two conventional method. Firstly, I use an evolutionary algorithm i.e., multi-objective particle swarm optimization (MOPSO) to select a predefined number of features (spectral bands) and then a well-known classifier i.e., support vector machines (SVMs) is deployed for classification.**

***IndexTerms*—Accuracy rate (AR), Bhattacharyya distance (BD), binary multi-objective optimization (BMO), classification, feature selection, hyperspectral imaging, mutual information (MI), particle swarm optimization (PSO), support vector machines (SVMs).**

I. Introduction

Hyperspectral sensors are able to capture images of a scene in hundreds of narrow contiguous frequency bands throughout electromagnetic spectrum [1]. This attribute facilitates identification and detection tasks according to electromagnetic interactions with matter because materials has different behavior in different wavelength. However there are various serious challenges such as dimensionality curse, spectral variability, and atmospheric compensation [2], so solutions have been proposed in different taxonomies. [1][3][4]

High dimensionality cause trouble for hyperspectral data processing so that it is unavoidable to reduce dimensionality.

Feature selection is a sophisticated method for declining some dimensions without substantial deficiency of information or transforming the original samples into different spaces which makes the interpretation of the outcomes more complicated and difficult. [5] Feature selection is a treatment for Hughes effect, happening when the ratio of training samples to the number of bands is small so that the spectra cannot be modeled accurately. [6]

By means of feature selection run time might be shortened [7], there will be chances to save storage space, computational cost and communication bandwidth. [8]

Among algorithms, the evolutionary ones belonging to the wrapper methods [9] is preferred to take their advantages especially wide search space which gives you the opportunity to check many random arrangement of features rapidly. Furthermore they are capable for modifying to multi-objective approaches which let you to optimize some fitness functions simultaneously. [10] Feature selection techniques have two parts: search algorithm and criterion. [11] Among search algorithms, particle swarm optimization (PSO) was designated for its great propensities to converge rapidly and its compatibility for dealing with binary data which is required for band selection. [12], [13]

By all above mentioned, a binary multi-objective particle swarm optimization (BMOPSO) with two fitness functions as criteria is going to be introduced. The mutual information (MI) between each spectral band and ground truth which is a good criterion for finding maximum relevance between them, is one of the criteria. [14] Bhattacharyya distance (BD) is the second fitness function as a proved class separability criterion [15].

In terms of classification, support vector machines (SVMs) is chosen for:

1) Less sensitivity to data's dimensionality, the size and combination of training samples [16]-[17]

2) Inferior execution time [17]

3) Regard to the conviction that only support vectors (among all training samples) contribute to estimate optimum hyperplane, so SVMs classifiers may save operating cost. [18]

The reminder of this paper is arranged as bellow:

Section II is a description for commonly used hyperspectral dataset, i.e., AVIRIS 92 AVC3C. Section III describes BMOPSO algorithm. Section IV expresses how to select optimum predefined number of Bands using BMOPSO algorithm. Section V presents a review of hyperspectral classification using SVMs. Section VI illustrates experimental results and section VII is for conclusion.

II. AVIRIS 92AVC3C

In this paper, a part of a scene including 220 spectral bands which has been taken by the airborne visible/infrared imaging spectrometer (AVIRIS) over NW Indiana's Indian pine, is exploited. 20 channels were discarded due to being influenced by atmospheric absorption.

III. Multi-Objective Particle Swarm Optimization (MOPSO)

A brief explanation of a MOPSO algorithm which has been proposed by Carlos A. Coello Coello and *et al.* in [19] is presented as a platform for the feature selection step as below:

1) Particles should be populated initially

2) Find particles which are not dominated by others so that:

Assume x and y are two particles so xf and yf would be their associated fitness functions vectors.

$$\forall i \quad xf_i \leq yf_i \quad , \quad i,j = 1,\dots,n$$
$$\exists j \quad xf_j < yf_j \qquad (1)$$

Then

$x \preceq y$ (Where $\preceq$ is the sign of domination.)

3) Create grids in the search space to pinpoint particles' location according to their fitness functions values as described in [19].

4) Implement the main loop (for it=1: maxiteration)

5) Within the main loop update all attributes of the particles

(a) for i=1: popsize

$$\text{pop(i).velocity} = w \times \text{pop(i).velocity} + c_1 r_1(\text{pop(i).Best.position} - \text{pop(i).position}) + c_2 r_2(\text{leader.position} - \text{pop(i).position}) \qquad (2)$$

Where w is inertia weight, r_1 , r_2 are random numbers in the range [0,1]. c_1 is cognitive acceleration coefficient and c_2 is social coefficient, the first term is called 'inertia', , the second term is called 'self-knowledge' and the last term is called 'social knowledge'. [20]

Leader is a member of the repository which is selected as below:

Suppose that n_m particles reside in the *m*th cell of the grid. P_m is the probability of the selection of the cell, so that:

$$P_m = \frac{e^{-\beta \times n_m}}{\sum_{l=1}^{N_c} e^{-\beta \times n_l}} \qquad (3)$$

β is the parameter of selection pressure and N_c is the number of occupied cells.

Now it is the turn of Roulette Wheel to select a cell according to the probability and you can choose one particles in the cell randomly.

(b)

$$\text{pop(i).position} = \text{pop(i).position} + \text{pop(i).velocity} \qquad (4)$$

(c) Evaluate all particles in pop

6) Let non-dominated particles become a member of the repository and then run a competition among the newcomer members and the old members of the repository to eliminate dominated members if there is. A temporary task would be done to eliminate those particles located in highly populated regions when the repository is stuffed full.

7) Reallocate pop(i).Best.position with the winner of a contest for domination between pop(i).Best.position and pop(i).position.

End

IV. Band Selection Using BMOPSO

Manipulated version of the MOPSO algorithm in [19] is required to deal with binary problems. Particles with binary bits play the band selection role in the PSO part. Furthermore there is different relation between binary particles and their velocity so that we must convert particles' position vectors obtained by Eq. 2 and Eq. 4 which are in continuous space to binary particles by calculating probabilities in Eq. 5, and then spinning a roulette wheel k times to select predefined number of features :

$$p_{ij} = \frac{{x_{ij}}^{\alpha}}{\sum_{j=1}^{n} {x_{ij}}^{\alpha}} \qquad (5)$$

p_{ij} is the probability of the jth bit in the ith particles and x_i is the position vector of the ith particles and k is the number of features which must be selected [21].

However, my initiative is to deploy BMOPSO for optimizing two objective function simultaneously in order to select optimum spectral bands and finally different arrangement and combination of k features. The first one is the average MI between each band and ground truth:

$$AMI = \frac{1}{|S|}\sum_{i \in S} MI(B_i, G) \qquad (6)$$

Where S is the set of the selected bands, B_i is the ith band, G is ground truth and:

$$MI(B,G) = \sum_{m=1}^{M}\sum_{n=1}^{N} P(B_m, G_n) \log \frac{P(B_m,G_n)}{P(B_m)P(G_n)} \qquad (7)$$

M, N are the number of Bands' levels. There is 256 level for all bands and 9 level for ground truth indicating class number. P(.,.) is the joint probability. In this regard optimum AMI leads to maximizing relevance between each band and ground truth. Maximum BD between two classes is the second fitness function which is a class separability index:

$$BD(C_1, C_2) = \frac{1}{8} \times \left[\left(\frac{(\mu_1 - \mu_2)}{\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)}\right) \times (\mu_1 - \mu_2)'\right] + \frac{1}{2} \times \ln\left(\frac{\left|\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)\right|}{\sqrt{|\Sigma_1| \times |\Sigma_2|}}\right). \qquad (8)$$

Where μ_1 , μ_2 are the mean vector of class1 and class2 and $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ are the covariance matrix of class1 and class2 respectively.

Maximum relevance between each band and ground truth and maximum class separability is desirable, so AMI and BD must be inversed to be compatible with MOPSO procedure.

$$I = \frac{1}{AMI} \qquad J = \frac{1}{BD} \qquad (9)$$

V. Classification Using Support Vector Machines

The linear SVM classification approach involves finding an optimal hyperplane to maximize the separating margins, whereas in the nonlinear case, optimal hyperplane is the one that minimizes a cost function:

$$\Psi(\omega, \xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \quad (10)$$

Subject to the constraints:

$$y_i(\omega^T.\phi(x_i) + b) \geq 1 - \xi_i \qquad i = 1,2,\dots,N$$

And

$$\xi_i \geq 0$$

Where ξ_i is called "slack variable" that measures how corresponding data points are misclassified by the hyperplane, C is the regularization parameters that determines a trade-off between margin maximization and the minimization of the classification error, $y_i \epsilon \{-1, +1\}$ are the correspondent targets and there are N vectors of the dataset as the training part so that d is the number of bands (dimensions).

$\varphi(x) \in R^{d'}(d' > d)$, is a kernel function that maps data into higher dimensional feature space where is a decision rule is carried out according to the sign of a distinguishing function:

$$f(x) = w^T\varphi(x) + b \quad (11)$$

And $w \in R^{d'}$, is the weight vector and $b \in R$ is the bias term.

However the above mentioned optimization problem is not solvable, but there is a dual optimization which leads to a quadratic programming (QP) solution i.e.:

$$\max_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \quad (12)$$

By the constraints: $\alpha_i \geq 0$, $i = 0,1,2,\dots,N$. And $\sum_{i=1}^{N}\alpha_i y_i = 0$.

Where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$, is a vector of Lagrange multipliers and then the decisin function is:

$$f(x) = \sum_{i\in S}\alpha_i y_i k(x_i, x) + b \quad (13)$$

Where S refers to the set of support vectors and K(.,.) is a kernel function such as: gaussian radial basis function:

$$K(x_i, x) = e^{-\gamma\|x_i - x\|^2} \quad (14)$$

Where γ is a parameter inversly proportional to the width of the guassian kernel and x_i , x are data vectors. [22]

VI. Experimental Results

The results of my proposed method is going to be reported in this section with the aim of evaluating its performance in terms of accuracy rate (AR) and execution time. The experiments includes two phases:

A. Feature selection phase

Feature selection phase has been run as preprocessing step based on BMOPSO method. The relevant repository has 11 members which are various combination of feature set with 87 numbers of bands. For each solution (particle) in the repository, there are two values corresponding to its fitness functions. Table I illustrates the repository members.

B. Classification phase

The fourth and ninth member of the repository is given to the SVMs classifier as a features set.

The first row of Table II includes the results of the second phase for Aviris dataset with all 200 spectral bands. Note that the second row in Table II belongs to the fourth member of the repository and the third row belongs to the ninth row of the repository.

TABLE I. The Repository including two criterion

NO	I	J
1	1.30667967	1.437152491
2	1.30667967	1.437152491
3	1.319650434	1.324035325
4	1.32161914	1.309764633
5	1.310288187	1.377966551
6	1.319049971	1.360561708
7	1.313632556	1.361887549
8	1.319347221	1.347017088
9	1.340858724	1.297801807
10	1.321770486	1.303543269
11	1.319347221	1.347017088

As mentioned in section I, the feature selection mission has been accomplished successfully because regarding Table II, if we tune C and gamma parameters properly, we can save time about 1200s and storage space around 2.12 MB without decrease in overall AR.

VII. Conclusion

To my best knowledge, this paper is the first one so far, in which a combination of new or conventional methods i.e., BMOPSO and SVMs is deployed to select optimum predefined number of features (spectral bands) by evaluating two criteria simultaneously for hyperspectral images. The results of the previous section proved that the proposed method has been able to reduce the dimensionality of data while saving execution time and storage space without loss in overall AR.

Acknowledgment

I am inclined to express my immense gratitude to Prof. D. Landgrebe for providing the AVIRIS data [23].

TABLE II. CLASSIFICATION RESULTS FOR AVIRIRS DATASET

C	γ	AR1	AR2	AR3	AR4	AR5	AR6	AR7	AR8	AR9	Overall AR	Overall Time	No of Band	Storage Space (MB)
40	0.2	91.47	83.41	94.93	99.72	100	88.56	91.90	92.23	9828	92.89	5495	200	3.8
25	1.7	91.04	85.46	97.47	99.72	100	88.77	91.58	89.46	98.44	92.92	4182	87	1.68
30	1	90.75	84.44	95.78	99.44	100	88.77	92.39	89.96	98.13	92.87	4251	87	1.67

REFERENCES

[1] N. M. Nasrabadi, "Hyperspectral Target Detection: An Overview of Current and Future Challenges," in *IEEE Signal Process. Mag*, vol. 31, no. 1, pp. 34-44, Jan. 2014.

[2] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral Image Processing for Automatic Target Detection Applications," in LINCOLN LABORATORY JOURNAL, vol. 14, no. 1, pp. 79-116, 2003.

[3] D. Manolakis, E. Truslow, M. Pieper, T. Cooley and M. Brueggeman, "Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms," in *IEEE Signal Process. Mag*, vol. 31, no. 1, pp. 24-33, Jan. 2014.

[4] C. I. Chang, Hyperspectral Data Processing: Algorithm Design and Analysis. JohnWiley, 2013.

[5] C. Conesea and F. Masellia, "Selection of optimum bands from TM scenes through mutual information analysis," ISPRS J. Photogramm. Remote Sens., vol. 48, no. 3, pp. 2–11, 1993.

[6] M. Pal and G. M. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," in *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297-2307, May 2010.

[7] Pasi Luukka, "Feature selection using fuzzy entropy measures with similarity classifier, "in *Expert Syst. Appl*, vol.38, no.4, pp.4600-4607, Apr, 2011.

[8] B. Guo, S. R. Gunn, R. I. Damper and J. D. B. Nelson, "Band Selection for Hyperspectral Image Classification Using Mutual Information," in *IEEE Geosci. Remote Sens. Lett*, vol. 3, no. 4, pp. 522-526, Oct. 2006.

[9] G. Chandrashekar , F. Sahin, "A survey on feature selection methods," in *Comput. Electr. Eng*, vol.40, no.1, pp.16-28, Jan 2014.

[10] B. Xue, M. Zhang and W. N. Browne, "Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach," in *IEEE Trans. Cybern*, vol. 43, no. 6, pp. 1656-1671, Dec. 2013.

[11] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," in *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360-1367, Jul 2001.

[12] S. T. Monteiro and Y. Kosugi, "A particle swarm optimization-based approach for hyperspectral band selection", *Proc. IEEE Congr. Evol. Comput.*, pp. 3335-3340, 2007.

[13] A. S. J. Tj iong, S. T. Monteiro, "Feature selection with PSO and kernel methods for hyperspectral classification," in Proc of IEEE Congress on Evolutionary Computation, pp. 1 762-1 769, Jun 2 0 1 1.

[14] M. Kamandar and H. Ghassemian, "Maximum relevance, minimum redundancy feature extraction for hyperspectral images," 2010 18th Iranian Conference on Electrical Engineering, Isfahan, Iran, 2010, pp. 254-259.

[15] D. Landgrebe, "Hyperspectral image data analysis," in *IEEE Signal Process. Mag*, vol. 19, no. 1, pp. 17-28, Jan 2002.

[16] M. Pal, P. M. Mather, "Assessment of the effectiveness of support vector machines for hyperspectral data, "in *Future Gener Comp Sy*, vol.20, no.7, p.1215-1225, Oct 2004.

[17] L. Bruzzone, F.Melgani, "Classification of hyperspectral images with support vector machines: multiclass strategies," *in Proc. SPIE 5238, Image and Signal Processing for Remote Sensing IX, 408 Feb 2004.*

[18] A. J. Plaza,"Recent developments and future directions in hyperspectral data classification," in*Proc. SPIE 6748, Image and Signal Processing for Remote Sensing XIII, 67480A*, Oct 2007.

[19] C. A. C. Coello, G. T. Pulido and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," in *IEEE Trans. Evol. Comput,* vol. 8, no. 3, pp. 256-279, June 2004.

[20] A. R. Jordehi and J. Jasni, "Parameter selection in particle swarm optimization: a survey," J. Exp. Theor. Artif. Intell, vol. 25, Issue 4, pp. 527-542, Jun 2013.

[21] D. K. Agrafiotis and W. Cedeno, "Feature selection for structure-activity correlation using binary particle swarms," Journal of Medicinal Chemistry, vol. 45, pp. 1098–1107, 2002.

[22] Y. Bazi and F. Melgani, "Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images," in *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374-3385, Nov. 2006.

[23] Pursue's University Multispec Site: AVIRIS image Indian Pine Test Site, Jun. 12, 1992. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

A Sentence Selection Model and HLO Algorithm for Extractive Text Summarization

Rasim Alguliyev
Institute of Information Technology
Baku, Azerbaijan
r.alguliev@gmail.com

Ramiz Aliguliyev
Institute of Information Technology
Baku, Azerbaijan
r.aliguliyev@gmail.com

Nijat Isazade
Institute of Information Technology
Baku, Azerbaijan
nijatisazade@gmail.com

***Abstract*— In paper text summarization represented as a sentence scoring and selection process. The process is modeled as a multi-objective optimization problem. The proposed model attempts to find balance between coverage and redundancy in a summary. For solving the optimization problem a human learning optimization algorithm is utilized.**

***Index Terms*—Text summarization, sentence scoring, sentence selection, multi-objective optimization model, human learning optimization**

I. Introduction

Despite the fact that the first work dedicated to automatic text summarization (ATS) is published in 1958 [1], interest to this problem has gained increasing attention in the last two decades, after appearing the Web and e-government. Because of the large amounts of text documents are created in the web and e-government and their volume increases exponentially along years. In result, increasing the volume of text documents has made difficult for users to read and extract useful information from them. So, with continuing increase of the text documents ATS has become very important research direction and consequently attracted the attention of many researchers for the last few years [2–5].

ATS is defined as a process of automatically creating a condensed version of a source document preserving its information content. ATS is one of the most important technologies which extract relevant information from large text databases and present it to users. It helps users to access the information more easily by reducing the time [6–9].

This paper is focused on the extractive text summarization where a summary is generated by scoring and selecting the sentences in the source text. At first it assesses the score of each sentence and then selects the most representative sentences from the text by taking into account that semantic similarity between selected sentences will be low. For scoring the sentences a new formula is introduced. For selection of the sentences an optimization model is proposed. This approach allows create an optimal summary by avoiding redundancy and providing wide coverage in the summary under the assumption that summary is upper bounded by a specified length. In other words, the model guarantees that the selected sentences will convey the main content of the source text, on the one hand, and on the other in the summary will not sentences conveying same information. For solving the optimization problem a human learning optimization (HLO) algorithm is developed.

II. Preliminaries

Let $\mathbf{D}=\{d_1,...,d_{ND}\}$ be a set of documents, where ND is the number of documents. Suppose that the $\mathbf{D}$ is presented as a set of sentences $\mathbf{S}=\{s_1,...,s_n\}$, where s_i is the i th sentence and n is the total number of sentences in $\mathbf{D}$.

Sentence representation. Let $\mathbf{T}=\{t_1,t_2,...,t_m\}$ be the distinct terms appeared in $\mathbf{D}$, where m is the number of terms. Using vector space model each sentence s_i is represented as a vector in m-dimensional Euclidean space, $s_i=[w_{i1},...,w_{im}]$, ($i=1,...,n$), where w_{ij} is the weight of term t_j in sentence s_i.

Term weighting. For weighting the terms the TF-IDF scheme is used:

$$w_{ij}=tf_{ij}\times isf_j. \quad (1)$$

In Eq. (1), tf_{ij} is the frequency of term t_j in sentence s_i

$$tf_{ij}=m_{ij}/m,\ i=1,...,n;\ j=1,...,m \quad (2)$$

and isf_j is the inverse sentence frequency:

$$isf_j=\log(n/n_j),\ j=1,...,m\ , \quad (3)$$

where m_{ij} is the number of occurrences of term t_j in sentence s_i and n_j is the number of sentences in which the term t_j appears.

Similarity measure. Cosine similarity is the most popular measure used in text summarization to evaluate semantic similarity of texts. It calculates similarity between two vectors $s_i=[w_{i1},...,w_{im}]$ and $s_j=[w_{j1},...,w_{jm}]$:

$$sim(s_i,s_j)=\frac{\sum_{l=1}^{m}w_{il}w_{jl}}{\sqrt{\sum_{l=1}^{m}w_{il}^2\cdot\sum_{l=1}^{m}w_{jl}^2}},\ i,j=1,...,n. \quad (4)$$

III. Modeling Text Summarization

In this section, to create an optimal summary we propose a multi-objective optimization model. The objective function aims to balance coverage and redundancy in a summary under the assumption that length of the summary is upper-bounded by a predefined value.

Coverage: A created summary should cover the major topics and subtopics of the source text.

Redundancy: In summary should not be presented the sentences conveying similar information.

Length: A summary should be upper-bounded in length by a predefined value.

Nowadays, optimize all abovementioned properties jointly is a challenging task in text summarization. So, in this study we aim select from the set of sentences $\mathbf{S}=\{s_1,\ldots,s_n\}$ a subset of sentences so that the created summary has satisfied the abovementioned properties. The aim has been reached by modeling text summarization as optimization problem. The optimization problem can be modeled as:

maximize

$$f(x)=\lambda\cdot f_{\text{cov}}(x)+(1-\lambda)\cdot f_{\text{red}}(x), \tag{5}$$

subject to

$$\sum_{i=1}^{n} len(s_i)x_i \le L_{\max}, \tag{6}$$

$$x_i\in\{0,1\},\ i=1,\ldots,n, \tag{7}$$

where x_i is a variable which is defined as follows:

$$x_i=\begin{cases}1 \text{ if } s_i \text{ is selected}\\ 0 \text{ otherwise}\end{cases} \tag{8}$$

$L_{\max}$ is the predefined length of the summary, $len(s_i)$ is the length of sentence s_i; $f_{\text{cov}}(x)$ and $f_{\text{red}}(x)$ are the coverage and redundancy functions, respectively.

In Eq. (5), a parameter $\lambda\in[0,1]$ combines the two objectives $f_{\text{cov}}(x)$ and $f_{\text{red}}(x)$ into a scalar objective $f(x)$. The weighting parameter λ specifies the relative contributions of the $f_{\text{cov}}(x)$ and $f_{\text{red}}(x)$ functions to the objective function $f(x)$.

We define the objective functions $f_{\text{cov}}(x)$ and $f_{\text{red}}(x)$ as follows:

$$f_{\text{cov}}(x)=\sum_{i=1}^{n} Score(s_i)x_i, \tag{9}$$

$$f_{\text{red}}(x)=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(1-sim(s_i,s_j))x_i x_j. \tag{10}$$

The Eq. (6) guarantees that length of the created summary will be upper-bounded. The aim of the model (5)-(7) is to find the binary assignment $\mathrm{X}=\{x_i\}$ (Eq.(7)) with the best content coverage ($f_{\text{cov}}(x)$) and less redundancy ($f_{\text{red}}(x)$) such that the summary length is at most $L_{\max}$ (Eq.(6)).

In Eq. (9), $Score(s_i)$ denotes the score of sentence s_i, which is defined as follows [10]:

$$Score(s_i)=\sum_{j=1}^{I_i}\mathrm{M}(s_j)\times\exp\left(-\left(\frac{1-sim(s_i,s_j)}{\sigma}\right)^2\right), \tag{11}$$

where $i=1,\ldots,n$; s_j is a sentence having semantic similarity with the sentence s_i, i.e. $sim(s_i,s_j)\neq 0$; and I_i is the number of sentences which have semantic similarity with the sentence s_i, $1\le I_i\le n-1$; $\mathrm{M}(s_j)$ is the "mass" of sentence s_j; generally speaking, it is set to 1; σ is a parameter used to control the influence scope of sentence. The optimal value of σ can be obtained by the method proposed in [11].

Higher value of $Score(s_i)$ corresponds to higher content coverage of sentences s_i. In Eq. (10), higher value of $(1-sim(s_i,s_j))$ corresponds to lower overlap in content between sentences s_i and s_j. In other words, higher value of the function (10) provides minimum redundancy, i.e. high diversity in the summary.

In Eq. (11), the sentence "mass" $\mathrm{M}(s_j)$ is an important parameter, which directly affects the value of $Score(s_i)$. In this study, the "mass" of any sentence is defined as follows:

$$\mathrm{M}(s_i)=\frac{sim(s_i,O)}{\sum_{j=1}^{n} sim(s_l,O)},\ i=1,\ldots,n, \tag{12}$$

where O is the center of $\mathbf{S}=\{s_1,\ldots,s_n\}$.

The l th coordinate o_l of the center O calculate as:

$$o_l=\frac{1}{n}\sum_{i=1}^{n} w_{il},\quad l=1,\ldots,m. \tag{13}$$

IV. Human Learning Optimization Algorithm

This section describes HLO algorithm which is utilized for solving the optimization problem (5)-(7).

HLO algorithm adopts the binary-coding framework in which each bit corresponds to a basic component of knowledge for solving problems [12–14]. The binary-coding framework is adopted in HLO, and consequently an individual is represented by a binary string as follows:

$$Y_p=[y_{p1},y_{p2},\ldots,y_{pn}], \tag{14}$$

where $y_{pi}\in\{0,1\}$, $p=1,2,\ldots,N$, $i=1,2,\ldots,n$.

Y_p denotes the pth individual, N is the size of population, and n is the dimension of solutions.

Each bit y_{pi} of solutions Y_p stands for a basic element of the knowledge or skill that humans need to learn and master, which is initialized as ''0'' or ''1'' randomly as there is no prior-knowledge for problems initially.

An initial population is filled by randomly generated N individuals:

$$\mathbf{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \\ \vdots \\ Y_N \end{bmatrix}=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1i} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2i} & \cdots & y_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pi} & \cdots & y_{pn} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{Ni} & \cdots & y_{Nn} \end{bmatrix} \quad (15)$$

For generating candidate solutions and finding the optimal solution HLO algorithm applies the random exploration, the individual, and the social learning operators [12–14] which are explained below.

Random exploration *learning operator* (RELO). When individuals start learning, they usually learn at random as there is no prior knowledge of problem. At the next stage of learning, due to interference, forgetting, disturbance, only knowing partial knowledge of problem and other factors, humans cannot fully replicate previous experience and therefore they still learn with certain randomness. For imitation of these phenomena, HLO algorithm performs the random learning operator [12–14]:

$$y_{pi}=\text{RELO}\,(0,1)=\begin{cases}0 & \text{if } 0\le \text{rand}\,()\le 0.5\\ 1 & \text{else}\end{cases} \quad (16)$$

where rand() is a random number lying between 0 and 1.

Individual learning operator. At individual learning stage, for avoiding mistakes and improving performance, humans usually use their own experience and knowledge. For imitation of this process, HLO applies the individual learning operator. This operator, according to the best solutions stored in the individual knowledge database (IKD) generates new alternative solutions as follows [12–14]:

$$y_{pi}=ikd_{pki} \quad (17)$$

$$ikd_p=\begin{bmatrix} ikd_{p1} \\ ikd_{p2} \\ \vdots \\ ikd_{pk} \\ \vdots \\ ikd_{pK} \end{bmatrix}=\begin{bmatrix} ikd_{p11} & ikd_{p12} & \cdots & ikd_{p1i} & \cdots & ikd_{p1n} \\ ikd_{p21} & ikd_{p22} & \cdots & ikd_{p2i} & \cdots & ikd_{p2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ ikd_{pk1} & ikd_{pk2} & \cdots & ikd_{pki} & \cdots & ikd_{pkn} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ ikd_{pK1} & ikd_{pK2} & \cdots & ikd_{pKi} & \cdots & ikd_{pKn} \end{bmatrix} \quad (18)$$

Where $1\le k\le K$; ikd_p represents the individual knowledge of p th individual, Y_p; K is the pre-defined number of solutions that saved in the IKD, ikd_{pk} represents the k th best solution of individual Y_p, and k is a random integer, determines which individual in the IKD is adopted for individual learning.

Social learning operator. Humans usually directly or indirectly transfer knowledge and skills and therefore the efficiency and effectiveness of learning can be significantly improved by sharing experience. To model of this mechanism HLO algorithm performs the social learning operator. This operator yields new candidate solutions as follows [12–14]:

$$y_{pi}=skd_{ki} \quad (19)$$

$$SKD=\begin{bmatrix} skd_1 \\ skd_2 \\ \vdots \\ skd_s \\ \vdots \\ skd_S \end{bmatrix}=\begin{bmatrix} skd_{11} & skd_{12} & \cdots & skd_{1i} & \cdots & skd_{1n} \\ skd_{21} & skd_{22} & \cdots & skd_{2i} & \cdots & skd_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ skd_{s1} & skd_{s2} & \cdots & skd_{si} & \cdots & skd_{sn} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ skd_{S1} & skd_{S1} & \cdots & skd_{Si} & \cdots & skd_{Sn} \end{bmatrix} \quad (20)$$

Where the SKD is the social knowledge database which is used reserve the best knowledge of the population; $1\le s\le S$; S is the size of the SKD, and skd_s is the s th solution in the SKD.

Implementation of HLO. New solutions are generated by the random learning operator, individual learning operator, and social learning operator in HLO, which can be simplified and formulated as follows [12–14]:

$$y_{pi}=\begin{cases}\text{RELO}\,(0,1), & \text{if } 0\le \text{rand}()\le pr\\ ikd_{pki}, & \text{if } pr\le \text{rand}()\le pi\\ skd_{si}, & \text{else}\end{cases} \quad (21)$$

where pr, (pr - pi) and (1- pi) are the probabilities of performing of random learning, individual learning and social learning, respectively.

Based on experiences [12] the probabilities pr and pi set as: $pr=\frac{5}{n}$ and $pi=0.85+\frac{2}{n}$, where n is the dimensionality of the solution.

Fitness function and constraint handling. Since (5)-(7) is the constrained optimization problem, then HLO algorithm cannot be directly applied to solve it. Therefore, for handling the constraint, i.e. to deal with infeasible solutions of which the total length of the selected sentences exceeds the pre-specified limit, the penalty function method is applied [4]. This method applies an algorithm for unconstrained optimizations to the penalty function formulations of the constrained problems [15]. The penalized objective function is defined as:

$$F(x)=f(x)+\beta\cdot\ln\left(L_{\max}-\sum_{i=1}^{n}len(s_i)x_i\right) \quad (22)$$

where $\beta>0$ is the penalty parameter.

In Eq. (22), the first term is the objective function defined by the Eqs. (5), (9) and (10) and the second is the penalty term. As seen from the Eq. (22), the function $F(x)$ is undefined if $L_{\max}-\sum_{i=1}^{n}len(s_i)x_i\le 0$. Therefore, the penalized function $F(x)$ is valid only for the feasible solutions such that the constraint is strictly satisfied, $L_{\max}-\sum_{i=1}^{n}len(s_i)x_i>0$. However, even for the feasible solutions the penalty term is

non-zero, but it becomes an "anti-penalty" if $L_{\max} - \sum_{i=1}^{n} len(s_i)x_i \geq 1$.

In this study, the penalty parameter β is dependent on the current generation:

$$\beta = \beta^{-} + (\beta^{+} - \beta^{-})\frac{t}{t_{\max}}, \quad (23)$$

where t is the current generation, $t_{\max}$ is the maximum number of generations, β^{-} and β^{+} are start and the end values of penalty parameter β, $0 < \beta^{-} < \beta^{+} \leq 0.5$.

Termination criterion. The HLO algorithm terminates when the maximum number of generation $t_{\max}$ is met.

REFERENCES

[1] H.P. Luhn, "The automatic creation of literature abstracts", IBM Journal of Research and Development, vol.2, no.2, pp.159–165, 1958.

[2] R.M. Alguliyev, R.M. Aliguliyev, and N.R. Isazade, "A new similarity measure and mathematical model for text summarization", Problems of Information Technology, no.1, pp.42-53, 2015.

[3] R.M. Alguliyev, R.M. Aliguliyev, and N.R. Isazade, "Formulation of document summarization as a 0-1 nonlinear programming problem", Computers & Industrial Engineering, vol.64, no.1, pp.94-102, 2013.

[4] R.M. Alguliyev, R.M. Aliguliyev, and N.R. Isazade, "An unsupervised approach to generating generic summaries of documents", Applied Soft Computing, vol.34, pp.236-250, 2015.

[5] R.M. Alguliyev, R.M. Aliguliyev, and N.R. Isazade, "CDDS: constraint-driven document summarization models", Expert Systems with Applications, vol.40, no.2, pp.458-465, 2013.

[6] R. Tucker, "Automatic summarizing and the CLASP system", PhD thesis, University of Cambridge, 1999, 190p. http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-484.pdf

[7] K.S. Jones, "Automatic summarizing: the state of the art", Information Processing & Management, vol.43, no.6, pp.1449–1481, 2007.

[8] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey", Artificial Intelligence Review, pp.1-66, 2016.

[9] A. Nenkova and K. McKeown, "Automatic summarization", Foundations and Trends in Information Retrieval, vol.5, nos.2–3, pp.103–233, 2011.

[10] Z. Wang, Z. Chen, Y. Zhao, and S. Chen, "A community detection algorithm based on topology potential and spectral clustering", The Scientific World Journal, vol.2014, Article ID 329325, 9 pages, 2014.

[11] W.-Y. Gan, N. He, D.-Y. Li, and J.-M. Wang, "Community discovery method in networks based on topological potential", Journal of Software, vol.20, no.8, pp.2241-2254, 2009.

[12] L. Wang, H. Ni, R. Yang, M. Fei, and W. Ye, "A simple human learning optimization algorithm", LSMS/ICSEE, CCIS, 2014, Part II, vol.462, pp.56–65.

[13] L. Wang, R. Yang, H. Ni, W. Ye, M. Fei, and P.M. Pardalos, "A human learning optimization algorithm and its application to multi-dimensional knapsack problems", Applied Soft Computing, vol.34, pp.736-743, 2015.

[14] L. Wang, H. Ni, R. Yang, P.M. Pardalos, X. Du, and M. Fei, "An adaptive simplified human learning optimization algorithm", Information Sciences, vol.320, pp.126-139, 2015.

[15] Ö. Yeniay, "Penalty function methods for constrained optimization with genetic algorithms", Mathematical and Computational Applications, vol.10, no.1, pp.45-56, 2005.

Authorship Identification of the Azerbaijani Texts Using n-grams

Aida-zade K.R.
Numerical decision-making methods in deterministic systems
Institute of Control Systems of Azerbaijan NAS
Baku, Azerbaijan
kamil_ayda-zade@rambler.ru,

Talibov S.Q
Numerical decision-making methods in deterministic systems
Institute of Control Systems of Azerbaijan NAS
Baku, Azerbaijan
saxavat@yahoo.com

***Abstract*—The purpose of this study is to show how n-grams are used for author recognition in the Azerbaijani language. As attribute vectors for analyzing of authorship are taken monogram and digram. We have developed a new approach to the determination of the attribute vectors for recognition of the author of an unknown text.**

***Index Terms*—digram, monogram, auther audentification, recognition.**

I. Introduction

It is known that one of the most important problems of text processing is their classification by the authors, ie, determining which one of a predetermined group of authors is the alleged author of this particular text.

Automation solutions to this problem began to develop most intensively in the 70s of the last century. Initially, the methods to solve this problem based on the use of specialized glossaries created for keywords.

Mosteller [1] was one of the first to use Bayesian analysis in order to solve the authorship recognition problem. Then Tan [2] and Doğan S., & Diri B. [6] used the N-gram (2-gram and 3 grams) recognition of authorship in Turkish. In [3], a new method is introduced for the reduction of the dimension by attribute extraction and evaluation of its impact on text classification.

Primary research in Azerbaijan investigated the frequency of the use of letters and word length by the authors [4] - [5] for recognition of authorship of Azerbaijani text, but nevertheless, there is not a computer system to recognize text attribution in the Azerbaijani language yet. In this paper we study the problem of identification of authorship based on the analysis of opinion articles of small size. The main difficulty in the recognition of authorship of the text of small size in the Azerbaijani language is that, the large number of uninformative suffixes, endings are used in words, and the automatic analysis of words into their component parts for the Azerbaijani language. These difficulties have not been solved yet.

II. Statement of The Problem

Formally setting texts authorship identification tasks can be described as follows. The database contains some texts of n authors, each of them has m_i number of texts of D_{ij}, j = 1, ..., m_i, i = 1, ..., n. The class (group) text of the i-th author is denoted by Y_i. The goal of this article is to determine the auther of a new entered text D among the n authers, in other words to which class Y_i it belongs. Each of the texts of D_{ij}, and D is associated with a set of values attributes

$$M_{i,j}^s,\ s = 1, \dots, k_i, \text{ and } d_s,\ s = 1, \dots, K, k = \min_{1 \le i \le n} k_i,$$
$$j = 1 \dots, \quad m_i, \quad i = 1, \dots, n$$

on the basis of which there is a classification of texts by authors where k_i - number of features to determine the authorship of i-th author $i = 1, \dots, n$.

We introduce the following notations, definitions and formulas:

$M_{i,j}^s$ - denote the value of s-th feature of j-th article by i-th author, $N_{i,j}$ − is the length (size) j-th text of i-th author, m_i- is the number of i-th article(text) of the author. Then it is clear that the average value of the s-th feature of i-author in the j-th article is determined by the formula

$$\varepsilon_{i,j}^s = \frac{M_{i,j}^s}{N_{i,j}},\ j = 1, \dots, m_i,\ s = 1, \dots, k_i, i = 1, \dots, n, \quad (1)$$

the average value of s-th feature in all the articles of the i-th author equel

$$\xi_i^s = \frac{\sum_{j=1}^{m_i} M_{i,j}^s}{\sum_{j=1}^{m_i} N_{i,j}},\ s = 1, \dots, k_i, i = 1, \dots, n. \quad (2)$$

Let the average value of the s-th feature to the new article D, equel

$$x_D^s = \frac{m_D^s}{N_D}, \quad s = 1, \dots, k. \quad (3)$$

Here m_D^s- the s-th feature value in the new article D, and N_D - the length (volume).

It is obvious that the variance of s-th feature for the i-th author is

$$(d_i^s)^2 = \frac{\sum_{j=1}^{m_i} (M_{i,j-}^s \xi_i^s)^2}{\sum_{j=1}^{m_i} N_{i,j}}, s = 1, \dots, k_i,\ \ i = 1, \dots, n. \quad (4)$$

Variation of s-th feature to the i-th author is

$$v_i^s = \frac{d_i^s * 100}{\xi_i^s}, s = 1, \dots, k_i, \quad i = 1, \dots, n. \qquad (5)$$

Consider the value of

$$R_i = \sum_{s=1}^{k} \alpha_s (x_D^s - \xi_i^s)^2, i = 1, \dots, n, \qquad (6)$$

determining the proximity of (normal) values of attributes of the new text D to the values of features that characterize the i-th author; α_s- is the weight (importance) of the s-th feature to determine the authorship of articles.

III. The Used Methods and Algorithms for Text Authorship Recognition

Algorithms of functioning of authorship identification systems generally include executing of the following series of processes:

• Held initial processing of existing texts (articles, works) of different authors and, for each author defined numerical values of the selected attributes;

• The analysis of the characteristic values and the determined set of informative features for each author (the set of attributes may not be the same for different authors);

• determines the characteristic values represented by the new article by an unknown author;

• According to specific criteria based on known algorithms defined by the presumed author of the present article.

Attributes of authorship, based on statistical analysis of letter combinations.

Note that many well-known algorithms and text authorship recognition system used attributes, based on the analysis of the use of different combinations of letters of the n - letters, called in the literature of n-grams. Thus, in this case,"gram" means that a unit is taken as one letter, wherein words, sentences or paragraphs entire text depending on the combinations of letters are divided into n comprising n successive letters, n = 1, 2

Note that the number of letters and, consequently, 1 gram per 32 Azeri and practical number of possible 2-gram equal to 835. The following three algorithms use features based on n-grams.

Algorithm 1 (using monograms)

Step 1. For each of the text (articles) of the i-th author included in the class Y_i, as determined by the attributes of the frequency of the use of all the letters of the alphabet (1 gram).

Step 2: Combining all the articels of each author in accordance with the formula (2) calculated the average values of all the attributes.

Step 3. To study new article D by the formula (3) is calculated feature vector values x_D^s, s = 1, ..., k.

Step 4. In the formula (6) taking $\alpha_s = 1$, (all weights equal to 1) define such ν that $R_\nu = \min\limits_{1 \le i \le n} R_i$, therefore the author of article D is ν-th author.

Algorithm 2 (using digrams)

It is the same as monogramm, but instead of using the frequency of individual letters, it uses the frequency of various combinations of the two letters of the alphabet (digram), used by the author in their articles.

Algorithm 3 (**modified algorithm 1**)

The basic idea of the algorithm based on the modified monogram is that it uses sustainable features, not including uncharacteristic atributes of the author.

The proposed algorithm for each author by using the formula (5) is calculated variation of each characteristic. Weight s-th attributes α_s by the formula (6) are selected depending on the variation of s-th feature to all the authors in the following manner. Let us dinoted by:

$$v^s = \begin{cases} \min\limits_i v_i^s, & \min\limits_i v_i^s > \varepsilon \\ \varepsilon, & \min\limits_i v_i^s \le \varepsilon, \end{cases} \quad s = 1,2, \dots, k$$

Pozitiv number ε selected on the basis of uncharacteristic attributes of authors. Then

$$\propto_s = \frac{\frac{1}{v^s}}{\sum_{j=1}^{k} \frac{1}{v^s}}, \quad s = 1,2, \dots, k.$$

It is clear that , $s = 1,2, \dots, k$ satisfy the conditions:

$$0 \le \propto_s \le 1 \quad , \qquad s = 1,2, \dots, k, \quad \sum_{j=1}^{k} \propto_s = 1 \propto_s$$

The first two steps of the proposed algorithm with the same first two steps of the algorithm 1.

Step 3. Using (4) and (5) for each feature of class Y_i based on variation is checked stability and set values of the weights α_s.

Step 4. To study the new article according to the formula D (3) the calculated attributes value, x_D^s , s=1, …, k.

Step 5: Determine ν at which $R_\nu = \min\limits_i R_i$ (therefore the author of the article is ν-th author).

IV. The Results of Computer Experiments

Assume that the training data set consists of 50 newspaper articles by four authors, conditionally called the A1, A2, A3, and A4. As the test data are taken two articles of each author for identifying the authors and compare the performance of the above algorithms.

As an indication in the case of monogram (n = 1) were used 32 letters and in the case digram (n= 2) were used for the 835 really possible combinations of letters of the Azerbaijani language.

13 newspaper articles for author A1, 11 newspaper articles for author A2, 12 newspaper articles for author A3 and 14 newspaper articles for author A4 were taken respectively. The total number of letters in the examined articles is from 3438 to 6859 range.

To recognize the i-th author, we take 2 articles $z^i = (z_1^i, z_2^i)$ for each author. letters for table footnotes.

TABLE I. The Resualt of Alqorithm 1

	A1	A2	A3	A4
z_1^1	**2262**	2682	2676	2379
z_2^1	2573	2542	**2181**	2701
z_1^2	2937	2363	2426	**2272**

z_2^2	2563	**2243**	2642	2549
z_1^3	**2257**	2495	2454	2792
z_2^3	**2103**	2514	2333	3048
z_1^4	2477	2646	2662	**2213**
z_2^4	2819	2335	2580	**2264**

TABLE II. THE RESUALT OF ALQORITHM 3

	A1	A2	A3	A4
z_1^1	2268	2699	2584	2447
z_2^1	2195	2591	2403	2809
z_1^2	2853	**2292**	2541	2312
z_2^2	2747	**2245**	2500	2506
z_1^3	**2176**	2515	2411	2896
z_2^3	**2184**	2282	2396	3136
z_1^4	2575	2577	2611	**2236**
z_2^4	2829	2296	2779	**2095**

TABLE III. THE RESUALT OF ALQORITHM 2

	A1	A2	A3	A4
z_1^1	**231**	253	247	267
z_2^1	**224**	257	247	270
z_1^2	269	**226**	265	238
z_2^2	261	**232**	260	245
z_1^3	231	257	**224**	286
z_2^3	**226**	232	260	245
z_1^4	275	252	**234**	285
z_2^4	263	235	285	**204**

Table 1 shows the results of the algorithm 1, which uses features based on monograms. In the crossing cell (the i-th row and j-th column) given the value $R_i(z^j) * 10^3$. It is clear that correct recognition of the i-th authorship of the article corresponds to the case that the value of j-th diagonal element is the smallest among the elements of the j-line.

As can be seen from Table 1, it was correctly installed only authorship of the articles of the author of the fifth, and the quality of recognition was 50%.

In the Table 2 shows the results of the proposed modified algorithm based on the monogram, and used for attributes the weights.

As seen from Table 2, the recognition efficiency is significantly improved and 75%.

In the Table 3 shows the results of the algorithm 2, using digrams as attributes.

The obtained recognition results are more stable, i.e. $R_\nu = \min_{i \le i \le n} R_i$, and substantially greater than R_ν for i $\neq \nu$ R_j, and the recognition efficiency is 80%.

As can be seen from table 1, table 2 and table 3, the use of algorithms based of monogram, and modified monogram algorithm, inefficiently for the recognition of the author. The algorithm is based on a monogram identifies the author only in 62.5% cases and the modified monogram algorithm in 75% of cases correctly.

As can be seen from Table 2 using digrams allows determining the author no more than 75% accuracy.

The relatively low percentage of recognition of the authors in the two versions is mainly due to large dimension of the space characteristics and the proximity characteristic values to each other. This is due to the impossibility of separating the plurality of features (attributes) that characterize each of the authors of the feature space, the usual linear hypersurfaces. On the other hand, it is also due to the small size (volume) of records in newspaper articles and low information content of these articles.

Note that the use of modified algorithms based of monogram and digram first proposed by us. The aim is to choose the parameters, forming a more robust system. The use of this idea may be rationally applied in other algorithms for the recognition and classification.

CONCLUSION

In the paper, we have analyzed the methods of authorship recognition of Azerbaijani texts.

The main features of the analyzed methods are using n-grams with n = 1 and n = 2. The recognition algorithms are built using a statistical approach. As a research object, we have taken the recognition of the authors of low volume newspaper articles.

In the article, we have proposed a modified algorithm of the statistical pattern recognition method using the magnitude of variations of n-grams typical for articles of the considered authors.

Results of experiments on the recognition of authors of articles using the developed algorithms have been given.

REFERENCES

[1] G. Mosteller, F., Wallace, D. L., Applied Bayesian and Classical Inference: The Case of the Federalist Papers. Reading, MA: Addison-Wesley, 1984.

[2] Tan, C. M., Wang, Y. F., Lee, C. D., The Use of Bigrams to Enhance, Journal Information Processing and Management, Vol.30-4, 2002, pp.529-546.

[3] Biricik, G., Diri, B. and Sönmez, A: C., "A New Method for Attribute Extraction with Application on Text Classification", 5th International Conference on Soft Computing, Computing with Words, ICSCCW, North Cyprus, Famagusta, 2009

[4] Aida-zade K.R., Talibov S.G Analysis of the effectiveness of the methods of recognition of authorship of texts in the Azerbaijani language. The 5th International Conference on Control and Optimization with Industrial Applications-COIA-2015, pp.27-29, August, 2015 Baku, Azerbaijan.

[5] Sardar Gasimov, Ibrahim Ibrahimov. ANALYSIS OF SENTENCES AND WORDS USED IN AZERBAIJANI TEXTS The Second International Conference "Problems of Cybernetics and Informatics" September 10-12, 2008, Baku, Azerbaijan

[6] Doğan, S., & Diri, B. (2010). Türkçe Dokümanlar için N-gram Tabanlı Yeni Bir Sınıflandırma: Yazar, Tür ve Cinsiyet. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 3, pp.11-20.

Sentiment Analysis for Bank Service Quality: a Rule-based Classifier

Yuliya Bidulya
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
bidulya@yandex.ru

Elena Brunova
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
egbrunova@mail.ru

Abstract. **The paper considers the analysis of the subjective information from user-generated content. The purpose of this research is to develop a rule-based classifier for the sentiment analysis within the bank service quality domain. Our sentiment lexicon includes 286 positive and 385 negative words. Besides, three more lexicon classes are added; they are required to apply the rule-based algorithm. To test the algorithm, 200 reviews in Russian are analyzed. The experiment demonstrates that the efficiency of the rule-based classifier is higher as compared to the Naïve Bayes classifier. It is determined that the system generally detects positive reviews better than negative ones.**

Key words: **sentiment analysis, natural language processing, rule-based classifier, algorithm, user-generated content, lexicon.**

I. Introduction

People are always interested in the opinions of their families, friends, etc. The manufacturers and service providers are interested in the market expectations while the clients before choosing products or services are eager to know the opinions of other clients. Now instead of asking a friend or an expert, people often first visit the manufacturer's site or just type the name of the product in a search engine to read the reviews. Thus, the investigation in the field of sentiment analysis is one of the rapidly developing areas of applied linguistics.

The specific features of the domain under study may sufficiently affect the sentiment analysis application. First of all, some words from sentiment lexicons appear domain-specific [3], e.g. the word *long* can be included into the positive lexicon when evaluating the battery operation for the smartphone domain but it belongs to the negative lexicon when evaluating the client's time consuming for the bank service quality domain. These are parametrical words, i.e. the words denoting the amount of some domain-specific parameters [2]. However, the specific features may be observed not only on the lexical level but on the sentence and the text ones. A review of a product (a car, a camera, etc.) is generally a description, with target aspects to be rather explicit. A service review is often a narration where the author outlines facts and events which evoke positive or negative emotions. In this case, the target aspects are implicit and hard to be extracted. A review on bank service quality is an example of this second category.

Thus, to extract opinions from reviews on bank service quality, we propose the rule-based approach which implies the use of lexical and syntactic structures along with the positive and negative lexicons. These structures determined within a sentence are considered as the factors affecting its polarity.

The purpose of this research is to develop a rule-based classifier for the sentiment analysis within the bank service quality domain.

II. Related Work

Sentiment analysis is one of the rapidly developing methods of natural language processing. The first works were published in the early 2000s [8, 9, 11], and since then much has been done in this field. Sentiment lexicons have been built; algorithms have been developed [3, 5, 6].

In the sentiment analysis, two main approaches can be distinguished: the machine learning approach and the lexicon approach. The machine learning approach generally uses the supervised classifier [7]. Machine learning usually starts with forming the training dataset which is later used to train the classifier. Finally, the classifier is tested on the test dataset. Within the lexicon approach, the polarity is determined in the process of matching the review text with the positive and negative classes of the sentiment lexicon [1].

The performance of the sentiment analysis within the lexicon approach can be sufficiently increased with rule-based classifiers, which is particularly observed for rather narrow domains. Thus, a rule-based algorithm for sentiment analysis was used to classify financial news in [10]. The positivity / negativity ratio of each sentence was determined and the sentiment values of the whole text were calculated.

III. Dataset and Sentiment Lexicon

The data for research are client reviews in Russian on bank service quality from the bank client rating taken from the site *www.banki.ru*. These reviews are rated by the author from 5 (very good) to 1 (very poor). However, this rating can be approved or rejected by the moderator. The reviews should be informative, so the authors are expected to give as much details and arguments of their opinion as possible and to outline some facts and events. As a result, the typical review text from our dataset is a narration consisting of several sentences, sometimes of several passages. The shortest review from our dataset comprises three sentences. This is the essential difference of these data from such short messages as tweets where the review text often consists of a single sentence.

For the experiment 200 reviews were randomly selected according to the length criteria: min review length is 50 tokens, max review length is 500 tokens. The author's opinion is assumed positive if the review is rated 4 or 5, and it is assumed negative if the review is rated 2 or 1. The reviews rated 3, the reviews without a rating, and the reviews for which the author's rating has not been approved by the moderator are not considered.

To build the sentiment lexicon, the seed lexicon containing 100 words was built manually from 50 reviews (24 positive ones and 26 negative ones) which were randomly selected. Then the seed lexicon was extended, using synonyms, antonyms, and the sentiment consistency technique [5]. This technique first proposed in [4] uses the list of seed opinion adjective words and the set of linguistic constraints (*and, but, either-or, neither-nor*) to identify other opinion words and their polarity. For instance, in the sentence *This i-phone is beautiful and easy to use*, if *beautiful* is known to be positive, it can be inferred that *easy* is also positive. On the contrary, in the sentence *This i-phone is beautiful, but expensive*, if *beautiful* is known to be positive, it can be inferred that *expensive* is negative. As a result, the sentiment lexicon consists of 286 positive and 385 negative words.

IV. Rule-based Algorithm for Sentiment Analysis

We developed the set of rules which are applied to determine the final sentence polarity. The review polarity is determined as the sum of sentence polarities. Every rule can be presented as a pattern to extract specific lexical and syntactical structures from the sentences.

The review polarity determination includes preprocessing, calculating draft polarity of each sentence, correcting the polarity of each sentence through the application of the rules, summing the corrected polarities, and normalizing the final polarity of the document.

Preprocessing is arranged as below:

Step 1. To break the text into sentences, marking the beginning and the end of each sentence.

Step 2. To break each sentence into fragments, substituting the punctuation marks within the sentence, quotation marks, and the conjunctions.

Step 3. To break the text into words using spaces.

Step 4. To count the number of words in the text.

Step 5. To mark the words typed in capital (upper-case) letters.

Step 6. To substitute all the upper-case letters for lower-case ones.

Step 7. To stem all the words.

The draft polarity of each sentence is calculated after all the words from the positive and negative lexicons are substituted by the non-terminal polarity symbols (POS or NEG, respectively).

To correct the draft polarity and to apply the rules, the corrective lexicon is manually constructed. It includes three classes: polarity modifiers, polarity increments, and polarity anti-modifiers. Polarity modifiers are the words changing the polarity of the words within a fragment into the opposite one. Polarity increments are the words increasing the polarity count of the sentence. Polarity anti-modifiers are the words canceling the change the polarity of the words within a fragment into an opposite one though the fragment contains a polarity modifier.

A fragment is understood as a part of a sentence between the beginning of the sentence, or a punctuation mark, or a conjunction (and/or) and the next punctuation mark, or a conjunction (and/or), or the end of the sentence.

The non-terminal symbols for the corrective lexicon are presented in Table 1.

TABLE 1. The non-terminal symbols

Corrective Lexicon	***Examples***	***Non-terminal symbol***
Polarity modifiers	*не, нет (not, no)*	ALT
	без (without)	WT
	? ?! (a question mark, a combination of a question mark and an exclamation mark)	QM
	" " (quotation marks)	Q
Polarity increments	*очень, совершенно, никогда, нигде* (very, absolutely, never, nowhere)	INC
	a capitalized word	CAP
	! (an exclamation mark)	EM
Polarity anti-modifiers	*так, такой* (so, such)	TH
Other symbols	a sentence	S
	a fragment	F
	a conjunction *и* (and) or *или* (or)	CJ
	, : ; – (a comma, a colon, a semicolon, a dash)	Z
	an unmarked word	*w*

The rules for correcting the draft polarity in a sentence are presented below. Each rule is illustrated with an example.

Rule 1.1. <ALT>(<POS>){n}→<NEG>{n}

Доверия нет! (*There is no trust!*), count=-1.

Rule 1.2. <ALT>(<NEG>){n}→<POS>{n}

Сами работники банка не грубые. (*The bank officers are not rude.*), count=1.

Rule 2.1. (<INC><NEG>) →<NEG><NEG>

Сотрудники крайне некомпетентны. (*The employees are extremely incompetent.*), count=-2.

Rule 2.2. (<INC><POS>) →<POS><POS>

Большое спасибо сотрудникам банка (*Many thanks to the bank officers!*), count=2.

Rule 3.1. (<TH><ALT><NEG>)→<NEG>

Меня еще никогда так не обманывали. (*I was never cheated like this.*), count=-2.

Rule 3.2. (<TH><ALT><POS>)→ <POS>

Такой оперативности я еще не видел. (*I have not met such an efficient response.*), count=1.

Rule 4. <POS>{1}<QM>→<NEG>
Просто? (*Easy?*), count=-1.

Rule 5. <QM>→<NEG>
Чего теперь ждать-то? (*What shall I wait for?*), count=-1.

Rule 6. <Q><POS><Q>→<NEG>
«Надежный» банк. (*A "reliable" bank.*), count=-1.

Rule 7.1. <WT>(*w*|INC){0,}<POS>→<NEG>
Мое обращение оставили без внимания. (*My application was ignored.*), count=-1.

Rule 7.2. <WT>(*w*|INC){0,}<NEG>→<POS>
Это банк без неприятных сюрпризов. (*This is a bank without unpleasant surprises.*), count=1.

Rule 8.1. <POS><EM>→<POS><POS>
Спасибо Сбербанку! (*Thanks to Sberbank!*) count=2

Rule 8.2. <NEG><EM>→<NEG><NEG>
Я в шоке! (*I am shocked!*), count=-2.

Rule 9.1. <POS><CAP>→<POS><POS>
ХОРОШО, когда есть такие сотрудники. (*It is GOOD when there are such employees.*), count=2.

Rule 9.2. <NEG><CAP>→<NEG><NEG>
ОТВРАТИТЕЛЬНЫЙ сервис. (*It is an AWFUL service.*), count=-2.

The rules are applied in a certain order. The algorithm operation results in summing the corrected counts of all the sentences and normalizing to the number of words in the review. At the end, the system gives a conclusion ("Positive" or "Negative").

V. Results and Discussion

The results of the experiment were verified with the *Precision*, *Recall*, *F*-measure values.

Precision and *Recall* can be calculated [6] using Eq.1, Eq. 2, Eq. 3 and Eq. 4:

$$Precision(pos) = a/(a+b), \quad (1)$$

$$Precision(neg) = d/(d+c), \quad (2)$$

$$Recall(pos) = a/(a+c), \quad (3)$$

$$Recall(neg) = d/(d+b), \quad (4)$$

where *a* and *b* are true positive and true negative values for positive polarity prediction while *c* and *d* are true positive and true negative values for negative polarity prediction.

F-measure is calculated using Eq.5:

$$F=2*Precision*Recall/(Precision+Recall). \quad (5)$$

The efficiency of the proposed rule-based classifier was compared with the unigram-learnt Naïve Bayes classifier [12]; manually built sentiment lexicon was applied. The efficiency of the two classifiers is presented in Table 2.

TABLE 2. Results of Experiment

Polarity	Methods	
	Rule-based	***Naïve Bayes***
Positive		
Precision	0.82	0.75
Recall	0.93	0.85
F-measure (Positive)	0.88	0.80
Negative		
Precision	0.93	0.88
Recall	0.77	0.71
F-measure (Negative)	0.84	0.79
Average F-measure	**0.86**	**0.80**

As it can be seen from Table 2, the average *F*-measure value of our rule-based classifier is higher than the average F-measure value of the Naïve Bayes classifier, 0.86 and 0.80 respectively.

The *F*-measure values demonstrate that for both classifiers, the efficiency for positive reviews is higher than for negative ones. It means that the system generally detects positive reviews better than negative ones. While expressing their opinions about the service quality people consider positive opinions as a norm which is understood as something organized and standard. Thus, an author expressing positive opinions and emotions often uses standard, clichéd words and phrases. On the contrary, negative emotions are considered as anomalies which can be numerous and diverse. An author expressing negative emotions tends to use various words and syntactical structures. This feature can be also observed in the sentiment lexicon we extracted where the number of positive words is higher than the number of negative ones, 385 and 286 respectively.

VI. Conclusions and Future Work

We developed a rule-based algorithm for sentiment analysis which included nine rules applied in a specific order. These rules and the corresponding syntactic models simplify each sentence and present the text of a review as a formal model.

The dataset of 200 client reviews in Russian on bank service quality was used to test the algorithm. The experiment demonstrated that for the same set of reviews, the *Precision*, *Recall*, and *F-measure* values for the proposed technique were 5-8% higher as compared to the Naïve Bayes Classifier. This means that the use of lexical and syntactic structures along with positive and negative lexicon improves the efficiency of sentiment analysis. This is especially important for reviews that are rather narrative than descriptive.

We applied the rule-based approach to one domain and left its expansion to other domains for the future. Another promising way for our future experiments is to apply a machine learning approach for building sentiment lexicon when the words from the manually built lexicon can be used as seed words.

REFERENCES

[1] Annett, M., Kondrak G. (2008) "A comparison of sentiment analysis techniques: Polarizing movie blogs", Advances in Artificial Intelligence, Springer Berlin Heidelberg, pp. 25-35.

[2] Brunova, E. (2013) "Parametrical words in the sentiment lexicon", (IJCRSEE) International Journal of Cognitive Research in Science, Engineering and Education. Vol. 1, No.2, available at: http://www.ijcrsee.com

[3] Ganapathibhotla, M., Liu B. (2008) "Mining opinions in comparative sentences", Proceedings of the 22nd International Conference on Computational Linguistics. Manchester, pp. 241–248.

[4] Hatzivassiloglou V., McKeown K. (1997) "Predicting the semantic orientation of adjectives", Proceedings of the 35th Annual Meeting of ACL, Madrid, pp. 174-181.

[5] Liu, B. (2010) Sentiment analysis and subjectivity. Handbook of Natural Language Processing, available at: http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf.

[6] Manning C., Raghavan P., Schütze H. (2008) Introduction to information retrieval. Cambridge University Press, Cambridge.

[7] Medhat, W., Hassan, A., Korashy, H. (2014). "Sentiment analysis algorithms and applications: a survey", Ain Shams Engineering Journal, 5(4), pp. 1093-1113.

[8] Nasukawa T., Yi J., (2003) "Sentiment analysis: capturing favorability using natural language processing", Proceedings of the 2nd International Conference on Knowledge Capture. Florida, pp. 70-77.

[9] Pang B., Lee L., Vaithyanathan S. (2002) "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of EMNLP, available at: http://www.cs.cornell.edu/people/pabo/papers/sentiment.pdf.

[10] Tan Li Im, Phang Wai San, Chin Kim On, Patricia Anthony (2015) "Rule-based sentiment analysis for financial news", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, pp. 1601-1606.

[11] Turney P. (2002) "Thumbs up or thumbs down? Semantic orientation applied to supervised classification of reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp.417-424.

[12] Webb G., Boughton J., Wang Z. (2005) "Not So Naive Bayes: Aggregating One-Dependence Estimators", Machine Learning, 58, pp. 5-24.

Object Detection and Segmentation Using Data from Time-Of-Flight Cameras

Timur Luguev
Department of Discrete Mathematics and Informatics
Dagestan State University
Makhachkala, Russia

Aida Razakova
Department of Discrete Mathematics and Informatics
Dagestan State University
Makhachkala, Russia

***Abstract*— The problems of object detection on an image and labeling of pixels, corresponding to each detected object (segmentation) are solved. In the work we have examined such issues as filtering and preparation depth data, extraction of semantically rich features vectors from the RGB-D images and classification methods, allowing implementing the objects detection and segmentation.**

***Index Terms*—object recognition, RGB-D, image understanding, time-of-flight cameras, object segmentation, 3D scene recognition**

I. Introduction

The problem of object detection and segmentation on images is a fundamental and aims to detect object from specific category and mark each pixel that belong to it.

Object detection and segmentation problem is actively investigated now [1,2]. However, most of these studies focus on outdoor scenes, leaving more complicated case of objects segmentation indoors poorly explored. The main difficulties are that there is a need of considering a large number of possible scene variations and that low light impacts on the existing work methods [13].

The recent proliferation of available time-of-flight cameras (Microsoft Kinect, Intel RealSense 3D-camera, Asus Xtion) allows looking for significant progress in researches. In addition to the color information, time-of-flight cameras receive a scene depth map, providing so-called RGB-D images (Fig. 1).

Time of flight cameras obtain an image based on measurement of the time infrared laser light fly between sensor and every point of the object.

Taking into account information on the scene depth has greatly improved the accuracy of methods in a number of research studies (body parts estimation, 3D modeling, etc.) due to better resistance to an angle and lighting changing.

There is a significant amount of methods for object detection for RGB images. Unfortunately, most of them cannot be applied to RGB-D data.

In the present work both issues of processing of the RGB-D images and semantically rich features vectors extraction, allowing implementing objects detection and segmentation, are examined.

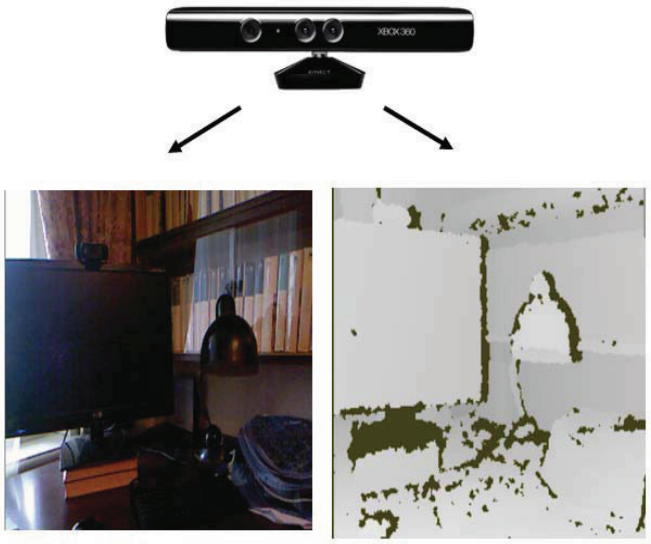


Fig.1. Color image and depth map, obtained from the Kinect sensor

II. Main idea and methods

In the present study, a common approach in the field of object recognition was used for object detection and segmentation on the RGB-D images: vectors of low-level features are extracted and coded for use by various classifiers, in particular, by support vector machines.

Our approach consists of the following steps: calibration, filtering of the depth data, encoding of the depth image, generation of the region proposal and pixel-wise segmentation (Fig. 2).

RGB-D data from Kinect. For obtaining images and depth maps, the time-of-flight Microsoft Kinect [4] camera was used.

Sensor Kinect consists of IR laser emitter, infrared camera and RGB camera. Triangulation is used for a depth measuring. Laser emitter emits one beam, which splits to few beams by diffraction grating, in order to create invariable image. Pattern of the image is obtained by capturing the plane on the known distance from the sensor and is stored in the sensor's memory.

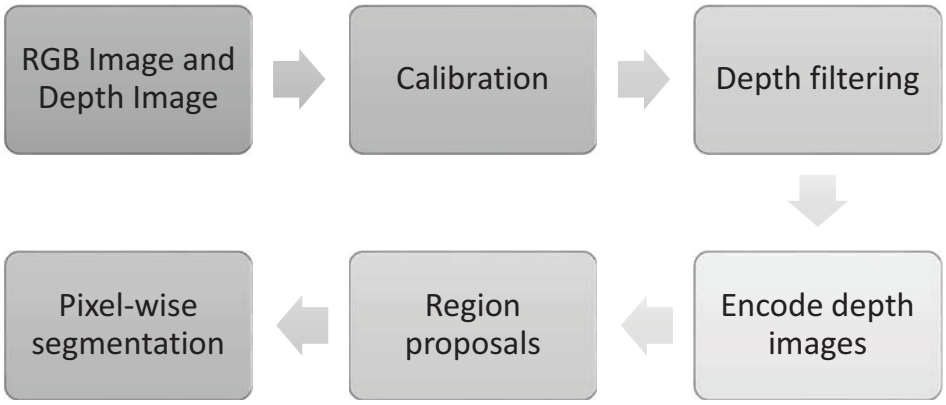


Fig.2. Object segmentation process

Suppose, the object is placed on the reference plane on the distance D_0 from the sensor. If object will be moved closer or further from the sensor, object position on the image will be changed along X axis. Such change correspond to the following equations:

$$\frac{Q}{b} = \frac{(D_0 - D_k)}{D_0} \quad (1)$$

and

$$\frac{d}{f} = \frac{Q}{D_k} \quad (2)$$

where D_k – depth of the point k in the object dimension, l – length of the base, f – focus distance of the IR-camera, Q – offset of the k in the object dimension, d – visible gap in the image dimension. Substituting Q from (2) to (1) we can get D_k:

$$D_k = \frac{D_0}{\left(1 + \frac{D_0}{fl} d\right)} \quad (3)$$

Equation (3) is a basic mathematical model for depth calculation if parameters D_0, f and l could be calibrated.

The Kinect sensor supports the operating range of distances from 0.5 to 3 meters and has a field of view of approximately 50 angular degrees in both directions.

Using the Kinect sensor for getting basic data for recognition has brought a number of benefits, including:

1) invariance with respect to position;

2) invariance with respect to illumination.

There are a number of difficulties when using the time-of-flight cameras that need to be overcome: lower resolution of the depth map and poor optics quality for RGB images (Fig. 1).

Calibration. For obtaining higher RGB images quality, the Kinect was synchronized with a high resolution camera. Depth maps calibration with high-resolution color images was carried out by means of searching locally invariant singular points, as it was performed in [5].

Depth filtering. When working with the depth map, which resolution is lower than that of the color images calibrated therewith, there are many pixels that do not have depth values. For smoothing and interpolation of missing depth values, approximation was performed [6].

Depth map contains noise and areas that don't contain data. Such areas caused by shadows that objects casts in infrared light of the Kinect sensors.

In order to improve quality of the depth map we uses method to exclude noise on the image and fill empty areas with data, which represent actual depth values with some accuracy.

The distance between infrared emitter and infrared camera can cause a shadow on the side of the object, which is closer to camera than emitter is.

For solving this problem for each pixel inside area without depth replaced with average of closest pixels from neighboring sides.

Encode depth images. The basis for a recognition system creation is extraction of features vectors from the RGB-D images. The used approach consists in calculation of a pixel attributes in a small window around it, in particular, gradient directions and magnitudes for a vicinity of 5x5 pixels. For computational optimization and representation convenience, the features vectors dimension is reduced by a principal component analysis (PCA) method [7].

Region proposals. Region proposal can be understood as an assumed object hypothesis. Many methods for proposal generation were created previously. We used MCG [8] because it outperformes others like Selective Search [9].

Region proposal generation algorithm begins with computation of segmentation hierarchy for various resolutions. At the next step, this hierarch is fused into one multiscale hierarchy. Single scale hierarchies are grouped with multiscale hierarchy in order to produce region proposals. Lastly, proposals are graded based on size, shape, position.

We generated 500 region proposals per image.

For each proposed window our method classify every pixel with a random forest classifier [10].

After calculating prediction for all pixels in the window smoothing step was made by averaging over superpixels.

For the received features vectors, a support vector machines (SVM) method is applied [11] that allows achieving detection and segmentation of objects on the RGB-D image.

Fig.3. Object segmentation results

III. Results

Testing of effectiveness of the obtained objects detection and segmentation methods was made on the RGB-D standard data set – RGB-D Object Dataset [12], containing images of various objects, as well as information on a depth map for each image. It was divided into training and testing part and contains densely labeled pairs of aligned RGB and depth images for 300 common household objects. The objects are categorized according to WordNet.

We used RGB-D Object Dataset and NYUD2 [13] to train our model. To generate additional data for training we used augmentation with synthetic data. We rendered 3D models of objects into the background and were able significantly increase our training database.

The results of the implemented method on testing images for three different objects are shown in the Fig. 3.

TABLE I. RESULTS ON THE TEST DATA

Object	RGB only	RGB-D
Bowl	0.46	0.74
Mug	0.48	0.68
Can	0.21	0.76

Recall values for some of the object categories from the test part of database are presented in the Table 1. This results were obtained for 0.5 precision.

IV. Conclusion

The result of our work shows significant improvement in detection and segmentation results when depth data used along with RGB image.

The implemented approach to object detection and segmentation can find application in the areas of computer vision, human-computer interaction, construction of intellectual video surveillance systems. The common distribution of time-of-flight cameras for both desktop computers and mobile devices (tablets, smartphones) reinforces relevance of the results obtained by the RGB-D images analysis.

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN X GPU used for this research.

This work was supported by Minobrnauki of Russian Federation.

References

[1] D.K. Prasad. Survey of The Problem of Object Detection In Real Images // International Journal of Image Processing, Vol. 6. I. 6. 2012. pp.441-446.

[2] S. Shantaiya, K. Verma and K. Mehta. A Survey on Approaches of Object Detection. International Journal of Computer Applications 65(18):14-20, 2013.

[3] P. Espinace, T. Kollar, A. Soto, and N. Roy, “Indoor scene recognition through object detection,” in IEEE International Conference on Robotics and Automation (ICRA), 2010.

[4] Zhengyou Zhang, Microsoft Kinect Sensor and Its Effect // IEEE Multimedia, Vol. 19, No. 2, pp. 4-10, 2012.

[5] T.S. Luguev, I.V. Luguev. Analysis of facial motions by means of RGB-D data // Pattern Recognition and Image Analysis. 2015. Vol. 25, I. 3, pp. 466-469.

[6] T. Luguev et al. "Some computational methods of bivariate approximation by discrete Chebyshev polynomials", Book of abstracts of 2nd Dolomites Workshop on Constructive Approximation and Applications, Alba di Canazei, Italy, 2009.

[7] I.T. Jolliffe. Principal Component Analysis // Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28.

[8] J. Pont-Tuset, P. Arbeláez, J. Barron, F. Marques, and J. Malik . Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016.

[9] Jasper R. R. Uijlings, Koen E. A. van de Sande, Theo Gevers, Arnold W. M. Smeulders. Selective Search for Object Recognition. International Journal of Computer Vision, Volume 104 (2), page 154-171, 2013.

[10] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends in Computer Graphics and Vision, 7:81–227, 2011.

[11] C. Cortes, V. Vapnik. Support-vector networks// Machine Learning. 1995. 20 (3). pp 273-297.

[12] K. Lai, L. Bo, X. Ren, and D. Fox. A Large-Scale Hierarchical Multi-View RGB-D Object Dataset // IEEE International Conference on Robotics and Automation (ICRA), May 2011. pp. 1817-1824.

[13] Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. ECCV 2012.

Heuristic approach to model of corporate knowledge construction in information and analytical systems

Victoria Bova
Department of CAD
Southern federal university
Rostov-on-Don, Russia
vvbova@yandex.ru

Vladimir Kureichik
Department of CAD
Southern federal university
Rostov-on-Don, Russia
vkur@sfedu.ru

Daria Zaruba
Department of CAD
Southern federal university
Rostov-on-Don, Russia
daria.zaruba@gmail.com

***Abstract* — Concerning intelligent information and analytical systems one of the most prospective areas is the construction of knowledge bases used ontological systematization as a tool for classification of corporate knowledge. The authors interpret a competence model, which can select significant features of classifiable objects, in terms of domain ontology. To classify corporate objects it is suggested a heuristic method of knowledge clusterization in multidimensional feature space in which a genetic algorithm is used to obtain effective solutions for classification procedure according to well-known criteria. The genetic algorithm is an iterative probabilistic search algorithm whose main feature is simultaneous using of a set of population from the space of potential solutions. A certain advantage of the method is guaranteed lack of intersections for all clusters and necessary to define the number of clusters. Experimental results were carried out on the basis of test tasks and confirmed a theoretical relevance and promising of the suggested method.**

***Index Terms*— Competencies models, classification and clusterization of knowledge, intelligent information and analytical systems, genetic algorithms, ontologies models, knowledge management systems**

I. INTRODUCTION

Today, a knowledge management is a powerful tool for organization focused on permanent changes of business processes. Its efficiency is defined by transformation of information into knowledge [1-4]. To knowledge management in corporate information and analytical systems (CIAS) the development of unified information environment for support a life cycle of organization knowledge and comprehensive knowledge model is actual and promising problem.

There are a lot of scientific works involves knowledge management systems (KMS), but the most of all focuses on the development of instrumental means and technologies supporting functions and services to process information and knowledge [1, 3-6]. There is very few experimental results deal with scientific substantiation of corporate strategies of knowledge management which can involve not only explicit but implicit knowledge in organization business processes. In the context of the modern economic implicit knowledge is a strategic and intelligent resource of organization but is not frequently used for knowledge management in CIAS [6-8].

One of the key problems here is an identification of specialists' competence, its formalization and creation of correspondence knowledge base (KB) in KMS. In this connection, the actual problem is the development of a knowledge management model based on heuristic methods of description and classification of specialists' competence profiles in multidimensional space of corporate competence features in CIAS.

II. THE PROBLEM FORMULATION

The use of a competence approach dictates by the need for creation of structural and information models in the KMS in which the main component is a KB of competence models. A lot of prosperous organization uses a competence model for integration the tactical work with special means of requirements unifying to competence of specialists [2, 9-11]. There are three main tasks solved with the use of competencies model in the KMS (Fig. 1).

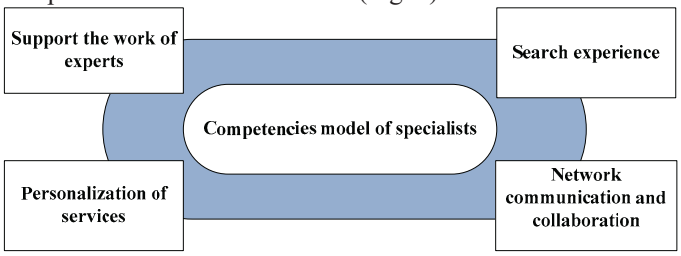


Fig. 1. Tasks solved with the use of specialist models in the KMS.

Competency is the term of the subject domain which have a behavior in space formed by thee vectors [10]: 1) human (personality) qualities, formed under the influence of activity, but extends to the whole person; 2) content of the activity, which demonstrate personality qualities; 3) assessment of activities quality. A competence behavior is a consistent changing in its properties through a mutual influence of personal qualities and maintenance activities. Thus, the description of requirements to the specialists' behavior with the use of competency modeling, its description and identification of the behavior, is a tool for processing of implicit knowledge within a corporate strategy of knowledge management [4-6].

A corporate knowledge model involves following elements (Fig. 2):

- competency clusters - sets of closely related corporate competencies;
- taxonomy of corporate competencies;
- profiles of specialists' competencies for specific user roles to solve practical problems of knowledge management;
- indicators of behavior.

One of the most important tasks of the model creation is to provide semantic functionality (interoperability) not only as part of the KMS model (Enterprise Information System) but also for collaboration with external models (branch and corporate information systems). The most promising approach for creation of corporate knowledge models is a semantic technology [12-14].

Ontology modeling techniques allows to specify a competency model [10,13] and provides an opportunity to use this model to solve a set of application tasks in CIAS. An ontology model of a corporate knowledge subject contains a glossary of competence taxonomy, a corporate knowledge model in projection of specific competencies and semantic metadata of a specialists' competence profile.

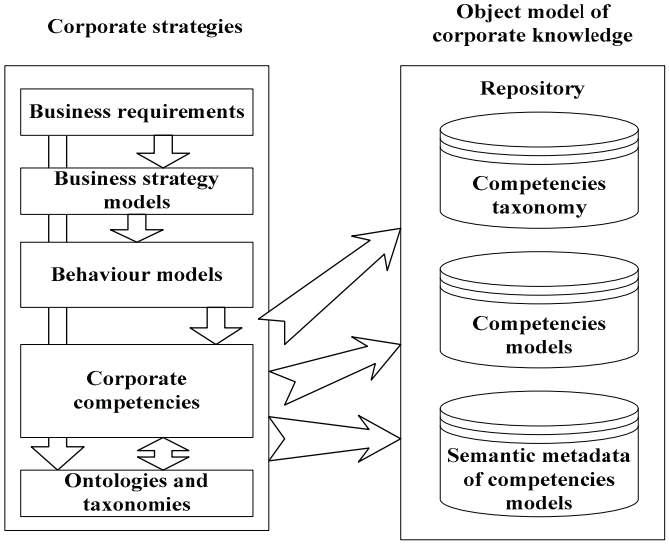


Fig. 2. Model of corporate knowledge representation.

A model of specialist's competency (Fig. 3) is represented as $MS = \{O,K,P\}$, where O is a general specialist's characteristics; K are indexes of specialist's competence in the relevant fields of knowledge; P is a characteristics of specialists' behavior in KMS.

General specialist's characteristics include declarative data about the user.

The specialist's competency $K = \{K_o, K_c\}$ involves indexes of specialist's competence (qualification, skills, experience) in knowledge fields K_o and its semantic description K_c on the basis of the knowledge field ontology; $K_o = \{(O_1, k_1), \dots, (O_n, k_n)\}$, where (O_i, k_i) is a knowledge field i and a level of specialist's competency in the knowledge field; K_c is a description of a specialist as a set of semantic metadata.

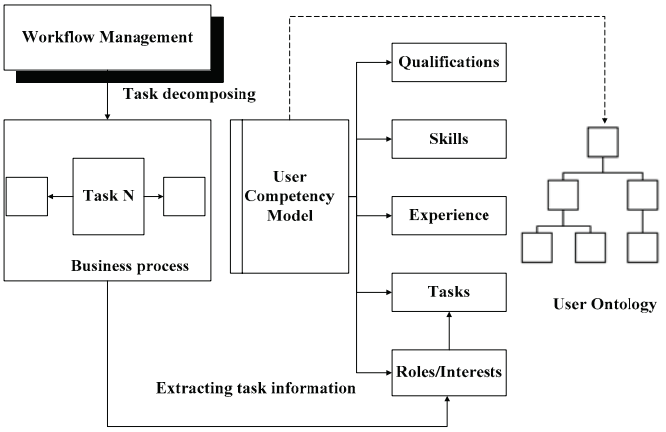


Fig. 3. Model of specialist's competence profiles

Semantic metadata are data describing a context and/or a content of an object within knowledge base KMS with the use of domain concepts defined by web ontology language [13]. Semantic metadata for the specialist S_i are defined as a finite set of ordered pairs (c_{ij},k_{ij}), i.e. $MetaD(s_i) = \{(c_{i1},k_{i1}), \dots, (c_{in},k_{in})\}$, where $c_{in} \in C$ is an ontology concept referring to the description subject s_i; $k_{in} \in [0,1]$ is a relevancy of the concept c_{in} to the subject s_i.

A specialist's behavior in KMS is a concept describing interconnection of specialists and KMS. Firstly, the authors denote a specialist's behavior by two parameters $P = \{R_i, Z_i\}$, where R_i is a role for performing specific tests of tasks corresponding to specialist's qualification; Z_i are practical tasks within business process KMS.

Simulation of competence profiles and creation of knowledge base KMS has a significant value to increase effectiveness of knowledge management in the CIAS [2, 5].

III. Analysis of Problem

Currently, creation of corporate models of competencies has a critical role for human resource management in modern enterprises [9]. Because of, the classification problem can be reduced to the clusterization problem. Competencies clusters are competencies models combines into groups on the basis of intention [11]. Each competence is a set of relative behavioral indicators united into blocks depending on semantic scope of competence.

The clusterization problem is to define groups of objects that are closest to one another according to some criterion. This problem determines the optimal number of clusters and its centers and boundaries. To formulate the clusterization problem initial data are parameters values and a method of decision making about the instance belonging to a particular class in conditions of uncertainty, i.e. under the influence of random factors which maximize a connection between features and the instance class [3]. Hence, heuristic algorithms for the clusterization problem solving are the most promising direction in the development of up-to-date artificial intelligent [15].

Clusterization can be considered as an optimal partitioning of objects into groups [7]. Optimality is maximization of

clusters density or minimization of mean-square distance between clusters center and other subjects:

$$f = \sum_{l=1}^{k} \sum_{i \in S_l} d^2(X_i, \bar{X}_l) \quad (1)$$

where l is a number of cluster ($l = 1,2,...k$), $\bar{X}$ is a center of the cluster l, X is a vector of variable values for *i-th* subject belonging to the cluster l, $d\ (X_i, \bar{X}_l)$ is a distance between *i-th* subject and a center of the cluster l.

Let us formalize the problem formulation. A user's competence model is represented as U_i=<V_i, Z_i, Ki>, where $V_i\ (u) = \{(t_1, w_1(u),(t_2, w_2(u),..., (t_n, w_n(u)\}$ is vector space of specialist's knowledge model including t_i – a competence term in the model, w_i is a weight corresponding to t_i. $R_i\ (u) = \{(z_1, q_1(u),(z_2, q_2(u),...,(z_m, q_m(u)\}$ are users' roles for solving specific problems z_i supporting by business processes, q_i is a competence indicator for supporting the specific role, K_i are clusters of competence model.

Let $V_i(u)$ be a point in n-dimension term space. For each term is calculate a weight w_i, i.e. for each t_i we can define its coordinate containing frequency of competency terms in user's role. In this case coordinate axes are terms. Their number is determined by a number of terms which are weighted, i.e. each t_i in the role R_i corresponds to non negative weight w_i, q_i is an indicator of competence to support specific role.

$$w_i = \frac{\delta_i \times \log(\frac{N}{n_i} + 0.01)}{\sqrt{\sum_{j=1}^{n} \left[\delta_i \times \log(\frac{N}{n_i} + 0.01)\right]^2}}, \quad (2)$$

where δ_i is a coefficient of proportionality to the frequency of competence term t_i as a user, N is a total number of tasks with t_i for the role R_i, n_i is a number of task z_i in which required competence t_i.

$$g_i = \sqrt{\sum_{t=1}^{n} (1 - Y_t(C_i^j))^2}, \quad (3)$$

where $Y_t(C_i^j) = K_t^i, B_t^j$ are expert assessments of u_i user according to the t_i competence. A vector $K_t^i = (K_1^i, \ldots, K_t^i, \ldots, K_n^i)$, $K_t^i \in [0,1]$ is a control parameter to define importance of competence t_i for the role R_i. A vector $B_t^j = (B_1^j, \ldots, B_t^j, \ldots, B_n^j)$, $B_t^j \in [0,1]$ is a control parameter to define score obtaining as a result of expert questionnaires u_i of user according to the t_i competence.

The competence term is a point within n-dimension space. Each cluster contains terms. Let assume that a center of cluster #1 is the first term. Next, it is calculated a sum of distance between the cluster and others terms containing in this cluster. Then, assume that that the second term is the center of cluster #1. There is calculated a distance to them. At the same way there are calculated all distances between each terms contained in the chromosome and in the cluster #1. As a result a center of cluster is a subject with minimum sum of distance to the others term in the cluster.

The fitness function (1) is a multimodal (multiextremal) function. For this reason an adaptive method of random search with the use of genetic algorithm (GA) allows to obtain quasi-optimal solution in polynomial time [8, 15]. Genetic search is a sequential transformation of one finite set of alternative solutions to another using genetic and natural mechanisms and principles [13].

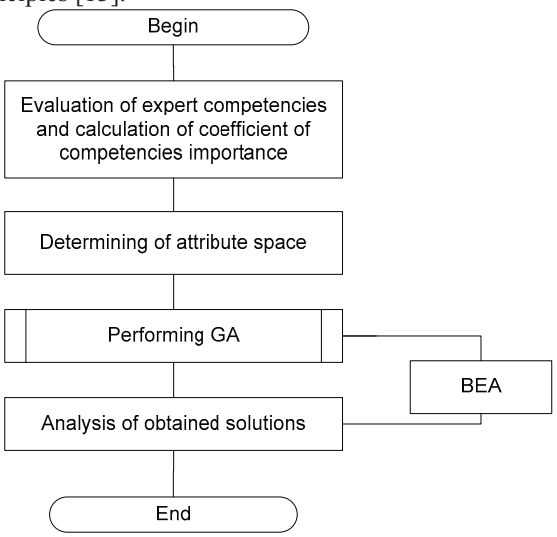


Fig. 4. The scheme of genetic search

The main aim of the paper is a development of a heuristic clusterization method based on the using of GA and meet the following requirements:

1. Method has to work with high dimensional metric space of features.
2. Developed algorithm has to be easy implemented and uses a minimum quantity of computational resources.
3. User's participation in the clusterization process must be minimized.

IV. Adaptation Method in Genetic Algorithm

Here we consider a strategy of GA adaptation to the clusterization problem. Hypothesis of this research is that the speed of global extremum search may be increased by reinforcement of population changeability and intensification of a leader selection.

Let us consider GA with a population in which each chromosome is an alternative solution of the given problem. In this case a solution is a partitioning of unordered set of competencies into clusters which define a competence profile and a relevant list of behavioral indicators (roles) for successful corporate knowledge management.

To solve the clusterization problem by GA following elements of algorithms need to denote:

- the mechanism of solutions encoding (chromosome);
- assessment function for each chromosome;
- selection, recombination and mutation operator;
- stop criterion;
- probabilistic parameters of evolution convergence.

A chromosome is an array (vector of specialist's competencies, behavior – role, cluster). An array length is equal to a number of corporate competencies partiting into clusters. This information is represented as a competence identifier and number of cluster in competence model database (Fig. 2).

A set of alternative solutions P generates on the basis of chromosome initialization. A chromosome contains values of threshold coefficients of the competence model for users U_i.

Each chromosome is evaluates by its fitness-function (FF) which determined as a sum of Euclidean distances between each term and center of the cluster and calculated as follows:

$$f = \sum_{l=1}^{k} \sqrt{\sum_{i=1}^{n} (\bar{X}_l - X_i)^2}, \quad (4)$$

where $\bar{X}$ is a center of cluster l, X_i is a coordinate of *i-th* term of the competence, k is a number of term determining a length of the chromosome, n is a number of coordinate axes.

The most adapted chromosomes have the most opportunity to produce offsprings. After the calculation of the FF for each chromosome, they are ordered according to its values. At top of list there are chromosomes with large values, at the end of list – chromosomes with low values. Next, a linearly decreasing function of the random number is used for the selection of potential parents [11].

Than for each chromosome a proportional selection assigns a probability *P(i)* which equal to dependence of its adaptation from sum adaptation of all population.

The next stage at the GA is a crossover operator. Here we use a multipoint crossover operator. A break point is a boundary between neighbor elements in an array. Parent chromosomes are broke down into segments at these points. After that corresponding segments from parents are united into offsprings.

Than a mutation operation is performed as the exchange of two random numbers of clusters. Numbers of corresponding competencies t_i are selected by a random way.

As a result we obtain an alternative solution (chromosome). To stop the algorithm there are analyzed obtained solutions by evaluation of its identity CS_i according to the density criterion cp_i (compaction measure – intracluster distance) and separability sp_i (intercluster distance).

Groups of obtained solutions can contain points within area of the same extremum. To compare solutions according to these criteria there are performed a following solutions:

$$sp_i = \min_l \left[\sum_{l_i, l_j \in L_k} d(x_i, x_j)^2 \right] \quad (5)$$

where $d(x_i, x_j)$ is a distance between neighbor points in clusters l_i and l_j.

$$cp_i = \max_l \left[\sum_{i,j \in L_k} d(x_i, x_j)^2 \right] \quad (6)$$

where $d(x_i, x_j)$ is a distance between крайними points in the cluster L_k.

$$CS_i = sp_i \times cp_i, \quad (7)$$

So, "the worst" cluster has the lowest value of CS_i.

For the purpose of a feedback an evolutionary adaptation (EA) is added in the search architecture. This process manages an adjustment of GA parameters and is based on interconnection with environment (decision making person) [14-16]. The evolutionary adaptation influences on reorganization of a current population of alternative solutions and creation of a new population.

V. Experimental results

A software implementation of the suggested method allows us to conduct computational tests on the basis of small and middle dimension problems. To compare the developed clusterization algorithm and k-means algorithm we use a mean-squared distance between the center of cluster and others subjects (4), the minimum intracluster distance (5) and the maximum intercluster distance (6). In these algorithms it is applied the same metrics of feature space and the fitness function.

The comparison of clusterization results for small and middle dimension problems by GA and k-means algorithms are shown on Fig. 5.

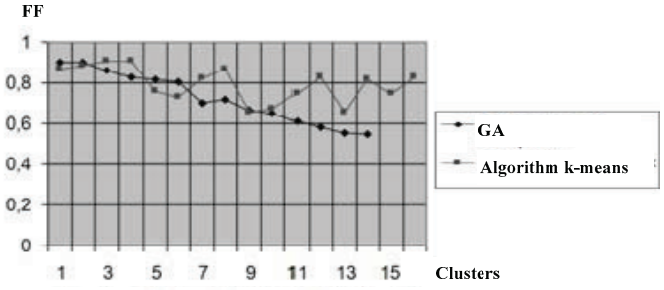


Fig. 5. The comparison of clusterization results for small and middle dimension problems

Increasing a number of clusters the GA has an advantage of the k-means algorithm. The results are shown on Fig. 6.

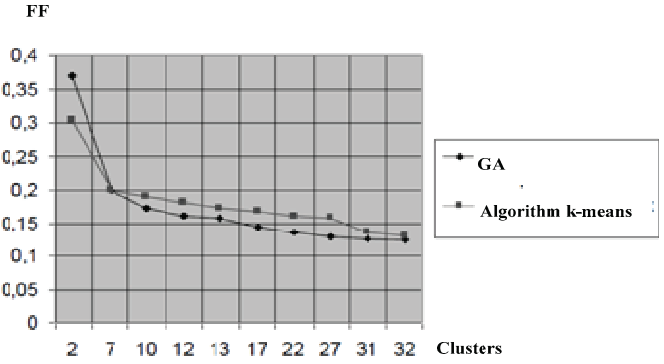


Fig. 5. The comparison of clusterization results for high dimension problems

Furthermore, the suggested method provides a clusters' separability in all coordinates that importance for recognition patterns problems. Note, the GA, as a tool for optimization problems solving, is not only accelerate a search process but allow to avoid stagnation of global search.

VI. Conclusion

In the paper it was considered a promising direction in knowledge base CIAS with the use of ontology systematization as a tool of subject classification of corporate knowledge. The authors suggest the competencies model interpreted in ontology domain terms and may select significant features if classified objects. To solve the corporate knowledge classification problem it was suggested the heuristic method of knowledge subject clusterization within multidimensional

feature space based on the GA which allows to obtain effective solutions for classification procedure. An advantage of the method is an absence of intersection for each clusters. Experimental results according to three criteria are demonstrated at test tasks. Also experiments have confirm a theoretical relevance and promising of suggested method.

ACKNOWLEDGMENT

The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

REFERENCES

[1] V. V. Bova, V. V. Kureichik, and A. A. Lezhebokov, "The integrated model of representation of problem-oriented knowledge in information systems," in *8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 – Conference Proceedings,* 2014. pp. 111-115.

[2] J. Koh, and Y. G. Kim "Knowledge sharing in virtual communities: an e-business perspective," in *Expert Systems with Applications*, vol. 26, 2004, pp. 155-166.

[3] S. I. Rodzin, and L. S. Rodzina "Search for optimal solutions and its application for the processing of problem-oriented knowledge," in *8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 – Conference Proceedings,* 2014. pp. 142–146.

[4] V. V. Kureichik, Y. A. Kravchenko, and V. V. Bova, "Decision support systems for knowledge management," in *4th International Conference on Computer Science On-line Conference, CSOC 2015* vol. 349, ed: Springer Verlag, 2015, pp. 123-130.

[5] A. Chua "Knowledge management system architecture: a bridge between KM consultants and technologists," in *International Journal of Information Management*, 2004, vol. 24(1), pp. 87- 98.

[6] Y. A. Kravchenko, and V. V. Kureichik, "Knowledge management based on multi-agent simulation in informational systems," in *8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014*, 2014.

[7] E.V. Kuliev, A.N Dukkardt, V.V. Kureichik, and A.A. Legebokov, "Neighborhood Research Approach in Swarm Intelligence for Solving the Optimization Problems," in *12th IEEE East-West Design & Test Symposium, EWDTS 2014,* 2014. pp. 112-115.

[8] V. V. Bova, A. A. Lezhebokov, and L. A. Gladkov "Problem-oriented algorithms of solutions search based on the methods of swarm intelligence," in *World Applied Sciences Journal*, 2013, vol. 27, pp. 1201-1205.

[9] B.Ya.Shvedin "Enterprise Ontology:expiriontological approach: Technology of ontological model creation", Moscow, Lenand, 2010, p. 240.

[10] A.I. Subetto, "Ontology and epistemology of competencies approach, classification and qualimetry of competencies", Mockow, Research Center of training quality problems, 2006, p. 72.

[11] S. I. Rodzin, and O. N. Rodzina "New computational models for big data and optimization," in *9th IEEE International Conference on Application of Information and Communication Technologies, AICT 2015 – Conference Proceedings,* 2014. pp. 3–7.

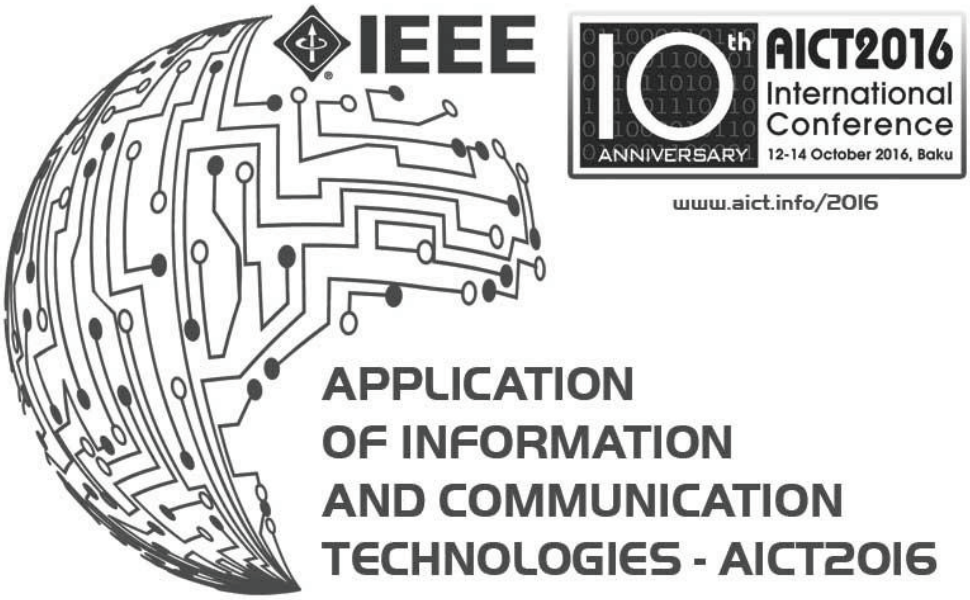
[12] V.V. Bova, V.V. Kureichik, and A.A. Legebokov, "Integration of ontologies in scope of model and conceptual semantics: modified approach", in *9th IEEE International Conference "Application of Information and Communication Technologies–AICT 2015,* 2015, pp. 156-160.

[13] V. V. Bova, Y. A. Kravchenko, and V. V. Kureichik, "Development of distributed information systems: Ontological approach," in *Advances in Intelligent Systems and Computing* vol. 349, ed, 2015, pp. 113-122.

[14] E.V. Kuliev, A. A. Lezhebokov, and B. Shkalenko "A New Approach for Software Development in Terms of Problem-Oriented Knowledge Search and Processing," in *Intelligent Information Technologies for Industry, IITI-2016*, 2016, vol. 1, pp. 297-306.

[15] D. Yu. Zaporozhets, D. V. Zaruba, and V.V. Kureichik, "Hybrid bionic algorithms for solving problems of parametric optimization," *World Applied Sciences Journal*, vol. 23, 2013, pp. 1032-1036.

[16] D. U. Zaporozhets, D. V. Zaruba, and V. V. Kureichik, "Representation of solutions in genetic VLSI placement algorithms," in *12th IEEE East-West Design and Test Symposium, EWDTS 2014,* 2014. pp. 1-4.



CONFERENCE PROCEEDINGS

SESSION 3.

CYBER SECURITY ISSUES

12-14 October 2016, Baku, Azerbaijan

Extension of access control policy in secure role-based workflow model

Ivan Tarkhanov
Federal Research Center "Computer Science and Control" of RAS
State Academic University for Humanities
Moscow, Russia
itarkhanov@isa.ru

Abstract — **Development of Task Based Access Control models have been recently progressing in workflow management systems. This paper describes the extension of the well-known secure role-based workflow model (RBWM) which allows flexible management of permissions to system objects through tasks. Nonetheless, RBWM does not address the issues of versioning of securable objects and dependence on the state of linked objects. Here we suggest addressing these problems with access control policies. We propose extension of security policy by additional set of permissions, schemes and algorithms allowing simultaneous execution of these policies. The research covers the implementation of this model in a running system and comparison with existing approaches.**

Index Terms — **access control policy, WFMS, RBWM, explicit permission.**

I. Introduction

Extensive usage of workflow management systems (WFMS) on different enterprises and industries resulted in a variety of approaches towards its design and implementation. One of the most challenging problems is that of access control in WFMS. Usually this challenge is met by management of task permissions which constitute central information object, irrespective of algorithms and notations used to describe workflow and business processes [1, 2, 3]. Task Based Access Control Model describes flexible change of task permissions [4, 5]. Other approaches are focused on the links between objects in WFMS including hierarchic links where management is executed through different delegations and inheritances [6, 7, 8]. Some works [9, 10], including the latest research of the author [11] are dedicated to the management of access control through policy algebra and algorithms of sharing policies. The model described in the present paper is based on the well-known Secure Role Based Workflow Model (RBWM) [5] and is extended by additional access control policy properties which allow to address the following issues:

- Management of versioning of securable objects. In WFMS new files are uploaded, documents are created, workflow processes are initiated. When users of the system perform certain tasks these objects are being changed. As a result, new versions of securable objects are created. Each version has its author, a set of permissions to authorized actions which depend on its state and the state of linked objects. Access control policy management for different versions of the objects is not elaborated enough and often is not taken into account in the existing access control policy models.
- Consideration of linked objects state. An object can simultaneously be part of several processes. For example, a contract is being created within one process in WFMS, while at the same time a financial manager is calculating the financial terms and conditions to the same contract. In this case there is a need of real time consideration of the object state. For instance, to disable the editing of object properties if the linked task has certain state.

The paper aims to offer the extension of the well-known RBWM model which allows to solve access control problem for different object versions and takes into account the state of linked objects in WFMS during permission check.

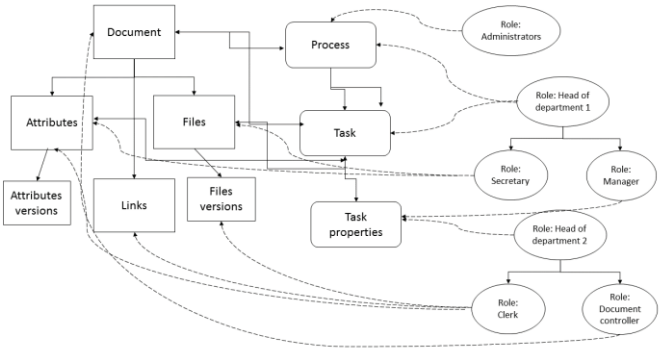


Fig. 1. Interaction of securable objects, roles and users in WFMS «Euphrates».

II. Extension of Secure Role-Based Workflow Model in WFMS

A. Securable object instance

In RBWM the terms task instance, explicit permissions and implicit permissions are used [5]. However, this set of terms is not enough to address the challenges described herein. In our model we introduced the term object instance.

Object instance is $oi \in OI$,

$oi = \{Attr', CurrentState, TI, EP, IP\}$, where

- $a' \in Attr', a \in Attr, a \to a'$, a' - is a version of an attribute (property, file) oi, created due to the change of a.
- $CurrentState(oi) = \{s \mid s \in S\}$. This is a set of object state defined by the states of attributes, properties and files of the object, and their versions defined by the state of tasks linked to the object.
- OO – is a set of WFMS objects, mapping of object instances $\delta: OO \to 2^{OI}$ - every document is mapped against the set of its instances, where $\delta(a) \cap \delta(b) = \emptyset$, if $a \neq b$ and $a, b \in OO$.
- TI – is a set of task instances linked to oi, the same as in RBWM. TT – is a set of tasks, τ – is the mapping of every task against the set of its instances. $\tau: TT \to 2^{TI}$, where $\tau(a) \cap \tau(b) = \emptyset$, if $a \neq b$ and $a, b \in TT$.
  - EP = EPT ∪ EPO, is a set of all explicit permissions, where
  - EPO = OPO ×OO – is the set of explicit permissions for performance of certain operations (OPO) on the set of objects (OO), versions of its attributes, files.
  - EPT = OPT ×TT – is the set of explicit permissions for performance of certain operations (OPT) on the tasks (TT).
- IP = IPT ∪ IPO – is a set of all implicit permissions, where
  - IPO = OPO × OI – is the set of implicit permissions for performance of certain operations (OPO) on the object instances (OI).
  - IPT = OPT ×TI – is the set of implicit permissions for performance of certain operations (OPT) on the task instances (TI).

The difference between explicit and implicit permissions should be noted. Explicit permissions are set before the object is created and processed within a workflow. Implicit permissions are introduced when there is already the object instance at place and instances of the tasks linked to it are created. This configuration of permissions allows to set initial state of permissions to a securable object and its versions, and then to change the permissions both directly and through tasks during the lifecycle of the object in WFMS.

B. Access control policy to the objects in WF

Usually at the description of access control policy the standard definition of the policy is used [10] - policy = <subject, object, operation> (subject – a subject seeking access, object – a securable object, operation – an operation performed by the subject on the object). In recent research papers the basic definition of the policy is often extended by additional set of permissions and constraints [9, 10]. In order to address the issues of complexity and variability of objects we suggest extending the definition of the policy by a specific type of permissions (both explicit and implicit) and scope of the permissions.

Access control policy to the object is

$policy = < subject, object_{all,my}, op, type >$, where

- *subject* – assignment (a user or a role).
- *object* – object or versions, attributes, properties and files constituent to it (a', f'). To consider versioning of secured objects in the access control policy we introduce additional scope labels according to which the permission is applied to the versions of child objects. In the present model these permissions have two types of labels: *(all, my)* – where permission $object_{all}$ is applied to all versions of the securable object, while $object_{my}$ is applied only to the versions authored by the current user.
- $op \in OP$ is an operation which can be performed on the *object*, including the creation, deletion or change of permissions.
- $type: (p \in EPO \mid p \in IPO)$ is a type of a permission described in the current access control policy, either it is explicit or implicit, belonging to *EPO* or *IPO* correspondingly.

Two types of permissions suggest the issue of their sharing. What happens with implicit permissions when explicit permissions take effect? What happens with implicit permissions when explicit permissions expire? To neutralize conflicts consequent to the performance of explicit and implicit permissions, we introduce the term access control policy sharing schema. The following sharing schemas are possible:

- When implicit permissions take effect explicit permissions:
  - Are cancelled: $P_{EPO} = \emptyset$.
  - Remain in force, are combined: $[P_{EPO}] \cup [P_{IPO}]$.
  - Are replaced by read permissions: $P_{EPO}\,\hat{}\,read = P_{EPO} \setminus \{(subject, object, read, type), read \in OP, read \neq o\}$
- When implicit permissions expire explicit permissions:
  - Are restored $P_{EPO} = P_{EPO0}$.
  - Are not restored $P_{EPO} = \emptyset$.
- When implicit permissions expire, they:
  - Are cancelled: $P_{IPO} = \emptyset$.
  - Remain in force: $[P_{EPO}] \cup [P_{IPO}]$.
  - Are replaced by read permissions: $P_{IPO}\,\hat{}\,read = P_{IPO} \setminus \{(subject, object, read, type), read \in OP, read \neq o\}$

.

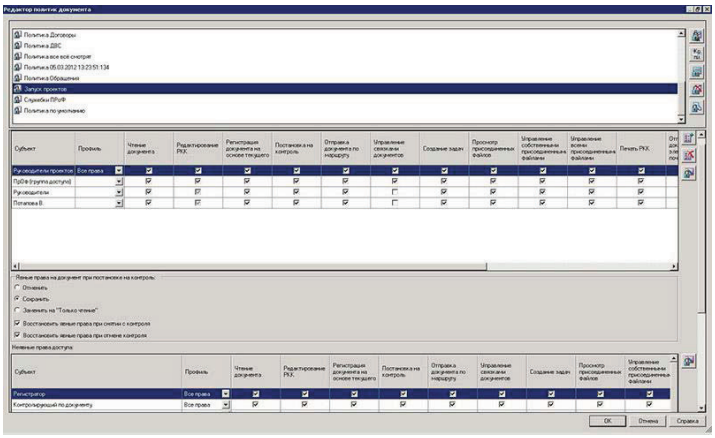


Fig. 2. WFMS «Euphrates» access control policies interface.

III. Implementation of RBWM Extension

To configure access control in WFMS it is necessary to define a set of policies to all securable objects taking into account the amount of different roles and operations on objects and tasks. There are defined about 100 assignments without consideration of possible users and about 500 of securable operations on documents, files, tasks and processes in the standard configuration of WFMS «Euphrates», where the suggested model is implemented. It is evident that not all the operations on the subjects are needed to be defined; it is enough to define only authorized operations. However, the amount of policies in WFMS configuration is one and a half times more than the amount of operations on the objects.

To simplify centralized management of access control policies we suggested to group policies defined in the system by certain set of rules. Grouping by documents of a certain type is implemented in WFMS «Euphrates». The rest securable objects (document properties, messages, files, processes and tasks) are linked to the certain document. Special labels help to manage versioning of the objects because there is no need to define a separate set of policies to the versions of securable objects. There is a "sharing schema of access control policy" for every type of document which corresponds to the three types of sharing described herein. Figure 2 represents the WFMS «Euphrates» access control policies interface and type of sharing grouped by the type of the object.

IV. Comparison with Releated Works

Assignments [7, 9, 10], tasks and objects [4, 5] are taken into account in Role Base Access Control, Task Based Access Control and their derivatives in the context of WFMS. However, versioning of a securable object is left unattended. In recent researches multiple links between objects are implemented in the form of complex constraints [7] or excessive policy algebra for access control [6]. The author applied policy algebra for access control to address some challenges in WFMS in his previous research [11], but practical implementation of this algebra revealed problems with simultaneous execution of permissions of different policies and difficulty of the centralized access management. The present approach suggests to consider the state of a securable object along with its versioning in the model, by analogy with the state of a task in RBWM [5], and to add an extension of policy definition to it. Besides, the sharing schema of simultaneous application of such permissions and difficulty of their management was not previously touched upon in other studies dedicated to constant and temporal permissions in WFMS [4].

V. Conclusion

This paper studies access control model which addresses two issues specific to WFMS: consideration of the linked objects state and versioning of objects. The suggested model is implemented in WFMS «Euphrates» which is integrated in more than 100 enterprises in Russia and CIS (http://www.evfrat.ru/). The model proved itself in cases where the central securable object is a document and links between objects may be represented in the form of a tree where the document itself is the root.

Notwithstanding the addressed challenges, the suggested model has some limitations. In particular, there are conflicts when process, task or file instance interact with several documents at a time. There is a need to simultaneously take into account a policy on several types of documents.

Besides the model does not identify special aspects of the integration with the role based model, access control through assignments and delegation which constitute an important part of WFMS. This is the subject for further research. This is where the author sees opportunities for the development of the present model.

References

[1] W. M. Van Der Aalst, A. H. TersHofstede, B. Kiepuszewski & A.P. Barros "Workflow patterns." Distributed and parallel databases, 14(1), 2003, pp. 5-51.

[2] E. A. Stohr, & J. L. Zhao, "Workflow automation: Overview and research issues." Information Systems Frontiers, 3(3), 2001, pp. 281-296.

[3] N. Lohmann, "Compliance by design for artifact-centric business processes." Information Systems, 38(4), 2013, pp. 606-618.

[4] S. Oh & S. Park, "Task–role-based access control model." Information systems, 28(6), 2003, pp. 533-562.

[5] S. Kandala, & R. Sandhu, "Secure role-based workflow models." Database and Application Security XV, Springer US, 2002, pp. 45-58.

[6] E. Bertino, E. Ferrari, & V. Atluri, "The specification and enforcement of authorization constraints in workflow management systems". ACM Transactions on Information and System Security (TISSEC), 2(1), 1999, pp. 65-104.

[7] J. Crampton, & H. Khambhammettu, "Delegation and satisfiability in workflow systems." In Proceedings of the 13th ACM symposium on Access control models and technologies, ACM, 2008, pp. 31-40.

[8] J. Wainer, A. Kumar, & P. Barthelmess," DW-RBAC: A formal security model of delegation and revocation in workflow systems". Information Systems, 32(3), 2007, pp. 365-384.

[9] H.Y. Balinsky & S.J. Simske, "Differential access for publicly-posted composite documents with multiple workflow participants." Proceedings

of the 10th ACM symposium on Document engineering, ACM, 2010, pp. 115-124

[10] H. Wang, L. Sun, & E. Bertino, "Building access control policy model for privacy preserving and testing policy conflicting problems." Journal of Computer and System Sciences, 80(8), 2014, pp. 1493-1503.

[11] I. Tarkhanov, "Policy algebra for access control in enterprise document management systems." In Application of Information and Communication Technologies (AICT), 2015 9th International Conference, IEEE, 2015, pp. 225-228.

Mobile ID – crucial element of m-Government in the Republic of Azerbaijan

Jana Krimpe
Founder
B.EST Solutions LLC
92 Nizami Str.
Baku, AZ1010, Azerbaijan
Tel: +994 50 245 79 77
yana.krimpe@bestsolutions.az

ABSTRACT
In this paper, we describe the implementation of the Mobile ID (Asan İmza) & e-signature project as a key component of m-Government concept in the Republic of Azerbaijan.

Keywords
Asan İmza, Mobile ID, e-Government, m-Government

1. INTRODUCTION

Technology of PKI based Mobile ID solution is known to everybody. Asan İmza helps Azerbaijan to skip all 20th century online communication platforms and goes directly to the most personal one: mobile. The real innovation relays on the business model – how Mobile-ID has been implemented in Azerbaijan based on Public-Private-Partnership (PPP). Using Mobile-ID in public services helped Azerbaijan to achieve significant results to boost the fight against corruption due to its transparency. Asan İmza connects citizen with their government.

Asan İmza (Mobile ID) is your mobile identity with which you can confirm your identity when accessing e-services and give digital signatures. Asan İmza (Mobile ID) enables to execute all available e-activities. It is essentially a SIM-card of your mobile phone connected to certificates, which constitute a document equal to a physical ID-card in the electronic environment. With Asan İmza (Mobile ID), you are able to prove your identity and digitally sign documents.

2. GLOBAL TRENDS TO KEEP AN EYE ON

So much is happening in the technology nowadays. It is worth to mention some key technology trends to keep an eye on.

First – people are going mobile. The numbers show that mobile usage growth is **very strong**, tablet units are growing faster than PCs ever did (+53%, 2013). This trend continues. So the message is that anything you do, think about mobile solutions (not just web, but tablets-smartphones). User interfaces are needed for everything – from booking a hotel to paying your taxes and making an appointment at your doctor.

As social media is boosting, this is a channel to use to share latest news and market new applications.

Second – big data generates service innovation. Data processing costs are reducing thanks to the cloud technology. It means that we have now much more capacity and opportunities to analyse the data and understand the patterns of behaviour behind. All this is a source of service innovation.

Third – one should not forget that cyber threats are actually growing and we need to look for such solutions, which are safer and which can minimize security risks (such as passwords and hacking).

3. LIFE IS BECOMING MORE DIGITAL…

Today life is becoming more digital. It all reflects in a way how we watch movies, book hotels, transfer money, do our shopping, buy tickets, or talk to each other.

The change is that in all these everyday aspects we use digital solutions – buying our tickets online, shopping online, using Skype or social media for talking to our family, friends and colleagues.

Yesterday these services moved to web. Now they are moving to mobile phones and tablets. Tomorrow more and more services will be designed for – and ONLY for – mobile.

4. WHY WE NEED E-GOVERNMENT?

As the President of Azerbaijan Republic H.E. Mr. Ilham Aliyev said – e- and m-governments help us to make our governments more trustworthy and efficient as it helps to fight against corruption and bribery.

Today telecom operators in Azerbaijan provide Asan İmza SIM cards. All business clients are adopted to the system automatically. Asan İmza helps to reduce their business costs regarding assigning contracts with their clients, partners and suppliers or exchange other formal documents (subscriptions, acts, invoices etc). Students are offered complementary use to get familiar with Asan İmza in an early stage of their life - to learn the benefits and use Asan İmza as a default choice for interacting with the state. Around 10 banks have already integrated their internet and mobile banks with Asan İmza service to make clients possible to sign loan contracts distantly. Integration of Asan İmza with any bank in Azerbaijan takes only 2-3 days.

For users it is more secure and convenient. Asan İmza is your identity on the mobile phone and mobile phone is always with

you. This means that all e-services and state services are always with you. You can use them where and when ever you like. And this one PIN for all different services, you do not need any more to remember lots of different passwords or write them down somewhere and put into your wallet.

5. E-AUTHENTICATION & E-SIGNING IN YOUR MOBILE PHONE

Here we have government and citizens hoping them to interact through e-channels. Now we have a new need – how to make sure that someone else is not using someone other's identity?

Here comes the linking element - which is crucial requirement in building electronic services – this is **secure electronic authentication and digital signing.**

Here we can consider Mobile ID as a cutting-edge solution for electronic authentication used for e-services. And in our case digital signing in Mobile phone as well.

6. GOAL OF SETTING-UP PKI INFRASTRUCTURE IN AZERBAIJAN

Although cloud technology is booming these days, it can't be used in all situations. Especially in those situations, where requirements for security and legal validity are the highest. Like the case when the government has to identify their citizens or residents. It has to be 100% secure as it deals with data which is one the focal component which defines a souvereign country – it's citizens. We can't imagine that issuing passports would not be 100% secure, legally valid and without a governmental guarantee. The same applies to Mobile ID.

That's why Mobile ID requires Private Key Insfrastructure, not any kind of cloud service. Every government who is interested in that, needs to create their own PKI.

7. WHAT IS MOBILE-ID?

Mobile ID is a very simple and secure way for authentication, which is a crucial element for service provision. It is so simple that almost anyone can use it and it is safer than a bank card.

Asan İmza is the Mobile ID service in the Republic of Azerbaijan. It is a SIM based personal identification for digital services.

Asan İmza (Mobile ID) is a secure authentication and digital signing Mobile phone and PKI based solution for public sector, e-service providers and Mobile Network Operators in Azerbaijan. Mobile ID works as "Single-Sign-On solution" – you can use one authentication solution for all different on-line services. For subscribers it means that they just have to enter one PIN-code to mobile phone and get access over 500 e-services (from tax office to banks) that support Asan İmza. Asan İmza can be used for legally binding digital signatures as well.

The simplicity also means that no card reader, no ID card, no e-token and special software is required. Only thing needed is a mobile phone with secure SIM card.

Asam İmza is the most secure on the market and it can be used with all phones: both simple ones and smart-phones.

Of cource, different clients have different needs – for an average citizen the Mobil ID is enough. However, there are also people who need to use the Digital ID in their work, giving big amount of signatures daily (a bank officer or civil servant). For them we have created ID card based solution.

Asan İmza can be integrated with different services. It can be used with PCs and smart devices also, meaning new services can be developed specially for smart devices (tablets/ smartphones).

More and more services can be built up for emotional decisions. It means "you see it, you like it, you authenticate yourself and you have it immediately!" You do not need to go to the nearest service desk or PC (and forgot this topic during that). Or you just see some service advertisement and you find it useful – then you just make a regular phone call with mobile phone (to 195 call-centre), authenticate yourself and it's done! Asan İmza also has iOS and Android applications enabling to manage signing documents directly at your mobile phone.

Asan İmza works everywhere you need to authenticate the customer. Just use your fantasy and provide/personalize your service – as a public sector service provider or private company.

8. ASAN İMZA USES HIGHEST SECURITY STANDARDS

Asan İmza is also simple, consistent and universal. Customers can use it everywhere and user experience is the same everywhere. All local interfaces/applications are standardized and available for all partners – MNOs, e-service providers including government e- & m-services. Big difference is that Asan İmza is a PKI based solution and PKI certificates are used. It means that it is more secure (private keys cannot be copied, they are stored on the SIM card) and Asan İmza supports legally binding digital signature according to the country based legacy. Asan İmza issuance process might be not so convenient, customers must get new SIM card. However, during that process all customers are identified face to face based on the state issued identity document. Asan İmza is a state-granted identity.

We follow international standards using EAL5 level security. It is higher than currently required by the international standards.

Enrollment is based for double authentication – two different companies must verify the customer. It means that it is nearly impossible to get fake identity or stole someone's identity.

For using Asan İmza you must have the handset and know PIN codes. If you lose your phone, nobody can use your identity.

For e-service providers it means that there is one state-granted identity solution and they can trust it. They do not need to invest to build up their own solutions. This is huge amount of

money what all e-service providers can save in total. And users have always same user experience with different service providers.

9. ASAN İMZA IN AZERBAIJAN

Mobile-ID (Asan İmza) system in Azerbaijan is based on the PPP model and the partners here are the government, all three national MNOs, Mobile ID service provider (B.EST Solutions LLC), governmental certification authority and e-service providers.

Asan İmza (Mobile ID), which means in Azerbaijani language „Simple Signature", was developed by the Azerbaijani IT company, B.EST Solutions LLC, and launched as a pilot system in 2013. It started fully operating in 2014 and now all major mobile operators are issuing secure SIM cards.

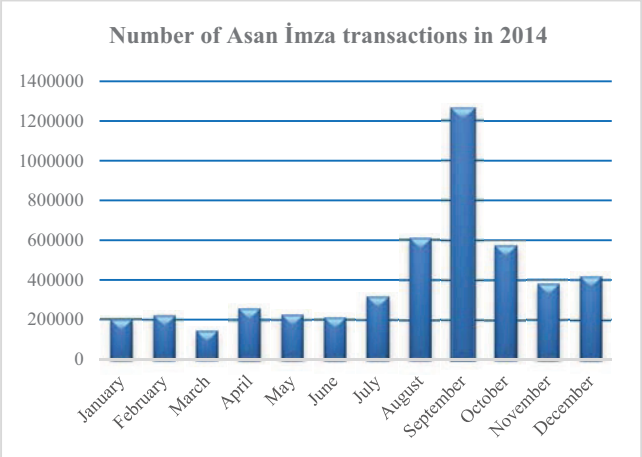
The first Asan İmza Mobile ID certificate was presented to the President of the Republic of Azerbaijan H.E. Ilham Aliyev.

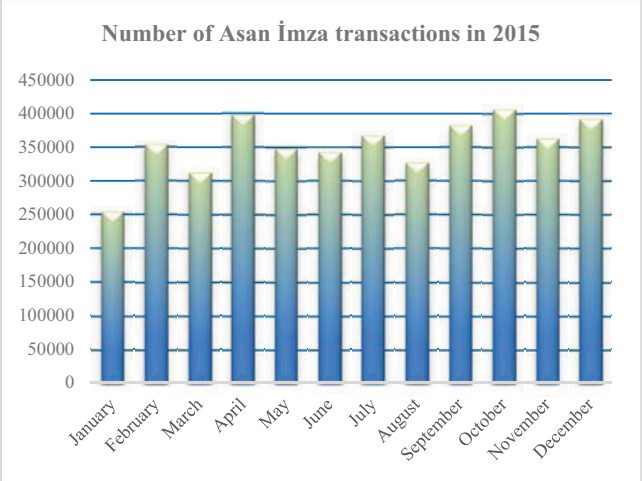
During the two years of the operation of the Asan İmza service in Azerbaijan, more than 250 000 Mobile ID certificates have been issued and over 15 mln. transactions performed – this is the quickest result of Mobile-ID user cases in all over the world and shows the high usability and satisfaction among users. The best indicator showing the success is not actually the number of issued certificates, but the number of real transactions. During this period over 90% of tax declarations in Azerbaijan were submitted electronically. In general, today over 500 public and private e-services are available to citizens with the help of Asan İmza in the Republic of Azerbaijan.

10. YEARLY STATISTICS

In 2014, 124 096 Mobile ID certificates were issued in Azerbaijan, 63650 of which were granted for individual usage, 54 425 – to legal entities and entrepreneurs, and 6 021 – to public servants. In 2015, there were issued 45 283 certificates; 22 703 of them were provided to individuals, 21 028 – to legal entities and entrepreneurs, and 1 552 – to public servants.

One of the significant indicators of successful operation of Mobile ID service is the quantity of transactions. The below graphs describe the number of transactions on yearly basis:





11. STAKEHOLDERS

Asan İmza involves several stakeholders, and these are the most important ones of them.

Certification Authority – **a government-licensed agency (either private or public)** responsible for legal and secure authorization.

Mobile-ID Service Provider – basically the process developer who makes it all work. In Azerbaijan the company „B.EST Solutions" was in this role.

Mobile Operators – cooperation partners whose involvement is needed as they issue secure SIM cards to the mobile phone users.

End-users – for whom it has to be safe and simple in order to become regular user.

End-user Service Providers – such as government agencies and banks who can use Mobile ID in order to safely offer electronic services to their clients.

We created a central OTA platform which functions as a Private-Public Partnership. In the centre of it is the Mobil ID Center and the Certification Authority. "B.EST Solutions" runs the Mobile ID Centre and takes care of the development of the whole ecosystem from client support to software upgrades. The Certification Authority is a state-authorized body which offers the certification service.

12. INTERCONNECTION SCHEME

In the center of the Azerbaijani Mobile ID ecosystem there are Mobile ID Centre and Central Certification Agency. We have been the first in the world who started using central approach. Before that, the mobile operators themselves offered this service, but real life showed weaknesses of such approach – for the mobile operators, it is just not their core business and they are not interested in developing all that enough.

Our company, B.EST Solutions LLC, runs the Mobile ID Centre and we take care of the development of the whole ecosystem – client support, integration of new service-providers (banks, e-services etc), software creation /upgrade, creation of apps etc.

In order to issue first time the sertificate for authentication and digital signing to physical persons, the client goes to one of the **registration centres** (State Agency for Public Services and Social Innovations or Centres for Service to Taxpayers). As a result, this person's SIM card is linked to him/her personally and legally. The registration centres both activate and deactivate digital ID certificates. Basically, the registration centres act like the authorites who issue passports.

In order to provide the services, **service providers** must be in place. They could be both private or public services.

And the mobile operator is connected to the whole system through their SIM cards. They issue secure SIM cards and provide end-users SIM-based services, collecting fees for that.

Implementing Asan İmza is simple for all parties - customer, operator and e-service provider (who does not need big investments for the technology). Mobile ID is provided as a service-based model. Operators must order new type of SIM cards, issue SIM cards to the Customer and open their SMSC gate for the Mobile-ID service provider. Customers have to get new SIM card and order the PKI certificates. The model is very secure – double authentications are made during the issuance process. First, when SIM is issued by the operator, and second, when certificates are issued by the State Agency. E-service providers can integrate Asan İmza into their services very easily using special toolkit for that and integration is free of charge.

13. E-SERVICES AVAILABLE VIA ASAN İMZA

Here are some examples which services already use Asan İmza.

It is worth to mention here the Ministry of Taxes which is actually one of the first agencies to use e-services in the country, as the volume of work is massive and the automatization adds lots of saving in time and money. Asan İmza is actively used in the e-tax operations in Azerbaijan. In 2014, the electronic system of the Ministry of Taxes of Azerbaijan registered almost 6.8 mln. transactions, 33% of which have been implemented by means of Asan İmza. In September-October 2015, there were registered approximately 7.2 mln. transactions in the e-tax system 32% of which were done with Asan İmza.

Starting from August 2014, Azerbaijan started to keep record of employment agreements electronically with the help of Asan İmza mobile e-signature service. This innovative service was launched by the Azerbaijani Ministry of Labor and Social Protection of Population which forced all entrepreneurs and legal entities in the country to conduct all employment agreements with their employees in electronic system of the Ministry by means of using of e-signature solution. This concept was introduced with the aim to strengthen the fight against illegal employment as well as to increase the tax and social insurance revenue to the budget. As noted by the Ministry, thanks to this system of e-registration of employment contracts, which is the first of such kind in the CIS region, the government managed to ensure in a more efficient way the public control over the fulfillment of the employment rights in the country.

Also, Asan İmza are used by the payment portals and many banks which use it in their digital banking platforms.

PPP is a good method, but here it is important to keep in mind that the government has a crucial role in the kick-off phase. Without governmental committment and interest, such innovations would not happen.

The ubiquitous usage of e-services and e-signature significantly increases the computer literacy of the population and is the important incentive to realize all advantages of e-government. It is a trigger for the rapid growth of the digital economy as well.

As of today, in the E-Government Portal of Azerbaijan Republic, there are more than 500 e-services available via Asan İmza. 343 of those services are rendered by the government.

14. USERS ABOUT ASAN İMZA

Satisfied Asan İmza users say…

"Today, the most important benefactor working with electronic authentication and signatures, is time. Time is a critical aspect for our customers. Timeframes for signature of Audit Reports and Contracts have reduced from 1-2 weeks into a matter of an afternoon. Cost and time saving with Asan Imza is literally immeasurable, and has paved the way for us to move wholly into paperless office."

Vugar Aliyev, Managing Partner - KPMG Azerbaijan

"Our company uses Asan İmza in the process of declaring and paying taxes. We are satisfied with this comfortable tool and wish more services would use Asan İmza."

Farid Ismayilzade, Founder & CEO - GoldenPay, Board Member - AMCHAM

"We are the first bank in Azerbaijan to introduce the mobile signature tool – Asan İmza into its internal document circulation system. Thanks to this innovation, we managed to optimize the internal business processes, save time and avoid risks and expenditures related with the document workflow."

Yuriy Yakovlev, Chairman of the Board – Bank VTB Azerbaijan

"A mobile phone and two PIN codes enable a customer to manage his/her accounts using the internet bank even while being out of the country. And what is most important, all these is ensured with maximum security and comfortability even for an ordinary "untechnological" person."

Nahid Zeynalov, 1st Deputy Chairman of the Board – Bank of Baku

"Paper was the bane of our lives, and we had been looking for a reliable and secure way to sign electronic documents in our inside document-management system of our bank. The ability to place multiple signature fields on the same page and to sign multiple pages with just a click makes the signing process fast and easy. We didn't have to change to or buy a specific HR system; it is an extension of our existing systems, so it was agile, quick, and easy to set up with great support from the Asan İmza team."

Emil Mammadov, Vice-Chairman - Nikoil Bank

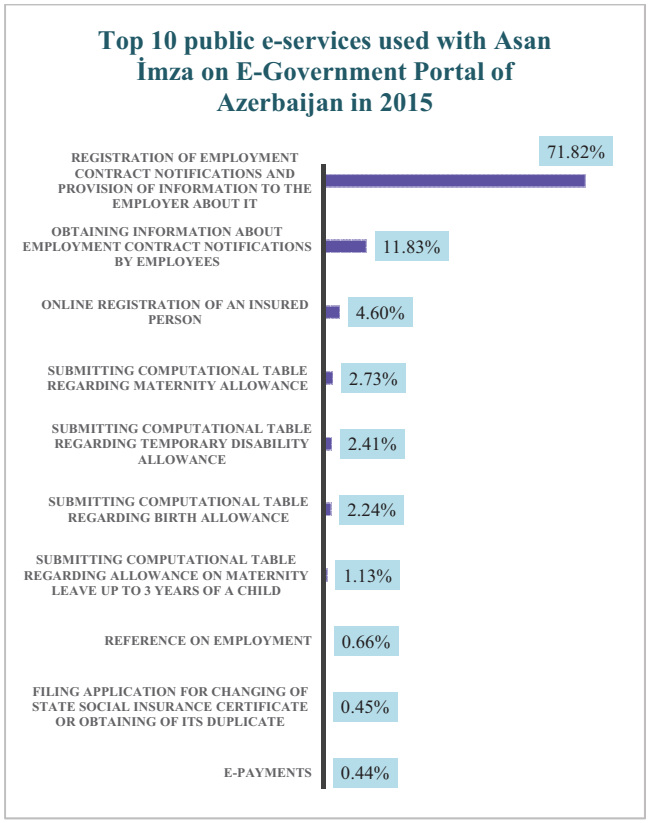
"Asan Imza has shortened and expedited my daily signing process to just a few minutes; this makes my business process much more efficient – with no environmental footprint.
Asan Imza service delivers a great solution to Azerbaijan businesses of all sizes by speeding up business processes and making all governmental and private services accessible to their customers anytime, anywhere."

Sabina Parvizi, Regional Marketing Communications Enterprise Manager - Microsoft Azerbaijan, Belarus, Georgia

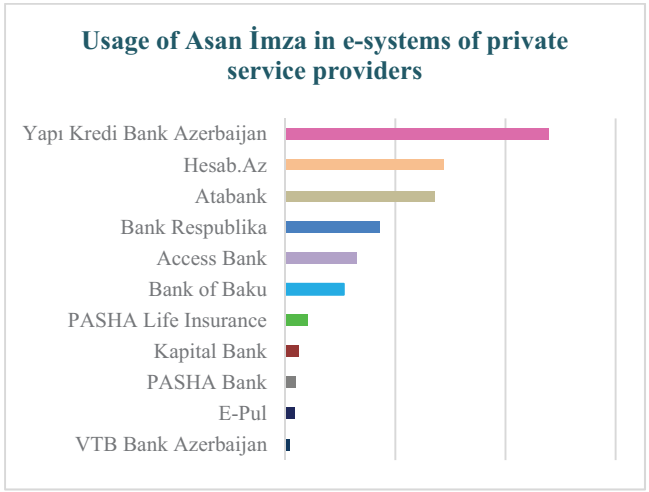
"Asan İmza has significantly eased the work of our respective department working with tax invoices."

Parviz Ismayilov, Business Development Department - Deloitte Azerbaijan

15. RANKING OF E-SERVICES USED WITH ASAN İMZA



Starting 2015, financial institutions operating in Azerbaijan such as banks and insurance companies started actively using Asan İmza in their activities for provision of online services in a more qualitative and secure way to the customers. During 2015, Asan İmza was integrated into the electronic systems of eight leading banks, two payment portals and one insurance company in the country. 10 more banks are in the pipeline for 2016. In 2016 all insurance companies are expected to be connected to Asan İmza.



As of today, Azerbaijani banks use Asan İmza in provision of such online services as money transfers, currency conversion, payment of loans, payment orders, currency buy/sale/conversion orders, card operations, payment of taxes etc.

It is worth to note that along with the internet bank, some banks started using Asan İmza in their internal document flow systems. At present, Asan İmza is the only e-signature tool used in the banking sector which is stipulated with its security, convenience and flexibility in operation. In its turn, the Asan İmza team provides continuous support to banks partnering with the Asan İmza service through all phases of mutual cooperation.

Usage of Asan İmza also enables banks save resources which they usually spend to support their physical branches. Introduction of such innovative solution as Asan İmza, significantly increases profitability of bank services and customer satisfaction especially in the crisis conditions.

Asan İmza is an efficient tool for saving public funds. According to the calculations of foreign experts, ubiquitous usage of mobile identity allows saving up to 2% of GDP on average.

16. ASAN İMZA IDENTIFIES YOU DURING CALL

Asan İmza also identifies you during a call. This is a brand new service, novelty in the world. Basically, it means that during an ordinary phone call a person can identify himself/herself through Mobile ID and a service-provider can help or assist the person to fill in some forms, submit a tax declaration etc.

It is meant for the people who are not computer-literate enough or who does not have access to the computer to use e-services.

So far, various private companies charged money for such service, now the person can directly address the public authority who is responsible for the service and do all that free of charge, as a part of the public service.

First time in Azerbaijan, especially for rural regions where access to the Internet is not very high and other communication problems may exist, the “195 Call Centre” of Taxes Ministry enabled citizens to authenticate and sign documents digitally – just make a call, identify yourself via Asan İmza and make all transactions from tax declarations to opening a bank account.

17. E-STAMPING

Electronic stamping is another service for digital signature which was developed for those service providers who need to sign massive amount of documents (like water or electricity bills every month). It would take ages to sign them separatately, one-by-one. Electronic stamping automatically signs the documents, once the fixed provisions are fulfilled. It adds value by improving productivity for such service providers.

18. VALUE FOR PUBLIC & PRIVATE SECTORS

Mobile ID creates value for both public and private sector.

In public services, it enables access to self-service portals, helps to easily pay taxes and dues, sign-up for services and apply for permits to name some examples.

In private sector, financial services benefit from Mobile ID very clearly – secure identification is very imporant there. As businesses need to sign agreements and contracts every day, digital signature makes it easier and faster. No need to send papers by post or rush to the other end of the city to sign something. Now you send the document by e-mail and just sign it electronically. It is legally valid and secure.

Currently Asan İmza can be used with more than 500 different governmental e-services in Azerbaijan, which are provided by various government organizations and represented on E-Government Portal (www.e-gov-az) and www.e-taxes.gov.az portal. Most popular services used via Asan İmza are the submission of e-tax declarations, registration of a new legal entity and registration of labor contract notifications. Statistics show that over 85% from digitally signed (more than 1 mln.) labor contracts have been signed using Mobile ID. Asan İmza also enables using online banking services as an option of logging in and conducting e-transactions on the Internet and bank portals via Asan İmza. Bank clients can apply for loans, make transactions and other bank services not visiting bank physically.

19. CALCULATOR FOR BUSINESSES

In order to demonstrate how much a company would save on using Mobile ID and other related services, we have developed a special calculator. It has become rather popular and makes clear to businesses how they could cut costs using Mobile ID.

20. OPENING WAY FOR CROSS-BORDER DIGITAL SERVICES – COOPERATION WITH ISESCO

All that does not have to remain inside one country. We are now opening ways for cross-border use of the Azerbaijani Mobile ID solution.

Now we are piloting a joint project with ISESCO which envisages for establishment of a single internal document management platform for ISESCO member states. The platform, titled as "ISESCO BESTDOC PORTAL", will be functioning on the basis of a unique digital identity tool – ISESCO ID, designated specially for ISESCO based on the Azerbaijani technologies of electronic and mobile identity – "Asan ID" and "Asan İmza", and will provide an opportunity to ISESCO member states to interact within single e-environment aiming to improve the process of document workflow. The platform will support such functionalities as strong authentication and validation of signatures of the users via ISESCO ID & m-ID, e-signing, storing and sharing of documents. In general, it is expected that launch of the platform would facilitate further development of the relations between ISESCO member states with more focus on IT innovations and electronic solutions.

21. VALUE FOR SOCIETY

E-services and mobile authentication creates value for the society in several ways.

First, the service quality improves. Besides the fact that it is easier to use for the clients, it also creates better results. Many actions which were previously being done manually are now automated. It means less mistakes, equal treatment of citizens, transparency and much less opportunities for corruption.

Second, it increases efficiency. Behind any e-service there is a streamlined and optimized process. It means that every step is well considered and unneccessary actions and burdens are eliminated. It means an ability to serve more citizens, less bureaucracy and paperless interaction.

Third, service innovation is a big value too. E- and m-services create opportunities for businesses, new applications which are better and smoother. Instead of going to three different governmental agencies for getting a construction permit, why not make it smooth whole without need to go to those three offices.

Another important aspect is that Asan İmza follows the EU standards; it means that these services can be used and developed across the borders.

22. CONCLUSION

How a Mobile ID infrastructure can be implemented in a country?! Good news is that it is not too hard, costly and complicated once you have a professional partner who knows how to do that.

Basically, you need two things – to set up a Certification Authority and have Mobile Network Operators on board. A good Mobile ID service provider can take care of everything else, from client support to developing necessary applications.

All of that is doable in half a year. The easier part is technological and technical. The hardest part is cultural and organisational change – moving to high-quality processes.

To conclude, instead of wasting millions contracting large, slow development projects that result in inflexible systems, it is much smarter to find systems that are already working, allowing for faster implementation.

Mobile ID is a proved solution which creates a lot of value for the society, both for citizens and businesses.

Carrier-Grade NAT - Is It Really Secure for Customers? A Test on a Turkish Service Provider

Kevser Ovaz Akpinar[1], Mustafa Akpinar[2], Ibrahim Ozcelik, Nejat Yumusak
Department of Computer Engineering
University of Sakarya
Sakarya, Turkey
kovaz@sakarya.edu.tr, akpinar@sakarya.edu.tr, ozcelik@sakarya.edu.tr, nyumusak@sakarya.edu.tr
[1]Correspoding author, [2]Member IEEE

***Abstract*—Dramatic rise in the user amount yields increase in the number of internet accessed devices within the last decade. Since most of the devices have internet connection, IPv4 space becomes inadequate. In order to avoid this situation, internet service providers focus on using their IP's within their IP pool, optimally. The most preferred approach to handle this problem is called Carrier Grade Network Address Translation (CGN). In this technique, a city, a neighborhood or a group of users could be configured as if they are in the same Local Area Network (LAN) and they have IPv4 Network Address Translation (NAT) connections for Wide Area Network (WAN) accesses. By applying this approach, IP costs are reduced and number of IP's in the pool is optimized. However, implementations done in recent systems could cause vulnerabilities as well. This work focuses on examining a part of CGN applied network that acts as LAN, by scanning, exploring users, devices and vulnerabilities for a specific neighborhood in Turkey. Users and devices are determined and since they are considered in the same LAN most of them are easily gained access and the insecurity of the system is proved. Also it is also observed that a user could stop or slow down the traffic by Denial of Service (DoS) or Distributed DoS attacks.**

***Index Terms*— Carrier Grade Network Address Translation, CGN, CG-NAT, security, vulnerability, IPv4, network security.**

I. Introduction

Number of end-user devices is increasing day by day. Computers, cell phones, tablets and a recent trend namely Internet of Things (IoT) concept also have undeniable contribution on consuming IPv4 addresses [1]. Past investments done on hardware have made difficult to convert many systems from IPv4 to IPv6. Also hardware costs for IPv6 transition have led experts to seek other solutions. One of these solutions is called CGN where the first research on this technique is done on 2008 by Durand [2]. Following this study, Maennel et al introduced an alternative method to CGN in their report which is claimed to be a better approach [3]. They showed harmful effects of CGN and gave brief information about CGN security. Kantola et al stated that user network connectivity is maximized by implementing CGN. They also applied an Ethernet based Internet architecture and examined various deployment scenarios on it [4]. Wing explained CGN and NAT in details and investigated recent studies done on related topic [5]. Bocchi et al examined CGN effects on web surfing [6]. They also stated matureness and stability of CGN approach. They mentioned that CGN existence could bring positive side effects such as it prevents home routers from unauthorized and malicious traffic. However, they did not mention about local attacks. Bocchi et al showed statistical methods for network monitoring on CGN [7]. They illustrated how to make preprocessing and showed resource saving policies. Richter et al made multi-dimension analysis of CGN deployment [8]. They gave detail information about CGN and detection CGN. CGN is a new technology instead of IPv6 and it has been accepted in the market rapidly.

Next steps of this work are studied as follows: second section called methodology gives information about the infrastructure, third section gives brief information about tools used and pre-process steps, forth section gives the results and the last section concludes the paper.

II. Methodology

A. Network Address Translation

Network Address Translation (NAT) is an approach to optimally use the limited number of IP address [9,10]. NAT does variety of IP address translations such as a one-to-one or a many-to-one. The main objective of the method is mapping an inside (local) IP address to an outside (global) IP address, so that an internal IP address is substituted by the suitable external IP address, and vice versa. During Network Address Translation, the IP address in the frame is replaced with the new internal or external IP address. If port translation is activated, port field is also updated in the TCP/UDP header. After this, first IP header then TCP header checksum is computed one more time and checked for integrity facts. Since port, payload (not always) and IP address field are changed, TCP header checksum needs to be updated here [10].

B. Carrier-Grade NAT

Carrier Grade NAT is developed by IETF and also known as Large Scale NAT (LSN) or NAT 444. CGN uses IPv4 and allows local users to access internet through a NAT process

[1,3,5]. It is first introduced to solve the problem of inability of service providers to provide service to customers by depletion of IPv4 addresses.

Disadvantages;

- Lack of remote desktop or camera access
- Cannot open port, even if opened it cannot be used
- If user is in a week CGN IP pool, slow internet connection or high pings may occur
- Since extra hop node is added, latency increases
- For services receiving registration or login from only one IP, in case of more than one user entrance from the same pool, problems may occur
- Existence Security vulnerabilities

Advantages;

- Many end-points can be connected on one IP address
- Decreasing IP costs for service providers
- Management of networks is easier

For CGN approach, F5 Vipron 4480, Vipron 2400, A10 TH5430 and etc. devices are used. CGN systems can hold variety of security vulnerabilities. The first and the simplest one is DoS attacks. CGN devices include a table which maps internal and external IP and port. Two types of DoS attacks can be done on a CGN system. The first attack is DoS or DDoS attack from an outside user to a user inside. The second one is generating high frequency traffic towards internet by a malware residing at an internal user and attempting to swallow CGN state tables. Eventually, both victim on the other side and thousands of costumers being served on CGN system could be affected from the attack. The available port number in the table will decrease and unstable states will occur. Blackhole logic cannot be applied here. Access control list cannot be created as well.

Another security hole is attacking from an insider to other users within the same network. In case of this attack, modems, computers or printers of other users in the same LAN can be simply accessed. Within the scope of this study, instead applying DDoS attack, local network is scanned and end-user devices are accessed.

III. Tools & Pre-Process

The first, network should be scanned and online connections should be detected. Following, for each active connection an open port scan should be processed. Here, accessing devices through their open ports is attempted. There is variety of IP and port scanning tools both for Linux and Windows based systems. For Windows "SoftPerfect Network Scanner" (netscan) has ability to scan active IPs [11], while for Linux "nmap" handles this process [12]. For instance, features such as remote device name, description, whether it has shared folders or not, can be easily and quickly seen through "netscan". "Nmap" tool, however, shows all open ports of an active connection. Therefore, if a simple and superficial result is needed, "netscan" can be preferred. "Namp" is suitable for more detailed examinations. For the next step more sophisticated tools such as "metasploit", "netsparker" etc. can be studied. This work only focuses on examining the CGN network and accessing modem interfaces for proving system insecurity. Black hat hackers could do more complex researches for the next step such as data leakage, backdoor, or malware installation. For CGN investigation, user's directly connected modem should have NAT connection to Kali installed virtual computer.

In order to archive following steps internet outbound should be directed to source computer within the modem device. For this process, either Demilitarized Zone (DMZ) or NAT could be done. NAT is done only for specific ports while DMZ has ability to forward traffic from all ports to a single computer (Fig. 1). Another words, DMZ feature allows all ports of a LAN client to be accessible from the WAN and forwarding all WAN traffic to a specific LAN user. This process helps collecting all packets into the computer that will make scan later.

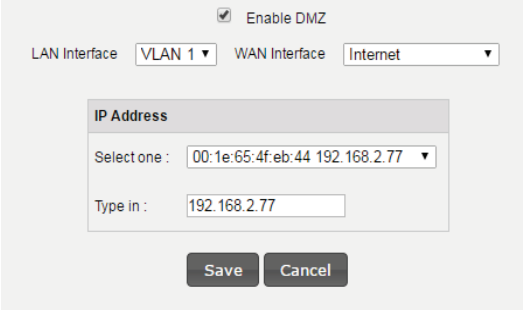


Fig. 1. DMZ assignments

Next step is detecting CGN applied IP address from the modem interface (Fig. 2). There is other information other then IP address here. Some of these information are WAN MAC address, Service Set Identifier (SSID), duration of the modem since power on and current time.

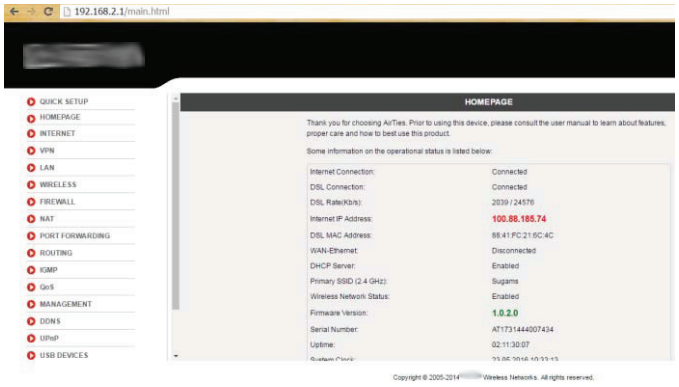


Fig. 2. Summary of modem interface

For the next step, gateway IP used for routing should be determined (Fig. 3). The routing table structure here which can be seen usually at gateway tab on most modems is generated in run-time. Thus, LAN and WAN sections are found and CGN scan starts.

Route Table

Connection	Destination	Gateway	Netmask	Metric

Route Table Run-Time

Connection	Destination	Gateway	Netmask	Metric
ppp0	default	100.88.128.1	255.255.255.255	0

Fig. 3. Gateway Address

IV. Results

The gateway IP 100.88.128.1/17 (Fig. 3) shows the first 17 bits are reserved for network and remaining 15 bits are reserved for computers in local. Thus, it is observed that IP addresses between 100.88.128.1 – 100.88.255.255 can be used as network scan range for searching active connections. In this study, "netscan" tool is used on a Windows computer [11]. Scan is done on 23.05.2016 and duration of around 2 hours. It is processed by 50 threads simultaneously. As showed in Figure 4, 2987 online connections are detected in the CGN network in local.

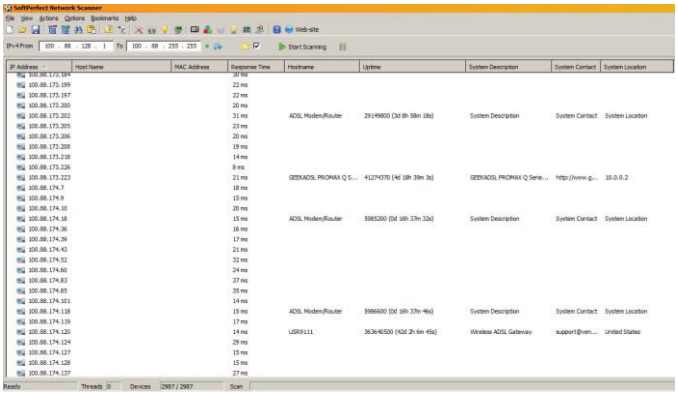


Fig. 4. Netscan results

By the guide of scan results, randomly picked number of modems are attempted to be accessed (Fig. 5). The devices whose default passwords are not changed before, are easily accessed.

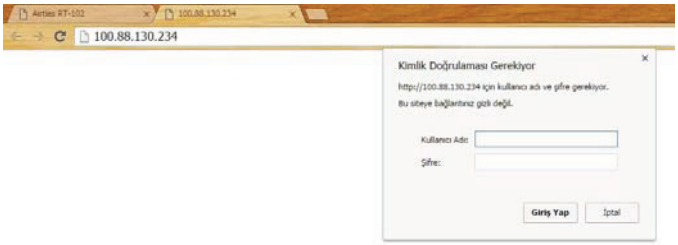


Fig. 5. Login authentication screen

As presented in Figure 6 targeted user's ISP user name and password, wireless connection password, device names and descriptions that are connected to their LAN network can be examined. After this phase, any requested computer connected to this device or shared folders (if any) can be accessed by applying DMZ or enabling port forwarding feature. Other than this, attacks can be realized to the devices that are not regularly updated or don't have accurately secured operating systems within the CGN. The next phase containing these types of attacks is out of scope of this paper since the limitation of legal responsibilities.

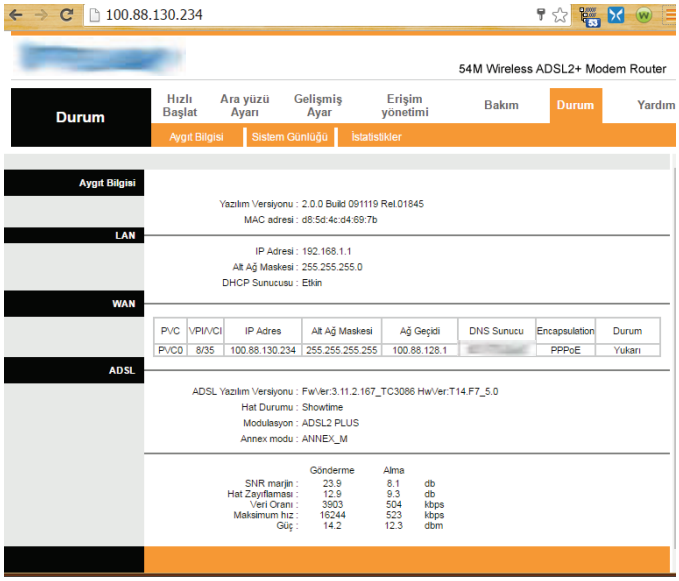


Fig. 6. Logged in modem interface

As showed above, instead modem configurations, shared printers are easily can be accessed as if they are in the same local network (Fig. 7). SCX-3400 network type of printer by a company is detected through this scan. For network printers, since their default passwords are rarely changed or even if changed they are assigned very simple new passwords (like company names), the passwords can be found either by a simple search or a brute force attack (Fig. 8).

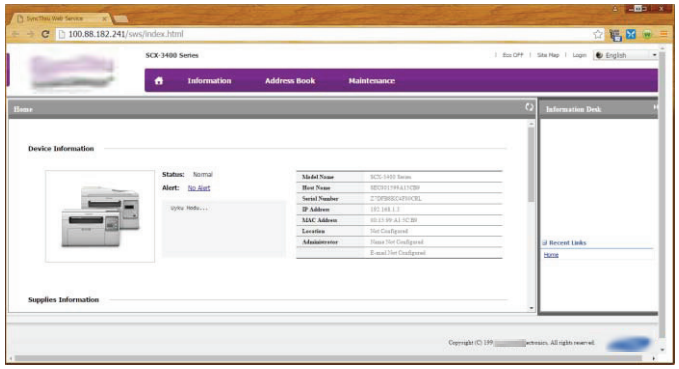


Fig. 7. Printer web interface in CGN

Eventually, hardware and software configuration details or a network printer can be inspected. For instance, printer found in this scan has 67% full of toner. An attacker can also send numerous documents to be printed and cause serious concern.

In addition to these simple vulnerabilities, there is also possibility to determine geographical locations of the nodes through CGN. For this, response time feature can be used. In

this research, response time changes between 959ms and 2ms. 2ms shows our own connection. The lowest response time after

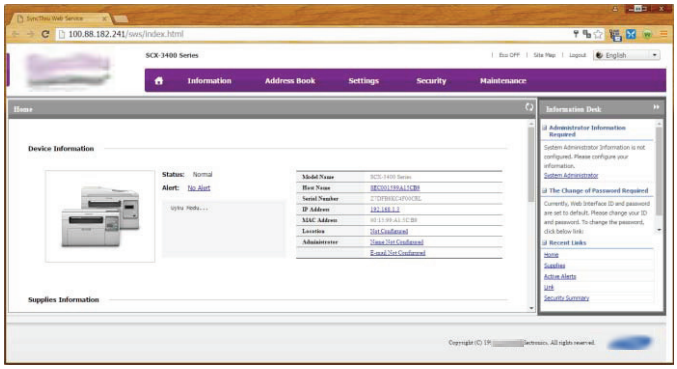


Fig. 8. Administrator logged printer web interface

this is 6ms by network IP. If these 2 values are ignored, the new lowest response time is observed as 7ms. and all end-users' average response time is 34.27ms. Figure 9 presents the state of response times less than 40ms. where there are 2334 connections within this range. There also are 653 different connections above 40ms. It can be observed that, end-devices with low response times are geographically close to the source computer where nodes with high latencies are farther. By this type of scan, close nodes to the source node can be predicted.

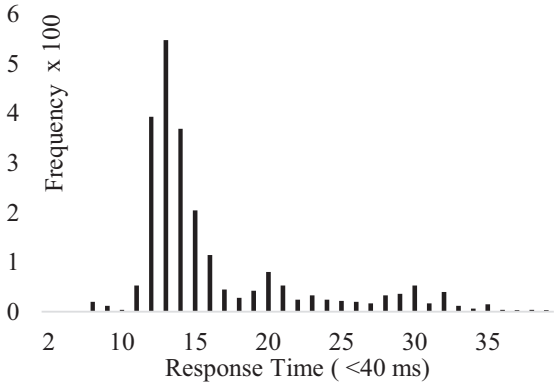


Fig. 9. Response time graph of end users in CGN

In order to avoid these kind of attacks, purpose-build DDoS mitigation devices or scrubbing devices can be used. Also for further protection, client's maximum number of session can be limited or in-flight traffic can be filtered.

V. Conclusion

This study focuses on security vulnerabilities arisen by switching the conventional Ipv4 service into virtual NAT called CGN infrastructure for internet access at most regions of the Turkey by a service provider company. These vulnerabilities cannot easily be predicted or solved by a regular/basic level end-user. As a result of CG-NAT implementation, all users are seen as if they are in the same local network, thus, this application brings some vulnerabilities which are presented throughout this paper. It is also observed that private configurations, NAT or sharing processes within users' devices can be easily accessed or changed by attackers. Not only file accesses but also modems are accessed in this research. Another point in this insecure CGN implementation is that geographical locations of devices can be superficially detected by examining assigned IP addresses' latency and ping respond values. The prevention solutions such as using mitigation or scrubbing devices or assigning authentication field to every end-user on CGN side, all depend on service providers.

References

[1] "RFC 2663 - IP Network Address Translator (NAT) Terminology and Considerations", Tools.ietf.org, 2016. [Online]. Available: http://tools.ietf.org/html/rfc2663. [Accessed: 12- Apr- 2016].

[2] D. L. Herbert, S. S. Devgan and C. Beane, "Application of network address translation in a local area network," Proceedings of the 33rd Southeastern Symposium on System Theory, pp. 315-318, Texas, USA, 2001.

[3] O. Maennel, R. Bush, L. Cittadini, and S. Bellovin, "A Better Approach than Carrier-Grade-NAT," Technical Report CUCS-041-80, September 2008.

[4] R. Kantola, M. Laoma and O. P. Lamminen, "Transport for Carrier Grade Internet," 2009 IEEE Globecom Workshops, pp. 1-8, Hawaii, USA, November 2009.

[5] D. Wang, "Network Address Translation: Extending the Internet Address Space," IEEE internet computing, vol. 14(4), 2010, 66-70.

[6] E. Bocchi, A. S. Khatouni, S. Traverso, A. Finamore, V. D. Gennaro, M. Mellia, M. Munafò, and D. Rossi. "Impact of Carrier-Grade NAT on Web Browsing," IEEE International Wireless Communications & Mobile Computing Conference, pp. 532-537, Dubrovnik, Croatia, August 2015.

[7] E. Bocchi, A.S. Khatouni, S Traverso, A. Finamore, M. Munafò, M. Mellia and D. Rossi, "Statistical network monitoring: Methodolgy and application to Carrier-Grade NAT," Computer Networks, June 2016, Accepted.

[8] P. Richter, F. Wohlfart, N. Vallina-Rodriguez, M. Allman, R. Bush, A. Feldmann, C. Kreibich, N. Weaver and V. Paxson, "A Multi-perspective Analysis of Carrier-Grade NAT Deployement," arXiv preprint arXiv:1605.05606, 2016.

[9] H. Ma, Y. Wu, Y. Ma, Z. Wang, "Optimization Scheme of CGN Logs," 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, pp. 817-820, Hangzhou, China, October 2012.

[10] E. Kohler, R. Morris, M. Poletto, "Modular Components for Network Address Translation," IEEE Open Architectures and Network Programming Conference, New York, USA, June 2002.

[11] SoftPerfect Co., "SoftPerfect Network Scanner Online User Manual," Brisbane, Australia, 2016 [Online]. Available: https://www.softperfect.com/products/networkscanner/manual/ [Accessed: 12- Apr- 2016].

[12] G. Lyon, *Nmap Network Scanning: Official Nmap Project Guide to Network Discovery and Security Scanning*, Insecure.Com LLC, USA, 2009.

New Approach To Software Code Diversification In Interpreted Languages Based On The Moving Target Technology

Mikhail Styugin, Vyacheslav Zolotarev,
Anton Prokhorov
Department of Information Security
SibSAU
Krasnoyarsk, Russia
styugin@gmail.com

Roman Gorbil
Department of Applied Mathematics and Information Security
SFU
Krasnoyarsk, Russia
rons@mail.ru

***Abstract* – The paper presents a method based on moving target technology for protecting software components of distributed systems. Practical implementation of this method is a code diversifier, which adds intermediary functions and inserts transformation of key variables into the program code. Theoretical analysis of the method presented here demonstrated its effectiveness even when an adversary had access to a application's program code. The diversifier presented below can also function in a mode of program code obfuscation, which was tested and demonstrated by the example of interpreted programming language functions. Metrics obtained from the program code after obfuscation, showed sufficient level of code modification for independent use in obfuscation mode.**

***Index Terms* – software security, diversification, obfuscation, moving target defense.**

I. Introduction

Algorithms for concealing data, as well as other information security algorithms are effective only when software components, which implement them, are secure. If adversary has the ability to make unauthorized code injections (attacks) into an application or to get reading access to key internal characteristics of a protocol then further security measures taken to protect the information are worthless. Nowadays, there is a lot of research carried out in the field of program code security. Main areas of the research are program code obfuscation and diversification. Currently, the most commonly used exploit is code injection attacks. Protection from those attacks is based on constant modification of an application using the moving target technology. However, existing protection methods are not always effective in case an adversary gains reading access to the executable code. Besides, existing methods usually involve overhead.

The problem solved by program code protection against research is about complicating injection of components into the code, which can lead to unauthorized access to processes or variables in a computer's memory. This vulnerability can be exploited for stealing confidential information from memory, spreading computer viruses and automated information theft systems, using somebody else's program code, and so forth. There are several approaches to such protection against research. The first and most well-known approach is the program code obfuscation implemented in machine codes or in a high-level language [1-3].

Although obfuscation is a technology for protection from program code research, but does not completely fulfil the concept of moving targets reviewed in this paper. It is also necessary to consider one more area which became very popular among researches over the last few years. This problem is referred to as Software Diversity.

Diversification addresses a problem of software monoculture. Software monoculture leads to a situation when once an intruder finds a solution for code injection into a specific version of a program it allows applying the same exploit to all programs of the same version that is commonly used. Similarly, a vulnerability found once in a program is exploited in all programs of the same version. This enables remote and automated cracking of information systems, as well as spreading computer viruses and worms.

The most popular area in the software diversification field is instruction-set randomization, ISR [4-6]. Constant modification of the code by randomization makes code injection to an application impracticable. Founders of the ISR technology claim that the idea comes from biology, where organisms respond to numerous threats of the environment through genetic changes. Similarly, the ISR mechanism changes the key executable instructions so that embedding external code in an application does not lead to correct execution. The ISR technology helps protecting against all types of remote code injection attacks regardless of the remote injection method by providing adversary's target displacement (attack plane displacement). Thus, the ISR technology employs the fundamental principles of more general technology of Moving Target Defence (protection based on a moving target).

It should be noted that ISR cannot protect a system from the most basic types of attacks, such as the case when an adversary gets the binary code of the required file and

examines the possibility of embedding code into it and run it on a victim's computer.

Another popular approach to software diversification, which became highly popular in last few years, is a mutation of the executable code. The most famous works in this area are [7-11]. The area of code mutation itself was originally considered for the purpose of automatic error correction and measurement of testing quality [7], [8]. For example, tests for a program and program copies with the same functions (mutants) were generated. Then, each mutant was checked with the test. If the mutant did not pass the test, it was deleted. Quality of a test or a group of tests was evaluated by the number of mutants remaining after such mutant test procedure. When the number of mutants was too high then the test was recognized as weak.

Further, we will analyze disadvantages of diversification solutions and how they can be improved by using solutions of the authors.

II. Program Code Diversification Method with the Use of a Self-Complicating Diversifier

The diversification technology can be used in a self-complicating version [12]. This allows improving program code's protection against research. This technology will be used for creating a solution that includes a code diversifier for program components which generally solves the problem of protection against system research.

Application consists of the main code and procedures, functions or interfaces called in by the main code. These elements can be placed either inside the application or in operating system's libraries, remote interfaces and so on. Dividing an application into the functional blocks, we can define input and output values for each of them. As a result, we receive a functional structure which can be made more complicated by adding functional elements. Consider the scheme in Figure 1 with an intermediary block, which was added for *F1* function.

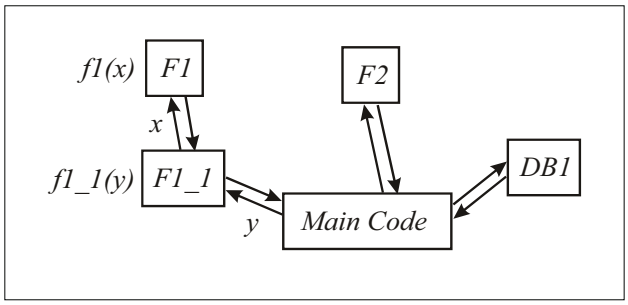


Fig. 1. Application's structure with an intermediary module.

It is assumed that the called function *F1* receives variable *x* as input and returns the result to function *f1(x)*. The intermediary module receives variable *y* as input and sends the result of function *f1_1(y)* to function *F1*.

Thus, the *Main Code* which initially called function *F1* with parameter *x* to get result *f1(x)* and now it should call function *F1_1* with parameter *y*, which calls function *F1* with parameter *f1_1(y)* and returns the result to the *Main Code*. Therefore, to achieve the same result we need to send the value of $f1\_1^{-1}(x)$ as input of *F1_1*. It is also practical to change the roles of functions *F1_1* and *F1* so that function *F1_1* replaces legal function *F1* in the program.

Process can be complicated by introducing additional intermediary modules as shown in Figure 2.

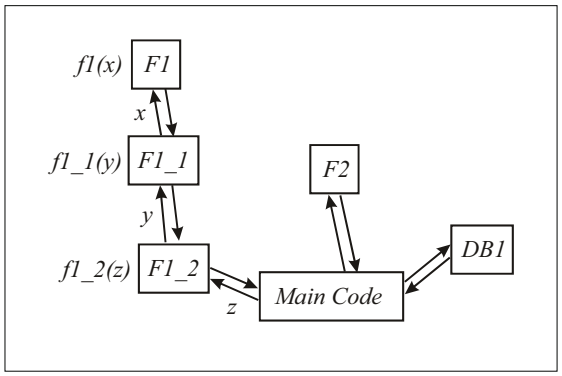


Fig. 2. Application's structure with a chain of two intermediary modules.

Here we get one more function *F1_2*, which sends parameters for function *F1_1*, which in its turn sends parameters for *F1*. In order to send the correct parameter to *F1*, the main application shall calculate parameter $f1\_2^{-1}(f1\_1^{-1}(x))$ and send it.

Previously mentioned examples demonstrate the main mechanism of the diversifier's operation, however these cases are not resistant to code injection attacks. Implementation of a resistant diversifier is reviewed in the next Section.

III. Implementation

A generalized technology of software diversifier based on the moving target technology was described in the previous Section. However, such generalized concept of a diversifier cannot guarantee solution of the above problem, i.e. protection against research of program components to prevent unauthorized code injection.

Generalized diversification algorithm:

1) Establishing the application's functional structure.
2) Creating random variables based on environment parameters.
3) Establishing a modified functional structure.
4) Application's code modification
5) Timeout of program error (go to step 2)
6) Authorized code modification (go to step 1).

Each step of diversifier's operation in details is described below:

1) Establishing the application's functional structure. An example of a functional structure is shown in Figure 1. Functional structure, which is practically suitable for modification should contain not only internal and external modules, which have inputs and outputs, but also a hierarchy of variables, functions and classes to define all the segments of the application where a particular piece of code is used.

2) Creating random variables based on environment parameters. Random variables here are a long sequence of

random or pseudorandom numbers. For generating such sequence a distributed random number generator is used, which allows setting a shortest unpredictable length of a pseudorandom sequence required. The resulting variables are used for generating intermediary blocks and parameter modification constants.

3) Creating a modified functional structure. This step is the most difficult in developing a diversifier. The modified scheme includes newly added intermediary blocks, as shown in Figure 1 and 2. For defining in which functional branches of the application's graph (obtained at step 1) additional blocks should be inserted, the two following parameters are considered: hacking criticality and overhead factor. For example, an application's database has maximum criticality, but in case of high load it also has high overheads when there is a high load on the system's resources. Using these two parameters and a pseudo-random sequence, diversifier generates a pseudorandom functional structure of the new application.

Special attention should also be paid to the logic of building intermediary blocks. In general, these functions use names of the original function and call it with a modified input parameter. Change of the input parameter is made with a pseudorandom key. The key is joined with the input parameter according to predefined conversion patterns. The most simple one of them is XOR operation.

In addition to the parameter-key in each intermediary function, conversion of types is also used, i.e. it is compulsory that the types of input and output parameters should not coincide. (E.g.) If the function should have a variable integer as input, then the intermediary function will receive a string type parameter or any other except for an integer. Predefined patterns are also used to make types equal.

Intermediary functions are always inserted by the code of the final execution environment. That is, if we call for a SQL database, then an intermediary function must be a stored procedure of the SQL language. This leads to the situation when even if the final conversion (original function call) exposes the injection code, it will be inserted as a variable, and injection will not happen. To make an injection one should have all the keys and data conversion patterns to expose a code in data transformation at intermediary function which is impossible for remote attacks.

Diversifier can also be divided by "fast" and "slow" of code diversification methods. The slow diversification methods is code modification with construction of a new functional structure. At the stage of such modification, constants are assigned into the environment (for example, conversion keys) which may be easily adjusted without code modification later on. Such transformations will be called "fast" diversification.

The whole method above is based on transfer of variables to external or internal functional elements. For transferring constants a more efficient technique was developed.

For example, there is a certain constant C which is sent to the input of the function $F(C)$. The task is to create an intermediary function $F1(x)$ so that $C = F_1(x)$ can be easily computed, but $x = F_1^{-1}(C)$ should be difficult to calculate. As a result, a potential adversary will not be able to compute the values of x for the require constants C, even despite the fact that he sees function *F1 ()* in the code. Thus, we can complicate the process of code injection for adversaries and, in some cases, make it impossible, even if an adversary has full access to all application files (but does not have access to internal parameters of the diversifier).

Operation algorithm:

1) Input data receipt. A file (a set of files) is sent as program input for diversification, cracking criticality factor c, value of which lies within [0, 1]. Cost factor is not required, because it can be simply calculated as $(1 - c)$. Cracking criticality factor is an external parameter in relation to the program.

2) Construction of program code's functional graph. It is noted as root vertex of graph point of input (main function or code block, which is performed outside functions and calls all declared functions). Identification of functions and its connections in the program code will be made by a determined finite-state machine.

3) Vector formation. Based on maximum quantity of functional blocks of program n_{max} and cracking criticality factor c, with the help of pseudorandom sequences M-PRNG generator, vector $N=(n_1, n_2, \ldots, n_m)$ is selected, where n_i is the quantity of added program functional blocks, m is the quantity of graph's vertexes (the root vertex is not included), which is built at stage 2, i is vertex number. Data structure, which describes the graph stores additional information about formal parameters of each node, which are accepted by functions.

The disadvantage of this implementation is the lack of recursion support and the relatively high complexity of the algorithm. Also, this implementation is not applicable to attached functions.

Below is an example of a function that is called from an entry point and two conversions of two formal arguments of the function. The source code of the program prior to conversion was as follows:

```
def func1(str1, str2):
    return str1 + str2
print func(str1, str2)
```

After conversion:

```
import ctypes
def fwecxjkhv(jewnkdsp, sdfjsoci4cj):
    return jewnkdsp + sdfjsoci4cj
def ofipcvibp3(ewchsdl, djf7dpuxc):
    return sdjfslkvuocx(ewchsdl / 226 + 784532,
(djf7dpuxc - 431245) * 84)
def sdjfslkvuocx(wenfiv47, fhwk32djc):
    return fwecxjkhv(ctypes.cast(wenfiv47 -
784532, ctypes.py_object).value,
    ctypes.cast(fhwk32djc / 84,
ctypes.py_object).value)
print ofipcvibp3(id(eweijvc) * 226, id(reopig3) +
431245)
```

As seen from the above, the initial code is quite distorted and it will create difficulties for an adversary in case of

attempts of automatic or manual analysis. It should be understood that the specified module implements only an extra level of code protection, which means that the results of diversifier obfuscation mode are shown in the present section. Additional protection should be implemented as specified in the technology presented above.

As it was stated above, practical implementation of the method for protection against research of program components of a distributed computer system based on the moving target technology can be used for protection against remote attacks attempting injection of unauthorized code, hence to increase security of a distributed computer system.

IV. Experiment

The program for searching a discrete logarithm will be taken as example.

We also will measure code metrics of the initial version and the obfuscated version of this program with *Radon*, a tool for initial code metrics measurement for Python programming language. Using metrics to evaluate the quality of the code is recommended in many research papers, as for example in [13].

Changes of quality and quantity metrics for a diversifier which operates in the mode of program code obfuscation and implements a mathematical function will be demonstrated.

A. Quantity Metrics

Amount of initial code lines.

This metric was originally developed for evaluation of labour input in a project. However, due to the fact that the same functionality can be split into several lines or written in one line, the metric was practically inapplicable with occurrence of languages in which one line can contain more than one command. Radon allows finding LLOC, i.e. the number of logical lines of initial code, i.e. lines containing only one operation.

TABLE I. LLOC METRIC

Metric	Source code	Obfuscated code	Ratio	Direction
LLOC	34	155	455.88%	Growth

The number of logical lines of initial code after obfuscation increased significantly because intermediary functions were added. As shown in the example above, this reduces the possibility of manual analysis, as well as semi-automatic construction of the program code flow diagram.

B. Halsted Metrics

Data is based on the following figures:

h1 the number of unique program operators, including delimiting characters, names of procedures and operation symbols (operator vocabulary),

h2 number of unique program operands (operand vocabulary),

N1 total number of operators in program,

N2 total number of operands in program,

$h = h1 + h2$ vocabulary,

$N = N1 + N2$ program length,

$N' = n1\log_2(n1) + n2\log_2(n2)$ calculated length,

$V = N * \log_2(h)$ program volume,

$D = (h1/2) * (N2/h2)$ program difficulty,

$E = D * V$ efforts required for program implementation,

$B = V / 3000$ quantity of potential errors.

TABLE II. HALSTEAD METRICS

Metric	Source code	Obfuscated code	Ratio	Direction
h1	6	7	116.6%	Growth
h2	28	638	2278.57%	Growth
N1	18	336	1866.66%	Growth
N2	36	672	1866.66%	Growth
h	34	645	1897.05%	Growth
N	54	1008	1866.66%	Growth
N'	150	5964	3976.00%	Growth
V	274	9407	3433.21%	Growth
D	3.85	3.686	95.74%	Decline
E	1059	34682	3274.97%	Growth
B	0.09157	3.135	3423.61%	Growth

Vocabulary, length, volume, efforts required for implementation and the number of possible program errors increased significantly after obfuscation. However, understanding the program complexity even decreased (against the program size).

C. Cyclomatic Program Complexity

Cyclomatic program complexity, or, as it is called, McCabe cyclomatic number, characterizes the resource input of software testing: $Z(G) = e - v + 2p$, where e is the number of arcs of a directed graph G; v is the number of vertexes; p - number of graph components.

The following evaluation grades are used:

TABLE III. CODE EVALUATION WITH Z METRICS

Z	Grade	Remark
1-5	A	Simple area
6 - 10	B	Well structured and stable area
11 - 20	C	Area of little complexity
21 - 30	D	Area of more complexity
31 - 40	E	Complicated area
41 - …	F	Very complicated, unstable area

McCabe cyclomatic number shows the required number of passes to cover all the contours of a strongly connected graph, or the required number of tests to cover each branch in the function.

The average cyclomatic complexity for the entire program will be considered, not complexity for the individual functions, as in this case general index increase or decrease is important.

TABLE IV. Z METRIC

Metric	Source code	Obfuscated code	Ratio	Direction
Z	2.333	1.063	45.56%	Fall

Average cyclomatic complexity reduction is due to appearance of a large number of functions, in which there are no conditional transitions or cycles, which are fundamental for determining cyclomatic complexity.

D. Composite Code Quality Index

This index was developed by experts from the Carnegie Mellon Software Engineering Institute. This metric can be ranged from 0 to 100 and indicates the relative complexity of code support. The higher the value of this metric, the easier to maintain a code.

The following assessment grading is used:

TABLE V. CODE ASSESSMENT BY MI METRICS

MI	Note	Assessment
0-9	Code very difficult to maintain	0-9
10-19	Code difficult to maintain	10-19
20-100	Code easily maintained	20-100

This metric is defined as:

$MI = MAX(0, (171 — 5.2 * ln(HV) — 0.23 * CC — 16.2 * ln(LoC) + 50*sin(2.4 * C)) * 100 / 171)$

HV – computational complexity (Halstead metrics);

CC – Cyclomatic complexity;

LOC – lines of code;

C – number of comment lines.

TABLE VI. MI METRIC

Metric	Source code	Obfuscated code	Ratio	Direction
MI	48.71	23.59	48.42%	Decline

The complexity of code support was increased after obfuscation, which follows directly from program size expansion according to the Halstead metric.

V. CONCLUSION

Protection of program components of a distributed computes system from research by the moving target technology presented in this Section can be applied to protect against remote unauthorized code injection attacks, and thus to enhance the security of a distributed computer system.

The proposed method of protecting software components against research has advantages as compared to the existing program code diversification technologies. These advantages include lower costs of implementing diversification by eliminating the need to use virtual computing environments or to modify the code runtime environment. Also, the presented diversification method is in some cases resistant to protect against code injection even when an adversary has full access to the original application code. The proposed method of protection from research of program components effectively solves the problem of the protection of program code in case of unauthorized code injection attacks.

The reviewed method of protection from research is of both theoretical and practical interest. Theoretically, the proposed scheme can be used in further studies to create continuously changing hardware and software systems, with the change of which the plane of intended attacks is continuously displaced.

As shown above, the method can enables this implementation to work in obfuscation mode and diversification mode for a specific class of applications in interpreted languages.

ACKNOWLEDGMENT

The research was done with the financial support of the Ministry of Education and Science of the Russian Federation in accordance with the agreement 14.574.21.0126, unique identifier RFMEFI57414X0126

REFERENCES

[1] C. Collberg, C. Thomborson and D. Low, “A taxonomy of obfuscating transformations”, Technical report #148, p.36, 1997.

[2] J. Cappaert, “Code obfuscation techniques for software protection”, Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering, Belgium, 2012.

[3] M. Popa, “Techniques of program code obfuscation for secure software”, Journal of Mobile, Embedded and Distributed Systems, vol. 3, issue 4, pp. 205-219, 2011.

[4] G. Portokalidis, A.D. Keromytis, “Fast and practical instruction-set randomization for commodity systems”, Proceedings of the Annual Computer Security Applications conf., pp. 41-48, 2010.

[5] G.S. Kc, A.D. Keromytis and V. Prevelakis, “Countering code-injection attacks with instruction-set randomization”, Proceedings of the 10th ACM conf. on computer and communications security, pp. 272-280, 2003.

[6] E.G. Barrantes, D.H. Ackley, T.S. Palmer and D.D. Zovi, “Randomized instruction set emulation to disrupt binary code injection attacks”, Proceedings of the 10th ACM conference on Computer and communications security, pp. 281-289, 2003.

[7] M. Azab, “Multidimensional diversity employment for software behavior encryption”, Proceedings of the 6th International conf. New Technologies, Mobility and Security, pp. 1-5, 2014.

[8] P. Larsen, S. Brunthaler and M. Franz, “Security through diversity”, IEEE Security and Privacy J., vol. 12, issue 3, pp. 28-35, 2014.

[9] B. Baudry, S. Allier and M. Monperrus, “Tailored source code transformations to synthesize computationally diverse program variants”, Proceedings of the International Symposium on Software Testing and Analysis, pp. 149-159, 2014.

[10] C.L. Goues, T. Nguyen, S. Forrest and W. Weimer, “GenProg: a generic method for automatic software repair”, IEEE Transactions on Software Engineering, vol. 38, issue 1, pp. 54-72, 2012.

[11] E. Schulte, Z.P. Fry, E. Fast, W. Weimer and S. Forrest, “Software mutational robustness”, Genetic Programming and Evolvable Machines J., vol. 15, issue 3, pp. 281-312, 2012.

[12] M. Styugin, N. Parotkin, “Multilevel decentralized protection scheme based on moving targets”, International J. of Security and Its Applications, vol. 10, issue 1, pp. 45-54, 2016.

[13] M. Shudrak, V. Zolotarev, “The technique of dynamic binary analysis and its application in the information security sphere”, Proceedings of 2013 IEEE EuroCon conf., pp. 40-45, 2013.

Privacy Protection in Nigerian E-Government Systems

David Alade
IS Security and Assurance Management
Concordia University of Edmonton
Edmonton, Canada
dalade@student.concordia.ab.ca

Sergey Butakov, Pavol Zavarsky
IS Security and Assurance Management
Concordia University of Edmonton
Edmonton, Canada
{sergey.butakov,pavol.zavarsky}@concordia.ab.ca

***Abstract*— The paper examines and evaluates specific features of the collection and transmission of personal information on some e-Government websites in Nigeria. It presents the adoption of a special publication on security and privacy requirement controls intended for the protection of information on e-Government web services. Among the websites assessed, only one transmitted the citizens Personal Identifiable Information (PII) data through a secured communication channel. The findings indicated that Nigerian e-Government policymakers/authorities would benefit from best practices on how they protect the collection, use, disclosure, retention, and disposal of recorded individual's personal information in their information systems. The scope of this paper was limited to only the Nigerian e-Government but could be used in other developing nations.**

***Index Terms*— Information privacy, Personally Identifiable Information, e-Government**

I. Introduction

Information privacy refers to an individual's ability to control or influence in some way processes of collection, use, disclosure, and retention of data about themselves [6]. Privacy concerns exist wherever sensitive information is collected, processed and stored either in a digital form or otherwise. With the growth of e-services in the daily life of individuals, the Personally Identifiable Information (PII) has been proven to have quantifiable value [12]. Electronic government (e-Government) refers to the use of IT by government agencies as the tool to support interactions between various government agencies and citizens (G2C) or government and business (G2B). With the increasing number of G2C services, the privacy and internet security of citizens' PII data quickly became crucial. Many countries are embracing the challenges of identity theft caused by data breaches and cyber-attacks by fine-tuning ways to protect PII of citizens and preserve the image of their government service through (potentially) effective legislations. In the world today, e-Government interactions between governments and people vary greatly within different regions, but most countries are making progress on providing greater access as revealed by 2014 UN e-Government survey [24]. In Africa, countries like Ghana, South Africa and Egypt have surpassed other African nations in privacy protection policies while one of the most populous country of the continent – Nigeria - still falls behind [2]. The improper handling or non-existent disclosure controls have been the root cause for privacy related issues [13]. Regulations and directives related to data protection are being constantly updated in developed countries, such as Canada or the United States. According to a report [4], as early as in 2002, the Government of Canada (GC) developed and designed a "secure channel" based on "Industrial Security Program" (ISP), with the aim to address G2C concerns about e-Government interactions for citizens' assurance that their sensitive information and online transactions with government web based access services is private and secured. Other listed aims are to "protect against IP network intrusions," "digital identification and authentication of individuals and business partners with which government conducts business." According to the report [19] introduction of public key infrastructure (PKI) as part of "authentication components" should serve as an example of such protection. Based on these example it can be easily seen how it is important for developing nations with a similar level of e-Government development to keep abreast any changes in the continual reassessment of compliance with data privacy and security regulations.

The Nigerian Draft on Computer Security and Critical Information Infrastructure Protection Bill was presented which contains some forceful data protection provisions that deal with identity theft and data retention [7]. However, this step does not set out any form of data protection issues such as data protection principles on PII, mandatory requirement of technical measures to mitigate data protection breaches, and adequate regulatory body to redress to individuals as can be seen for example, in Canada (i.e., Office of the Privacy Commissioner). Such regulatory bodies require organizations to implement and use framework for privacy policy principles in all information systems. These efforts should cover areas such as "… transparency, security, individual participation, purpose specification, data minimization, use limitation, data quality and integrity, and accountability and auditing" [10].

With the main goal to assess the landscape of PII protection in Nigeria and suggest potential initiatives to address the issues the objectives of this research include non-invasive content assessment on the selected e-Government websites accompanied by evaluation of processes of collection and transmission of sensitive personal information. The evaluation

is followed by recommendations from an international standard on security and privacy protection to guide the transmission of sensitive information such as PII over the network in which such information is being collected. Although the scope of this research is limited to the Nigerian e-Government websites the study could also be useful in other developing nations with similar e-Government circumstances.

II. Related Works

A. Information Privacy in E-Government Systems

The Nigerian National Information Technology Development Agency (NITDA) was established and empowered under the Ministry of Science and Technology (MoST) with the mandate to drive the development of Nigerian e-Government initiatives termed "National e-Government Strategy Limited" (NeGSt) [5]. The approach was to guide continuous advancement of digital government solutions with consistent standards, operating platforms and applications across all federal government systems and agencies. Furthermore, in a report by NITDA among its focus was to champion and implement best practice on national IT policies, presidential network project, and the National Information Communication & Education Program (NICEP), that will enable effective and efficient usage of ICT facilities and infrastructures in the country. This step by the government however did not cover all the related areas as it lacks the required regulations for the best practices in the use of internet including protection of information security and privacy. As reported by [22], 23 out of 60 Nigerian e-Government websites were hacked and defaced in 2012, including ones of MoST and the Central Bank of Nigeria. The authors confirmed lack and understanding of penetration testing as a vulnerability assessment in e-Government information systems. The increased public demand on the need to protect personal data of "mobile telephone subscribers made the Nigeria Communications Commission (NCC)" cite the Registration of Telephone Subscribers Regulation of 2011, which represents a wider perspective and attempts some protection of the data collected, collated, retained and managed by the telecommunication companies [2]. The regulations were found to be too general and poorly drafted as they tend to only address merely the purpose of levying penalty fines to the licensees, independent registration agents or any other entities that are not in compliance with the data protection [2]. According to [25], when a violation of subscriber's right to information privacy occurs, it is counted as a breach of data protection. The Data Breach Investigation Report (DBIR) recommends the data centric approach "to [manage] network security, [identify] and [classify] data," and keep rigor security control over who has access to sensitive information to mitigate the insider threats.

B. Privacy Protection in Nigerian E-Government Systems and Description of Personally Identifiable Information

As per NIST publication "Privacy, with respect to PII is a core value that can be obtained only with appropriate legislation, security policies procedures and associated controls to ensure compliance with regulations. Therefore, protecting PII that is collected, used, maintained, shared, and disposed by e-Government systems is a fundamental responsibility of all federal organizations" [17].

According to [14] the use of the internet by citizenry to acquire personal information from the e-Government information systems create great concerns on traceability, invasion and inadvertent leakage of PII; given that the network environment in which their private information is revealed are unsecured and vulnerable to the risk of identity theft threat through unauthorized third parties. In another related research [21] on balancing security and privacy in e-Government services, development of a "trustworthy system" was suggested to complement technical aspects of security. The findings in the same research state that only technical aspects do not provide a truly "trustworthy system" but rather must facilitate quantifiable, auditable technical and organizational aspects of the delivery: policies, architectures, SLAs, etc. [21]. Furthermore, according to [26], when developing e-Government infrastructures, focus must be given to "metrics, certification, standardization, governance and management as well as international agreements on interoperability to provide adequate privacy and security in e-Government web services". Same study emphasizes the security of e-Government web applications and suggests potential security mechanisms to be adopted in protecting e-Government websites from external attackers [26]. The other area that is being highlighted in related research is related to electronic identity management (eIdM) [14]. eIdM allows greater controls over each individual's choice/consent, when and whether to share PII, what is the minimum that must be shared, and the particular circumstances under which that PII can be shared. Obviously, organizations cannot have effective privacy without a reasonable foundation of information security [14] and having such mechanisms as national eIdM system may play critical role in the e-Government penetration.

Assessment of the current status of PII protection in e-Government cannot be done without defining PII itself. To be marked up as PII the collected information should fall under one of the definitions below:

- Information about a prospective, current, and former customers, employees and others with whom entity has a relationship, directly or indirectly identified individuals. Name, Home or e-mail address, identification number e.g. National Identification Number (NIN), Tax Identification Number (TIN), Physical characteristics, Consumer purchase history [9].
- Any information that can be used to distinguish or trace an individual. Any other information that is linked to an individual, such as medical, educational, financial, and employment information [8]...
- Name, such as full name, mother's maiden name, or alias Personal Identification Number (PIN), Passport number, National Identification Number (NIN), TAX Identification Number (TIN) or financial account or credit card number. Address information such as street

address or email address. Personal characteristics, including photographic image (especially of face or other identifying characteristics, fingerprints, handwriting, or other biometric data [18].

It has been indicated in many studies related to hacktivism and doxing that an aggregation of several pieces of information can produce more personalized information and therefore makes it possible to link up different fragments of data to an individual. Related study indicated trust is one of the major factors affecting effective deployment and implementation of e-Government [16]. The following section looks at the current status of PII protection on e-Government web sites to identify steps that could be helpful to increase the penetration of web based government services.

III. FINDINGS

To outline potential benefits of adopting information privacy assessment framework for e-government systems in the developing countries the authors have selected a sample of e-government web sites in Nigeria and applied the suggested privacy assessment framework to identify the potential gaps and potential remedies. The selected sample included five sites: *nigeriadriverslicense.org, jamb.gov.ng, lagosresidents.gov.ng, waeconline.org.ng , www.nysc.gov.ng g* listed in the table 2 below. The selection was based on the criteria of large audience and potential collection and processing of distinguishable and linkable PII. Large audience of these sites increases the potential risk of identity theft and other cybercrimes to the users' PII data [1]. Assessment of the web sites was done with the use of privacy related controls listed in NIST SP 800-53 Revision 4 document.

The selected sample of Nigerian e-Government websites allow citizens to register, access, track, and check for their personal data status online with their unique identification during and after registration processes. As discussed in the literature review, lack of adequate local regulations and policies on data protection led to the need to observe best practices on internet security and privacy protection as used in developed countries. The two documents used for the assessment included OWASP Application Security Verification Standard security requirements and NIST special publications on information systems security requirements. For the purpose of this paper, NIST SP 800-53 Revision 4 on security and privacy controls was selected.

The publication includes 18 security and 8 privacy family control enhancements, which are both interrelated as organizations cannot have effective privacy without a basic foundation of information security [17]. To bridge the gap in the information security and privacy of PII, the Nigerian e-Government initiative may need to adopt similar methods as adopted and used in the United States. The choice of the possible control identifiers enhancements provided in Table 1 was limited to publically accessible controls, as the authors did not have access to other internal details of the systems' configuration. For example, security awareness training (AT-2) and privacy awareness and training (AR-5) were near to impossible without having access to the system developers, owners or hosting service providers.

Table 2 lists the selected websites evaluated and outlines the information that is being collected on the web sites and processed by the respective services. The analysis of the connections and web site content indicated that only one out of five sites uses secured (HTTPS) connection and none of the sites have written policy related to the privacy of PII that is being collected and processed.

TABLE I. NIST SP 500-53 REVISION 4 SECURITY AND PRIVACY CONTROLS

Security Controls Control: SC-23 (Session Authenticity) Family: System and Communications Protection Class: Technical-Organizations ensure communications at sessions in service-oriented architectures providing web based services are secured and protected for confidence at both ends of communications sessions in ongoing identities of other parties and in the validity of information transmitted.
Privacy Controls Control AP-2 Authority and Purpose Purpose Specification- Ensures that organizations identify a particular PII collection or activity that impacts privacy and specify in their notices the purpose for which PII is collected. IP-3 Individual Participation and Redress Redress- Addresses the need to make individuals with access to PII and the ability to have their PII corrected or amended which enhance the public confidence. TR-1 Transparency Privacy Notice- Ensures organizations provide public notice of their information practices and privacy impact of their programs and activities.

*Summary description of NIST SP 800-53 Revision 4 on "Security" and "Privacy" controls can be found at; https://sites.google.com/site/nfrmtnscrt/Controls

TABLE II. SELECTED NIGERIAN E-GOVERNMENT WEBSITES.

Websites	Information Collected
www.jamb.gov.ng	First Name, Middle name, Surname, Date of birth, PIN, Registration number, Photographic image, medical/health information, postal/street address, biometric data.
www.lagosresidents.gov.ng	First Name, Middle name, Surname, Date of birth, National identification number, Photographic image, medical/health information, postal/street address, biometric data, employee details, passport number, driver's license number.
www.nigeriadriverslicence.org	First name, Middle name, Surname, Date of Birth, biometric data, medical/health information, photographic image, signatures.
www.waeconline.org.ng	First Name, Middle name, Surname, Date of birth, PIN, Registration number, Photographic image, medical/health information, postal/street address, biometric data.
www.nysc.gov.ng	Data page of the passport, Photographic image.

As shown in table 2, only the Nigerian driver's license (NDL) that allowed its users to track and access the status of their application or during renewal of their driver's license number, transmitted the citizenry's sensitive information on the web network environment through a secure communication

channel (SSL/TLS). The others respectively are not properly secured as they tend to transmit the stored personal information collected before and after registration in a plain text format. Furthermore, in similar process, among the unsecured websites, for instance, Joint Admission and Matriculation Board (JAMB) allowed its users to check for their personal data online with their unique identification. Those who may have forgotten their e-Registration number should retrieve it upon PIN validation and serial number with the provision of their surname, first name, or middle name and date of birth. By so doing, the users are redirected to a different web portal, which then displays the user's personal information. These findings indicated that the assessed web sites did not pay required attention to the issues related to PII protection and they may indicate the general tendency when the developers of e-Government initiative on the initial stages of the projects pay much more attention to the functionality but not the security of the web services.

IV. Discussion and Suggestions

A. Discussion

As can be seen in table 1, the security controls require US government institutions that collect and transmit sensitive information such as PII to deploy secured web based services and facilitate the establishment of protected communication channels (SSL/TLS) between the web clients (user) and web server (e-Government) to prevent cyber-attacks. Lack of transparency in PII protection on the web sites is surfaced as non-existing privacy policies on the websites. Such policies would typically state what pieces of personal information were collected; specifically, how they have been collected, used, shared, stored, disposed; what measures of security mechanism have been deployed to protect the stored sensitive information collected; and possibly if during redirection process, what assured the user's personal information privacy in such an unsecured/secured website. As described in NIST SP 800-53 Revision 4 "Appendix J" as standard requirement, the privacy controls require organizations that collect, store and transmit personal information over the internet to display a visible external written privacy policy statements on their websites for transparency that would reassure citizenry worried about information privacy. Although industry says that there is no "guarantee" that the secure (encrypted) websites connection is safe to use, it assures users the "integrity," "identity" and "trustworthiness" of an organization websites on the internet [15]. Consequently, users' privacy can still be compromised by the approach of how websites use personal information while delivering governmental information and services to citizens [11].

Furthermore, despite the laudable steps on e-Government development taken in Nigeria government, it can be said that the developers did not really focus on web site security, integrity and the user's personal information privacy. Obviously because these features were not included in the requirements.

Among many other possible cyber-attacks, a website that request, store and transmits unencrypted sensitive information may result in a loss of confidential information to unauthorized user. For instance, on the web site of Joint Admission Matriculation Board (www.jamb.gov.ng), after the successful submission of the login credentials to access the personal information, the user is redirected to another site. Therefore, during the communication transmission for the intended purpose, the unsuspecting legitimate user could be coerced to a malicious web page through session hijacking by an attacker intercepting the communication session in the ongoing identities of the validity of information transmitted. This would trick users to submit more private and sensitive information to carry out the variety of attacks on user identity. For identity thieves to successfully carry out their crimes, the main information they seek out is mainly the full address, mother's maiden name, driver's license number, passport number, personal identification number (PIN), national identification number (NIN) and tax identification number (TIN) [2]. Even if the public pages can be accessed via insecure connection on the e-Government web sites the parts that are dealing with PII should enforce security thorough HTTP connection.

B. Proposed Measures for PII Protection

The subsection below outlines suggestions for Nigerian e-Government websites but similar recommendations can be used by other developing nations with similar level of e-Government development to appear as nations that understand the need for information security and privacy in e-Government systems. This set of suggestions is being discussed in two parts: Technical suggestions and Policy suggestions.

1) Technical suggestions:

- The provision of adequate IT Security and Privacy standards, which will compel organizations deploying government e-services to comply with the security standards requirements as mandatory for all federal organizations such as NIST SP 800-53 in the US. This major move should present a list of system-level security principles to be considered in the early design, development or even in already deployed operation of an information system life cycle, where PII data involved.
- According to [3] for an organization to address variety of "security" issues that may arise in information systems, it is essential for the organizations not only depend on compliance/regulations but also security awareness training aimed to the intended employee. The policymakers may consider to enact data protection awareness programs for the government agencies that are processing PII as well as related contractors. The program would highlight the importance and value of PII protection mechanisms to be devised, implemented, and made available to ensure that personal data of citizens are treated and managed in accordance with best practices on data protection principles. Such initiatives may include for example, the creation of security awareness website(s), helpful hints via email or even visual aids like posters, display on logon screen/monitor when they start up and conduct of information security awareness events. These methods can help ensure employees have a solid understanding

of organization security policy, procedure and best practices [20].

• Establishment of national Public Key Infrastructure (PKI) as an enabler to support data encryption, strong authentication, data integrity and identity assurance. An example of authentication mechanism that can be used in e-Government could be introducing Comodo Personal Authentication Certificates (CPACs) as a two-factor identity authentication of genuine users logging onto websites online services, guarantees secure management and delivery of electronically transmitted confidential information to deter unauthorized access from legitimate user information.

2) Policies.

As website privacy statements specifically address the gathering of PII, the Nigerian e-Government administration may step forward to ensure services provide public notice of their information practices and the privacy impact of their programs and activities. External privacy policies should be displayed on websites for transparency. The suggestions on written privacy policy include:

• Indication on what is being collected and what is not being collected. For example, the statement may make it clear that names, mailing addresses, phone numbers, email addresses, and contact preferences can be collected but the service does not collect financial information such as credit card or bank account numbers.

• Indication on what pieces of information will be used for what purposes. For example, it may stay that service accessibility will be aligned with the age of users.

• Use of Cookies: A session cookie refers to a temporary cookie that only remains in the cookie file of a user's browser until the user leaves the site. For transparency, access to website or login to a secure area, send an encrypted session cookie to validate the user access to different parts of the site [23].

• Security of Personal Information: When personal data is stored by an organization, the organization should state the protection mechanism in-use for security of personal information during data transmission using industry standard encryption protocols such as Transport Layer Security (TLS/SSL) for information in transit or various encryption protocols for the data at rest.

• Third‐Party Sites and Services: States that the websites' services contain links to third-party websites, products, and services. If these cannot be avoided, the products and services may also use or offer products or services from third parties.

• Access to Personal Information: Ensure that the contact, personal information and preferences are accurate, complete, and up to date by logging into the account. If inaccurate access, correction, or deletion requests are made, it can be done, for example, through a privacy contact form on the website.

• Integrity and Retention of Personal Information: This makes it easy for users to keep their personal information accurate, complete, and up to date. An organization should retain personal information for the period necessary to fulfill the purposes outlined in the Privacy Policy unless a longer retention period is required or permitted.

• Changes in Privacy Policy: This policy may be amended from time to time. If an organization amends the Privacy Policy in any way, they should place an updated version on the page of the website. Suggestions to regularly reviewing policy page, ensures that citizens are always aware of what information is collected, how it has been used and under what circumstances.

V. Conclusion

For the Nigerian government to appear as a nation that understands the urgent need for information security and privacy in e-Government information systems, it will be a right step forward if the policymakers/authorities enact the provision of adequate IT Security and Privacy standards. Development of a nationwide data protection awareness programs would highlight the importance and value of personal information and help with e-Government penetration helping citizens to get better service. Other developing nations with similar e-Government landscape may adopt the same suggestions.

Similar to NIST SP 800-53 REVISION 4 such program may include control catalog, which would be directly linked to the current state of the threat space. The controls and enhancements are distributed throughout the control catalog in various families and provide specific privacy and security capabilities that are needed to support those computing technologies and computing approaches.

From the review and findings, lack of local regulations on PII protection and subsequently lack of websites security was confirmed. The observation shows that only one website use secure communication channel out of the four assessed e-Government websites in this paper. In general, none of those websites displayed PII policies. As this paper has listed, it is therefore suggestible if the respective organization websites step forward to display a written privacy policy i.e., to show and outline the specified purpose(s) for which PII is collected, and for the unsecured websites to consider web security on their websites. Suggested mechanisms may be also used in the developing countries with the same status of e-Government development.

References

[1] L. Adedayo, S. Butakov, R. Ruhl and D. Lindskog, "E-Government web services and security of Personally Identifiable Information in developing nations a case of some Nigerian embassies," Internet Technology and Secured Transactions (ICITST), 2013 8th International Conference for, London, 2013, pp. 623-629. doi: 10.1109/ICITST.2013.6750278

[2] Adeniyi, S. A."The Need for Data Protection Law in Nigeria: http://adeadeniyi.wordpress.com/2012/07/18/the-need-for-data-protection-law-in-nigeria-2/ (2012)

[3] Blevins B. "Security Awareness Training and Internal Threats-Information": http://searchsecurity.techtarget.com/news/2240234092/ Despite-skeptics-security-awareness-training-for-employees-is-booming (2014).

[4] Fraser, C. (2009). E-government: The Canadian experience. Charmaine F. "E-Government: The Canadian Experience".

Dalhousie Journal of Interdisciplinary Management. Volume 4 – Spring 2009

[5] Choudrie, Jyoti, Emeka Umeoji, and Cynthia Forson. "Diffusion of e-Government in Nigeria: A qualitative study of culture and gender." (2012).

[6] Bélanger, France, and Robert E. Crossler. "Privacy in the digital age: a review of information privacy research in information systems." MIS quarterly 35.4 (2011): 1017-1042.

[7] Franklin F. A."Data Protection Legislation for Nigeria: The Time is Now!: http://www.datalaws.com/common/pdf/Article02.pdf (n.d)

[8] GAO Report 08-536. Privacy Alternatives Exist for Enhancing Protection of Personally Identifiable Information: http://www.gao.gov/new.items/d08536.pdf (2008).

[9] Generally Acceptable Privacy Principles, CPA and CA Practitioner Version: http://www.cica.ca/resources-and-member-benefits/privacy-resources-for-firms-and-organizations/gen-accepted-privacy-principles/item10511.pdf (2009)

[10] Hugo Teufel III, Homeland Security; PRIVACY POLICY GUIDANCE MEMORANDUM: The Fair Information Practice Principles: Framework for Privacy Policy at the Department of Homeland Security: http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf (2008).

[11] Fonou Dombeu, Jean Vincent, and Nelson Rannyai. "African e-government research landscape." The African Journal of Information Systems 6.3 (2014): 2.

[12] Soma, John T., J. Zachary Courson, and John Cadkin. "Corporate Privacy Trend: The" Value" of Personally Identifiable Information (" PII") Equals the" Value" of Financial Assets." Rich. JL & Tech. 15 (2009): 11-13.

[13] Karen, A. W. e-Study Guide for Health Care Information Systems: A Practical Approach for Health Care Management 2nd Edition. (2008)

[14] Longe, O.B., and Chiemeke, S.C. Information and Communication Technology Penetration in Nigeria – Prospects, Challenges and Metrics. Asian Journal of Information Technology Vol. 6, No. 3 pp280-287 (2007)

[15] Microsoft, "How to know if an online transaction is secure"; Available online at: http://windows.microsoft.com/en-ca/windows/know-online-transaction-secure#1TC=windows-7 (n.d).

[16] Mundy, D and Musa, B. “Towards a framework for e-Government Development in Nigeria", Electronic Journal of E-Government, Vol. 8,no.2,pp. 148-161,2010 (2010)

[17] NIST SP 800-53 REVISION 4, Summary of NIST SP 800-53 Revision 4, Security and Privacy Controls for Federal Information Systems and Organizations; National Institute of Standards and Technology; Computer Security Division: http://csrc.nist.gov/publications/nistpubs/800-53-rev4/sp800-53r4_summary.pdf (2014).

[18] NIST Special Publication 800-122, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf (2010)

[19] SafeNet Report, "The Foundation of Information Security"; CASE STUDY; Canadian Government Secure Channel: http://www.safenet-inc.com/uploadedFiles/About_SafeNet/Resource_Library/Resource_Items/Enterprise_Data_Protection_(EDP)/SafeNet_Case_Study_Canadian_Government_Secure_Channel.pdf (2009).

[20] SANS Institute InfoSec Reading Room,"The Importance of Security Awareness Training: http://www.sans.org/readingroom/whitepapers/awareness/importance-security-awareness-training-33013 (2009).

[21] Sullivan, Kieran, and Jim Clarke. "Balancing security and privacy in eGovernment services." IST-Africa, 2010. IEEE, 2010..

[22] Techloy, Cyber Attacks At Nigerian Government Websites Increased By 60% In 2012 [REPORT]: http://techloy.com/2013/01/17/nigerian-government-websites-cyber-attack-report/ (2013)

[23] Use of cookies. The Chartered Institute for IT, Available online at http://www.bcs.org/content/conwebdoc/1481 (n.d)

[24] United Nations Economic and Social Affairs, E-government Survey: E-Government for the Future we want; http://unpan3.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf (2014)

[25] VERIZON DATA BREACH INVESTIGATIONS REPORT, INSIDER THREATS: http://www.tripwire.com/state-of-security/top-security-stories/2014-verizon-data-breach-investigations-report-insider-threats/ (2014)

[26] Willemson J and Ansper A., A Secure and Scalable Infrastructure for Inter-Organizational Data Exchange and E-Government Applications, "Third International Conference on Availability, Reliability and Security. 2008. ARES 08. vol. 4, pp.572, 577, 4-7 March 2008.

The Ways Assessment of Information Security in Organizations

Ganiev Abdukhalil Abdujalilovich, Yusupov Sabirjan Yusupdjanovich, Gulomov Sherzod Rajaboevich
Information Security Department
Tashkent University of Information Technologies
Tashkent, Uzbekistan
a.ganiev@tuit.uz, intdep-tuit@rambler.ru, sherzod.gulomov@rambler.ru

***Abstract*—This paper is about the ways assessment of information security, allows defines the criteria for auditing information security and performance of information security. Also the assessment model as a framework linking the need for measuring information security, information security audit specific objectives, with appropriate processes and represent the attributes processes is considered.**

***Keywords*— indicators, audit, quantitative, assurance system, vector, frequency characteristics.**

I. Introduction

Nowadays, providing information security (IS) of computer systems for various purposes continues to be an extremely serious problem and challenge. It can be noted that despite the efforts of numerous organizations dealing with this problem, the overall trend remains negative. There are two main reasons for the increasing number of incidents in the field of information security, with dire consequences for large organizations according to:

- increasing role of information technology in support of business processes, as a result of increasing demands on information security of automated systems. Price errors and failures of information systems is increasing;
- increasing complexity of information processes. This places high demands on the qualification of personnel responsible for ensuring information security. Selecting the appropriate solutions, ensuring an acceptable level of information security at an acceptable cost level, it is becoming increasingly challenging.

II. Evaluation of Information Security on the Basis of Indicators

The decisive element in the process of auditing IS organizations and systems is a model of assessment processes to ensure IS. The basis of valuation models and make a list of the model evaluated processes and a set of indicators that are used to collect data and to determine the extent to which attributes the process of establishing criteria for IS audit.

Assessment model describes how quantitatively measured attributes and how they are converted into indices, providing a basis for decisions about the degree of compliance with the criteria of IS audit, the extent of the right processes to ensure IS organization system [1]. In general, the model assessment processes to ensure IS organization can be represented by the structure shown in Fig.1.

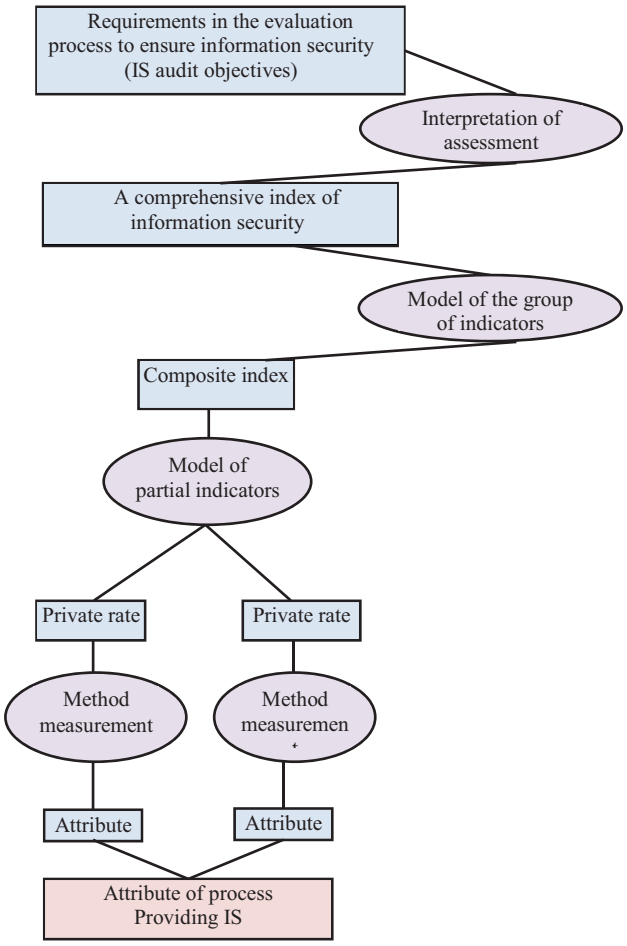


Fig.1. Process Assessment model for information security organization

III. Audit of Information Security Organizations and Systems

The method of measurement (assessment) - is a logical sequence of operations, which is used for measuring (estimating) attributes relative to a certain scale [3]. Operations may include activities such as event counting or monitoring of the passage of time. One and the same measuring method can be applied to many attributes.

Type of measurement method depends on the nature of the operations that are used to measure the attribute. One can define two types:

- subjective - quantification involving human judgment;
- objective - quantification based on calculations.

Possible examples include methods of measurement survey, observation, survey, assessment of knowledge, testing, data sampling. Some measurement methods may be implemented in many ways. Used measurement procedures reflect the specific implementation of the measurement method in the organization, this system.

Measuring method translates the value of the measured attribute to a value of the measurement scale. scale view depends on the relationship between character values on the scale. Examples of the types of scales:

- nominal: the measurement values are categorical. For example, the classification of defects by type does not imply having order among the categories;
- ordinal: the measurement values are ranked. For example, the distribution-division defects by severity is the ranking;
- interval: the measurement values have equal distances corresponding to equal quantities of the attribute. A value of zero is impossible;
- scale relations: the measurement values have equal distances corresponding to equal quantities of the attribute, where zero corresponds to zero attribute.

The measurement method generally affects the choice of type of scale that can be reliably used in the measurement of process attributes.

Index – is a measure that provides assessment of certain attributes using the selected measurement methods. Indicators are the basis for analysis and decision making. This is something that should be provided to users of measurement. The measurement is always based on imperfect information, so that the determination of the uncertainty the accuracy or relevance of indicators is an important component of the representation of the actual value of the indicator.

Consider the pricing model based on process performance indicators IS organization assurance system. W Indicator of functioning process is a measure of the extent to which the actual result of the process required.

The main requirement when selecting indicator operation is appropriate indicators objective process, which displays the requirement result $Y^{\text{Requirement}}$. For a description of the relevant real outcome Y of the process required to formally define a numerical function on a set of process results:

$$P = p(Y, Y^{\text{Requirement}}) \qquad (1)$$

which is a function of conformity, showing the degree of achievement of the process purpose. Thus, the indicator functioning process can be represented as:

$$W = p(Y, Y^{\text{Requirement}}) \qquad (2)$$

Internal audit of the information security of organizations and systems. However, in order to function (1) can be considered as index operation, in addition to the requirements of fitness for purpose of the process, it must meet the following requirements: content, measurability and interpretability. Pithiness means that when evaluating the index takes into account all the essential characteristics and properties of the process.

Interpretability is needed to understand the evaluation results. Measurable means that there is a method of measuring the indicator that provides reliable information and trusted method of estimation.

If the process is evaluated by a certain number of its attributes, the vector is introduced (generalized) measure performance, combining private data:

$$W_0 =< W_1, W_2, \dots, W_m \qquad (3)$$

where $W_i\ j = 1, m$ it is determined by (2) with the setting in place of y, $Y^{\text{Requirement}}$ quantities y, $Y^{\text{Requirement}}$ frequency characteristics of the process, i.e. $W = p(y_i y^{\text{Requirement}}), j = 1, m$.

Introduction of a vector operation indicator imposes additional requirements: Minimum number of private indicators and completeness.

Requirements minimum number of private indicators associated with the desire to reduce the complexity of evaluation, however, while maintaining full coverage characteristics and properties (attributes) of the process. Typically, a vector component is administered in cases where the process is achieved resolution of several problems, the effectiveness of solutions each of which is estimated corresponding private exponent $W_i,\ j = \overline{1, m}$. The size of the vector index is the number of estimated process attributes. Private indicators may have different dimensions. Therefore, the formation of a generalized indicator is necessary to operate with normalized values of parameters that need to match them [4]. The value of the private exponent can be represented as a percentage or share.

Measurement of IS can be based on imperfect information, so determining the accuracy or relevance of indicators is an important component of the representation of the actual value of the indicator. The accuracy of the measurement depends on the selected method, the data source and accuracy of the data provided. Subjective measurement methods depend on the expert interpretation of process attributes. The accuracy estimates may be improved if, in addition or instead of the numerical values of parameters used processes provide IS. For example, to estimate the private exponent can be applied method of calculation is to determine the proportion of employees, the assessment of skills and professional assessment of the suitability of which is carried out regularly.

In this case, private parameters are included in the metrics in the context sources and IS audit certificates, and index calculation method, for example, be presented in a table 1.

TABLE I. THE METRIC FOR A PARTICULAR INDEX

Parameter	Description
Private rate IS	Personalized if all roles in the organization and whether the responsibility for their implementation is set?
Measurement method	Share roles and personalized for the execution of which is set liability.
Evidence of self-esteem IS	It is documented to define all roles that exist in the organization? It determines whether the relevant instructions of staff responsible for the execution of roles? Is there documentary evidence familiarize staff with their duties for the execution of roles? How many personalized roles and their pursuant to establish liability? How many roles exist in the organization?
Method of calculation	The number of roles which are personified, and for the execution of which is set liability/number of roles in existing organizations.
Sources of evidence IB assessment	Information security policy credit organization. The provisions of the roles. Provisions on the allocation of responsibility. Orders on the appointment, the division of responsibilities between the staff of the organization.
Indicator	The purpose of indicator

The more indicators to computationally evaluate interesting processes attributes, the higher may be the objectivity of the evaluation process with subjective measurement methods [5]. With the help of private indicators IS valued attributes of processes to ensure information security systems, and with the help of generalized indexes IS processes are evaluated to ensure information security system. Model of particular indices - an algorithm or calculation of connecting private performance by a certain rule. The rule should be based on an understanding of or assumptions about the expected relationship between the private performance.

That rule may be the allocation of significant private performance by assigning to them the significance of the coefficients. The importance of particular indicators is determined by the degree of influence on the process attribute of the process result. In this case, the composite index is calculated as follows:

$$W_0 = \sum a_i x W_i, \quad (4)$$

where $a_i -$ boost factors of particular indices W_i;

$$\sum_{i=1}^{m} a_i = 1; \quad (5)$$

$m -$ number of individual indicators W_i summarized indicator W. Model of particular indicators may be based on utility theory, when a method of convolution vector (generalized) indicator with the preference system.

Association rule can also be based on the system preferences of some partial indicators over the other, which makes it possible to evaluate the process, focusing on the particular indicators chosen.

For example, if the installed system preferences indicates preference for the private index W_1 over W_2 and W_2 over W_3 ($W_1 > W_2 > W_3$), the process, the estimated generalized indicator can have an estimate equal to the most preferred private index (in this case W_1).

Model of generalized indicators also represents an algorithm or calculation, connecting generalized indicators for a particular rule. This rule can also be based on the preferences of the system.

In this case, the evaluation process will reflect the totality of the preferred assessment generalized indicators. The merger of generalized indicators IS is obtained an integrated indicator of the IS organization and system.

When selecting and shaping private, generalized and complex indicators should be taken into account, such as, the following criteria:

- the feasibility of data collection;
- availability of human resources for data collection and management;
- ease of data collection;
- the degree of interference in the activities of the staff;
- availability of appropriate tools;
- ensuring confidentiality;
- potential resistance data providers;
- ease of interpretation of the indicator customers and evaluation specialists.

IV. CONCLUSIONS

Interpretation of assessment results is an explanation that relates a quantitative assessment of indicators to measure the needs of the software processes of information security in the language of users of measurement results. Such an interpretation may reflect, for example, breach of information security properties, possible negative impact on the organization or functioning of the results assessment systems.

REFERENCES

[1] Dunkerley, K., & Tejay, G. (2012). The development of a model for information systems security success. In Z. Belkhamza, & S. Azizi Wafa (Eds.). Measuring organizational information systems success: New technologies and practices (pp. 341 366). Hershey. PA: Business Science Reference, doi: 10.4018/978-1-4666-0170-3. chO 17.

[2] Federal Information System Controls Audit Manual (FISCAM), Government Accountability Office, GAO-09-232G, February 2, 2009. Available: http://www.gao.gov/new.items/d09232g.pdf

[3] Catalog of Minimum Acceptable Risk Controls for Exchanges – Exchange Reference Architecture Supplement, Draft, Version 0.99, CMS, March 16, 2012.

[4] Gorniak, S., Ikonomou, D., Saragiotis, P. et al., Priorities for Research on Current and Emerging Network Trends. European Network and Information Security Agency. 2010.

[5] Hyoungshick Kim, Jun Ho Huh, and Ross Anderson. On the security of internet banking in South Korea. Technical Report RR-10-01, March 2010.

A New Approach to Detecting and Preventing the Worm Hole Attacks for Secure Routing in Mobile Ad-hoc Networks based on the SPR Protocol

Seyed Ali Sharifi
PhD. Candidate, Department of computer engineering
University of Kashan
Tabriz, Iran
miralisharifi@yahoo.com

Seyed Morteza Babamir
Associate Professor, Department of computer engineering
University of Kashan
Kashan, Iran
babamir@kashanu.ac.ir

***Abstract*— Ad-hoc mobile networks are free of any specific infrastructure in which, their nodes act in a moving and totally autonomous form and communicate with each other through radio waves. Communication among nodes is subject to change at any moment due to the emergence of new nodes in the network. Therefore, the security of these networks is of great importance. In this paper, we study different types of routing protocols, security problems in the routing of ad-hoc mobile networks, and various types of secure routing protocols, along with their advantages and disadvantages. By applying some changes over the SPR protocol, we attempt to solve problems associated with the Worm Hole attacks using a new combination of packet leashes techniques. That is by adding another technique as status leashes to the SPR Protocol which is based on the encryption information and security features, the exchange of packets in the route is tracked and controlled. Moreover, for key management and security enhancement we used Diffie-Hellman algorithm and finally for the implementation of the proposed method Matlab as a tool and C# programing language were used.**

***Index Terms*— Ad-hoc mobile networks, Secure routing, Security attacks, Key management, SPR protocol**

I. INTRODUCTION

Ad-hoc networks are independent sets of nodes which do not have any specific infrastructure. An ad-hoc network is divided into two categories of wireless sensor networks and ad-hoc mobile networks. The nomenclature behind the ad-hoc mobile network is its mobility and autonomy in nodes. For the case of ad-hoc communication, integrated networking has a low performance so in this case, the ad-hoc network is an excellent alternative. In the real world, networks continuously change, and alter their links to connect to new nodes, and this is why we are interested in these networks. Despite their security problems, ad-hoc mobile networks have many applications. In fact, their performance increases day by day and their price falls with implications for their market popularity.[1] Also, in the areas where communications' infrastructure does not exist or the existing infrastructure costs very expensive, wireless mobile users can communicate with each other through ad-hoc mobile networks. Each node in order to establish communication is equipped with a transmitter and a receiver, which communicates with the remaining nodes via radio waves in two methods: the peer to peer method and the broadcast method. In this network, the nodes have no knowledge regarding the network connections in which they locate, because of the dynamic structure of these networks. In order to send information to other nodes, each node must detect the route and keep the its track.[2].

II. ROUTING AND THE DIFFERENT TYPES OF ROUTING PROTOCOLS

Routing is an important challenge in the ad-hoc mobile networks and due to the dynamics of the environment the routing methods must comply with these conditions. With this type of networks, the route determination is a challenge, because each node can have random motion and sometimes can go outside the network. The ad-hoc mobile network is a connected network which uses several interface nodes to reach from one node to another.[3] Communication among nodes is done using different routing methods and in this way, a number of criteria, including delay reduction and increase in the Packet Delivery Fraction are considered. Some types of routing algorithms in ad-hoc mobile networks such as DSR, AODV, DSDV, and ZRP algorithms, and their methods of action along with their advantages and disadvantages are shown in Table I.

III. SECURITY PROBLEMS IN ROUTING MOBILE AD-HOC NETWORKS

Security issues in ad-hoc mobile networks are crucial because in these networks, the communication process is done wirelessly with the possibility of passive or active intrusion. Also, as the nodes themselves operate in the routing process, the penetration of a hostile node can totally destruct the network. Security in ad-hoc mobile networks revolves around key management, secure routing, authentication, verification, preventing attacks aimed at blocking service, and prevent misbehavior and detection of penetration.[8]

Attacks in ad-hoc networks can be classified in different perspectives. In the first categorization, the attacks can be in the form of internal attacks and external ones. Internal attacks

TABLE I. THE DIFFERENT TYPES OF ALGORITHMS, THEIR METHODS OF ACTION THE ADVANTAGES AND DISADVANTAGES OF EACH

algorithm	Protocol type	Method of action	Advantages	Disadvantage	Ref
DSR	demand -based	Consisted of two phases of route detection and keeping. first the source studies the available routes to the destination node, if no route is found the route discovery phase begins. The nodes of ad-hoc networks store all the information they know in their memory. During the process of discovering route the node can normally connect to other nodes. All saved routes have an expiration date after which the memory is cleared. If a node detects the transmission is failed it clears the related route from the memory. Likewise, if a middle (interface) node detects that it cannot complete the transmission to the next node, it sends an error message to the source to remove that route from the list of known routes .	Routing is simple and effective and certainly has solution	The network load is high-a lot of bandwidth is consumed-increased number of headers when the distance between the source node and the destination node gets longer.	[4]
AODV	demand based	In this algorithm routes are discovered only when needed and are kept only during the time of being used. This Protocol also is composed of two phases of route detection and route keeping. The route detection process is done in the flooding form. In this algorithm in order to correct the pre-made routes a sequence number is used in route request packets. Furthermore this algorithm has another message named as (RERR) which is sent in case of routing error.	Minimum control overload-minimal process overload - dynamic and multi-step routing capability – dynamic maintenance of topology	Enhanced redundancy in request, response and error packets	[5]
DSDV	table based	based on the classic Bellman Ford algorithm 9. In this case, each node keeps the list of all destinations and also the number of steps to any destination. Each list entry is numbered accordingly	Avoids routing loops in networks including moving routers.	This Protocol requires parameters such as information update time intervals and number of needed updates.	[6]
ZRP	The combinational	The total area is divided into different areas. The radius and size of the area does not depend on distance of, but on the number of mutations. In order to identify the neighboring nodes, ZRP uses MAC protocol and to identify nodes in the area's neighborhood uses the route detection Protocol.	used for mobile networks with multiple displacements in long time-increased the effectiveness quality of route detection using the ask / answer mechanism	For small areas with fast-moving nodes and where route demand is high, is not suitable	[7]

are made by the allowed nodes inside the network and usually are difficult to be prevented. External attacks are made by one or more nodes outside the network and most of security measures are carried out and applied against such attacks. The second categorization is from active and passive points of views. In passive attacks, the attacker only listens to the channel information or eavesdrops but an active attacker (intruder) not only has access to the channel information but also can alter the information flow as well.[9]

IV. THE FAMOUS ATTACKS OF MOBILE AD-HOC NETWORKS:

A. The worm whole attack

This type of attacks is specific to ad-hoc networks. In this attack, two hostile nodes, in cooperation with each other create a short connection in the network topology. Then, when one hostile node receives a routing request from another node, this hostile node posts the request via a private network to the second hostile node. Now, if these two nodes do not change the step counter of routing request, a majority of the route is passed by the private network without increasing the step counter. Thus, that packet reaches at the destination only with two counters instead of dozens of counters and this route is certainly considered as the shortest route. This problem can be solved using the Packet Leashes.[10]

B. The Rushing Attack

In this type of attacks, suppose that during a route detection operation, received packets from attackers, are the first received packets by neighboring nodes of the destination node. In this case, all other requests received later by these nodes are ignored and it only sends the route request received from the attacker to the destination. This culminates at a situation in which the obtained routes have at least one attacking node. [11]

C. The black hole

in this kind of attacks the hostile node penetrates into the network, locates on the shortest route to the destination, gets the packets sent from this route, removes them and reduces the rate of packet delivery.[12]

V. THE SECURE ROUTING FOR THE AD-HOC MOBILE NETWORK

Good routing in the ad-hoc mobile network should lead to the correct detection and saving of a route so that the hostile nodes will fail in intruding in (and preventing from) the formation and keeping of correct routs. In the Table II, different types of security protocols in ad-hoc networks routing are tabulated.

TABLE II. THE DIFFERENT TYPES OF SECURITY PROTOCOLS

Protocol type	Method of action	Advantages	Disadvantage	Refere nce
SPAAR	uses GPS to determine the current position of nodes -each node can only receive the transmitted packets from its single step neighbors –if it not be aware of the position at all, the flooding algorithm. Is used	protected data against nodes that not authenticated	Failure to receive packets sent by multi-step neighbors	[14]
SEAD[3]	this Protocol is based on the DSDV Protocol where each node has a routing table including the list of all possible destinations in the network. To update the routing table, each node sends a routing request message to all its neighbors in order to enter new routes in the table-to create the security it uses unilateral hash functions instead of non-symmetrical cryptographic functions	There is no ring on the route when adding a sequence number to each of the routing table entries	The rashing attack-problem in key agreement	[10]
SRP[2]	This Protocol is based on DSR routing algorithm and a six-word part is added to its header it that in addition which includes ID and the sequence number as well as the message authenticate code. This Protocol will consider a security dependency between the source and destination nodes & based on it the source node detects received responses for this operation and ignore them	guaranteed comprehensiveness and message authentication	There is no approach to track the packets on the network and along the route passed by them – probable failure in understanding the network topology imposed by hostile nodes (worm hole	[13]
SAODV	founded to secure AODV protocol and uses hash functions – uses step counting to measure the number of steps passed by a packet is the counted steps outnumber from a pre-defined max count the packet will be ignored	Resistant against the black hole attacks	Specifying the value for the max count variable which a packet can pass	[5]
Ariadne	used to secure the DSR Protocol - instead of using the public key it engages the symmetric encryption and the authenticate code to verify the message authentication which is built by a hash function	Secure against the worm hole attacks	The need to exchange encryption keys among the network nodes before starting the Protocol	[9]
ARAN1	used to create security in the AODV protocol -based on public-key encryption, as well as the use of digital certificates – consists of the certification process - and guarantees end-to-end acknowledgment -in this method the source disseminates a route detection message to be answered by the destination node in a unicast mode so that routing messages, both at each step along the route from source to destination and also on return routes are authenticated	Resistant against counterfeiting attacks to ID's	Suffering from the lack of resistance against the black hole attack -high energy and CPU consumption - identical response to any unpredictable behavior either from hostile nodes or from a friend node which does not function correctly	[8]
LHAP	This Protocol is based on a medium layer between the network layer & MAC and uses TESLA techniques to authenticate the messages - in this technique a hash function is utilized. In this Protocol, a random key will be considered and the remaining chain keys are replicated from it.	Fully respected message authentication	The problem against the worm hole and internal attacks – fails to receive the key updates for more than one time period ,	[15]

[1] Authenticated Routing for Ad hoc Networks
[2] Secure Routing Protocol
[3] Secure Efficient Ad-hoc Distance vector routing protocol

VI. THE PROPOSED METHODOLOGY

One of the most famous attacks on the routing of ad-hoc mobile networks is the worm hole attack where several hostile nodes may penetrate into the network and in cooperating with each other impose changes in the network topology and when the authorized nodes send the route request packet in the network (RREQ), this packet is delivered to a hostile node during the transmission process and this node resends it through a wired or wireless private network to other hostile nodes without changing the step counter of the packet route request. So the step counter field in the packet has the lowest value and therefore this implies for the existence of the shortest route, so this route is selected. According to the Table 2, the action method of each of these protocols along with their advantages and disadvantages are specified. In this paper, we attempt to change the SPR Protocol and solve the problem of the worm whole attacks by using new combinations of the Packet Lashes techniques. One way to avoid attacking the worm hole is using the Packet Lashes technique.

In this technique, some information will be added to the sent packet first, in order to prevent sending packets for distances more than a threshold and second to prevent the worm hole attacks for distances less than a threshold. This technique has two categories.

Temporal Leashes technique:

This technique is based on the precise synchronization of the source and destination nodes relying on the use of a time stamp inside the packets. Thus, with the reduction of the time stamp in value, the time duration the packet has been in travel is estimated.[10]

The technique of Geographical Leashes

This technique is based on the Geographical information in which , due to speed limitations, the destination node can measure the estimated distance of the source node from itself and therefore prevent unreasonable routes.[10] In this paper, we add another technique as the technique of Status Leashes to the SPR protocol based on the status of each packet related to the security features and encryption information in which the timestamp value and the physical address tag of the transmitter node of the route request packet are encrypted by a private key and all these values are submitted inside the header packet. Using this technique, the exchange packets in the route are tracked and controlled. In case of receiving the route request packet by any of the intermediate nodes, by using the send and receive timestamps and calculating the difference between these times and according to the transmission speed of the packet in the media, the delay of the packet is extracted from Equation. (1)

$$D_{TOTAL} = H.(D_{SWITCH} + D_{QUEUING} + D_{PROPAGATION}) \quad (1)$$

Where h is the step number and D_{switch} is the switching delay in route selection in each node and $D_{Queuing}$ is the queuing delay for to be processed and $D_{Propagation}$ is the Propagation delay of the media.
Then, Equation (2) will be examined.

$$(T_r - T_s - D_{total}) <= h * tr_{max} \quad (2)$$

If equation (2) is established, it means that all existing nodes in the route are valid, and the packet will be accepted and this route will be recognized as the safe route while if the abovementioned equation is not established, there should be hostile nodes in the discovered route, so it is not safe and hence the packet will be ignored.

$$A \rightarrow B: g^a \bmod p = X \quad (3)$$
$$B \rightarrow A: g^b \bmod p = Y$$
$$A \rightarrow B: Y^a = g^{ab} = K$$
$$B \rightarrow A: X^b = g^{ab} = K$$

For key management and security in the proposed method the Diffie Hellman algorithm is used. In this algorithm, taking A and B as nodes the agreement on key is realized through the equation (3)
In the abovementioned equation, by having p, g, X, and Y, only a and b cannot be calculated and this problem is known as the discrete logarithm problem
Also by having only p, g, X, and Y, the key (K) cannot be obtained and this problem is known as the Diffie Hellman problem.
The workflow of the proposed procedure is shown in the Fig.1

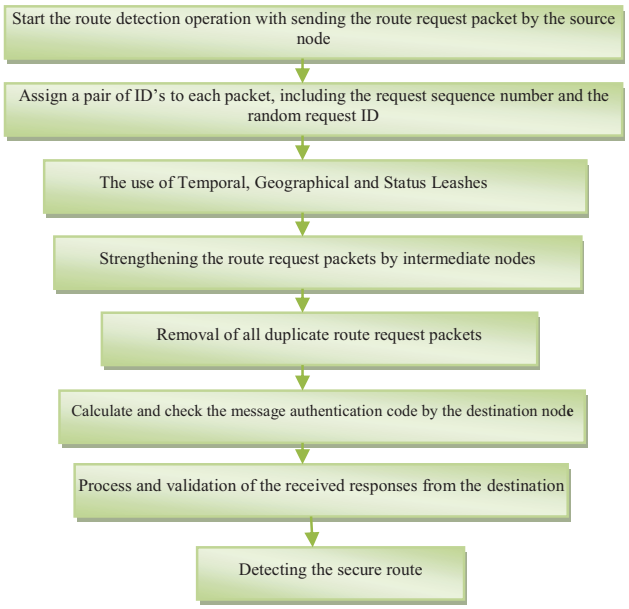


Fig.1. the workflow of the proposed procedure

VII. The Simulation Results:

For simulation the MATLAB as a tool is used and the simulation parameters are shown in the Table III.

TABLE III. The Simulation Parameters

Parameters	Value
Number of Mobile Nodes	50
Number of Worm Hole Nodes	2
Time	300m
Topology	1000m*1000m
Routing Protocol	DSR
Transmission Range	200m
Packet Size	512 bytes

In the following diagrams, ADSR is the proposed method.
In Fig.2, the impact of the mobility speed on the packet delivery fractions of nodes in the proposed method is demonstrated.

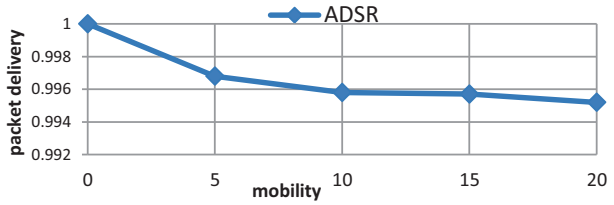


Fig.2. the impact of the mobility speed on the packet delivery fraction

In Fig.3 with the addition of a new technique called the status leashes technique overload traffic changes has been shown.

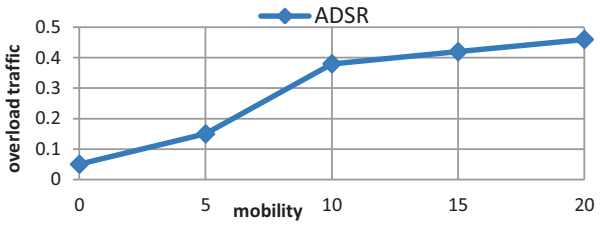


Fig.3. the impact of traffic on the resultant overload from the leashes

Fig.4 executes the ADSR algorithm and shows the effect of node speed on the number of route request messages in which by increasing the mobility speed of the nodes the number of packets rises drastically.

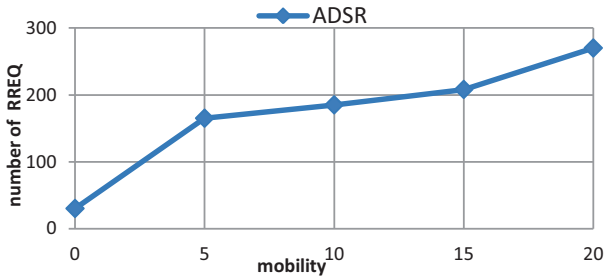
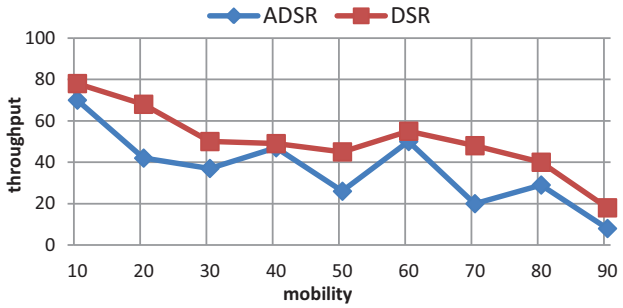


Fig.4. the impact of speed on the number of route request messages

As seen in Fig. 5, the throughput of the ADSR algorithm is raised compared to the DSR algorithm because in the ADSR

algorithm, malicious nodes are detected and ignored in the route.

Fig.5. the impact of the worm whole attack on the network Throughput



In Fig.6, the impact of the worm whole attack on the packet delivery fraction of the network is shown; in spite of malicious nodes, this fraction is increased.

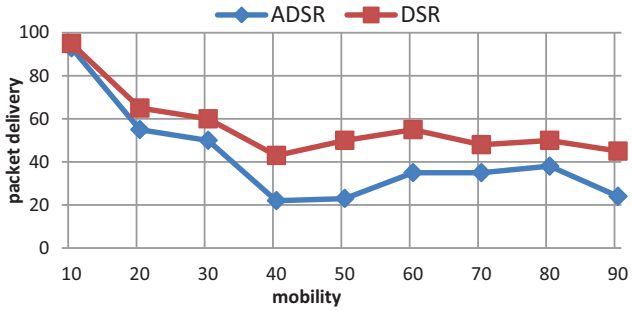


Fig.6. the impact of the worm whole attack on the packet delivery

Given that the existence of malicious nodes in the network causes the formation of short routes and because these nodes cooperate with each other to build a local network so the end to end delay in the DSR algorithm is reduced when compared with the ADSR algorithm. Fig 7 shows the impact of the worm whole attack on this delay

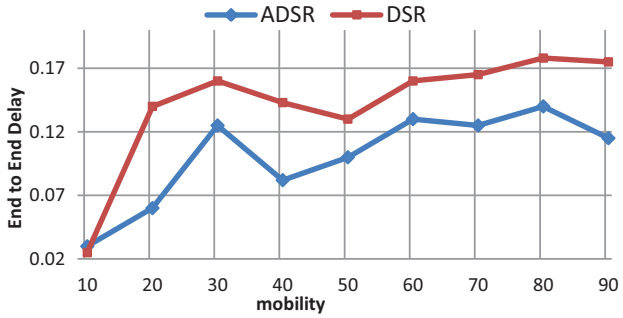


Fig.7. the impact of the worm whole attack on the end to end delay

VIII. Conclusion

As ad hoc networks are widely used nowadays in many applications, security is essential in these networks. Therefore, their security should be guaranteed. Recently, different methods have been introduced to make these networks secure. In this paper, firstly, routing algorithms and secure routing are studied followed by studying the effect of Worm Hole attack over DSR based on the DSR routing algorithm. Finally a new method of detection and prevention of worm whole attack in secure routing of ad hoc

Mobile based on The DSR protocol is suggested. And after finishing the simulation in Matlab, it was tested with various criteria. According to obtained results, it can be concluded that the ADSR algorithm outperforms the DSR algorithm against the attacks of Worm Hole.

References

1. Kale A., Ruchia Ms., and Gupta SR., “An Overview of MANET Ad Hoc Network,” International Journal of Computer Science and Applications, vol. 6, no. 2, pp. 257-264, 2013.
2. Singh G., and Singh J., “MANET: A Study of Challenges and Routing Principles,” International Journal of Advanced and Innovative Research, vol. 1, no. 1, pp. 514-523, 2012.
3. Bakshi A., Sharma A.K., and Mishra A., “Significance of Mobile Ad-Hoc Networks (MANETS),” International Journal of Innovative Technology and Exploring Engineering, vol. 2, no. 4, pp. 899-912, 2013.
4. Tai W.Y., Tan C.T., and Lau S.P., “Towards Utilizing Flow Label IPv6 in Implicit Source Routing for Dynamic Source Routing (DSR) in Wireless Ad Hoc Network,” Computers & Informatics (ISCI) on IEEE Symposium, pp. 101-106, 2012.
5. Cerri D.,Ghioni A., “Securing AODV: the A-SAODV secure routing prototype” Communications Magazine in IEEE, pp. 120-125, vol. 46, Issue 2, 2008.
6. Perkins C. E., and Bhagwat P., “Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers” Conference on Communications Architectures, Protocols and Applications, pp. 234-224, 1994.
7. Nicola Milanovic et al, “Routing and Security in Mobile Ad Hoc Networks“,IEEE Computer, vol. 37, no. 2, pp. 61-65, 2004.
8. Bridget Dahill et al, “A Secure Routing Protocol for Ad Hoc Networks,” MobiCom 2002, Atlanta, Georgia, USA, September 23-28, 2002.
9. Yih-Chun Hu and Adrian Perrig, “A Survey of Secure Wireless Ad Hoc Routing,” IEEE Security and Privacy 2004, Editorial Calendar, vol. 2, no. 3, PP. 94-105, May/June 2004.
10. Stefano Basagni et al, Mobile Ad-hoc Networking, IEEE press, John Wiley and Sons publication, PP. 329-354, 2004
11. Yih-Chun Hu, et al, “Rushing Attacks and Defense in Wireless Ad Hoc Network Routing Protocols,” Proceedings of the 2003 ACM workshop on Wireless security, San Diego, USA, pp. 30-40, 2003.
12. S. B. S. Jamali, "A Survey over Black hole Attack Detection in Mobile Ad hoc Network," International Journal of Computer Science and Network Security, vol. 15, pp. 44-51, 2015.
13. Panagiotis Papadimitratos and Zygmunt J. Haas, “Secure Routing for Mobile Ad hoc Networks,” SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, TX, January 27-31, 2002.
14. Stephen Carter and Alec Yasinsac, “Secure Position Aided Ad hoc Routing,” Proceedings of the IASTED International Conference on Communications and Computer Networks (CCN02), Nov 3-4, 2002.
15. S. Zhu, S. Xu, S. Setia and S. Jajodia, “LHAP: A lightweight network access control protocol for ad hoc networks,” Ad Hoc Networks, vol. 4, no. 5, pp. 567-585, Sep. 2006.

A Cryptography Approach on Security Layer of Web Service

Arezoo Mirtalebi
PhD Candidate, Department of Computer
University of Kashan
Kashan, Iran
mirtalebi@grad.kashanu.ac.ir

Seyed Morteza Babamir
Associate Professor, Department of Computer
University of Kashan
Kashan, Iran
babamir@kashanu.ac.ir

***Abstract*— Nowadays, web services turned in to a hot spot for hackers as they include valuable information. WS-Security is a standard that states a strategy and specifications to bring different security technologies together as a result of investigations about the safety of web services, especially, the confidentiality of exchanged messages. In this paper, we concentrate on the security challenges of WSDL file of web services and present a new cryptography approach to secure this via WS-Security standard. As an unwelcome side effect, the response time increase in deals with such services, so applying this security layer is suited for web services that provide specific types of services and the information contained in their WSDL files are valuable resource for attackers.**

***Index Terms*— Web Service Security, Cryptography, WSDL Threats.**

I. Introduction

By increasing the use of Web Services more threats would be planed to attack them and therefore more security policy is needed to defend Web Services against hackers. XML[1] encryption and XML signature are introduced by W3C[2] in 2002, to protect the XML documents against hackers by use of cryptography. In 2007, the outcome of investigations about the security of web service has been gathered in WS-Security standard[1]. This standard concentrates on the security of messages and use XML encryption and XML signature to get confidentiality and integrity[2]. WSDL[3], which is a document that provide the needed information for calling methods of a web service[3], is a vulnerable point of web services. In some cases, hackers pars WSDL to obtain necessary information in order to attack to web service .In this paper, a new cryptography layer on WSDL has been offered to enhance the security of this file. It should be noted that we have presented an article about this issue already in [4]. Now, we tend to complete it in this paper.

[1] eXtensible Markup Language
[2] World Wide Web Consortium
[3] Web Service Decryption Language

II. Statement of the Problem

This section, first, defines the security by considering different aspects, then explains the kinds of attacks to the Web services, and finally, discusses the security problem that this article focuses on comparing with related works.

A. Security

Security is examined from different angles. In actual and practical perspective, the security should cover two categories[3]:

- Functional aspects: This part of the definition applies in observing the security standards and used in traditional applications. These are as follows:
  - Authentication or identifying the consumers.
  - Authorization or deciding on the authority to do an act, by an identified consumer, on a specific source.
  - Data confidentiality or to protect the data.
  - Data integrity or detecting the falsification in data and ensure that any sender or receiver cannot modify the data.
  - Protection against attacks or ensures that the attackers cannot control the data.
- Non-functional aspects: these features, in contrast to functional, are used to ensure that service-oriented solutions work properly in organizations. Some of them are the followings:
  - Interoperability: this concept is particularly for service-oriented architecture and implies that different security solutions should not violate the adjustment of the services.
  - Simplicity in the development: this aspect is common to all security solutions. As higher the complexity of the development of solution gets, the adjustment with the related architecture will be less.

B. Threats to web services

Ten threats have been identified to web services that hackers use these threats to attack web services[5]. These include: Coercive Parsing, Parameter Tampering, Recursive Payload, Oversize Payload, Schema Poisoning, Parsing WSDL, External Entity Attack, Routing Detours, Replay Attack, and SQL Injection.

This article has provided a way to repel the WSDL parsing attacks. WSDL contains information about methods which offers in web services. In this type of attack, hackers may be able to obtain secret information about web services by examining all the given data and analyzing them. For example, by discovering the flow of called methods, it can reach the other methods that have not been registered in the public list of procedures [6].

C. Related Works

The article named "Assessing the security of web service frameworks against Denial of Service attacks"[7] tries to protect web service against denial of service attacks, but didn't talk about other types of attacks. In a study named "possible attacks on Web services" [8] general suggestions for dealing with each of these attacks have been reported. This survey has proposed encryption technology to stop hijacking the parameters. In addition, the limiting of WSDL has been suggested in order to confront its attacks. Another idea in relation to securing WSDL is an article named "web services security"[9] which tries to propose a comprehensive model to improve the security of all these three technologies, includes UDDI[1], SOAP[2] and WSDL, using encryption to provide confidentiality and hiding important information that WSDL contains, but it does not explain details of implementation. One of the models recommended for improving web service security that is suggested in other studies, is the model known as IAPF[10] which proposed an abstract solution to protect UDDI, WSDL and SOAP.IAPF suggests to enhance the security of WSDL by limiting the access of users to UDDI.

In this article we will offer a practical approach to protect WSDL by applying cryptography on critical parts of WSDL to minimize the limitation on it.

III. Suggested Approach

To create a security layer on WSDL, we should encrypt some specific XML tags by cryptography algorithms which prepared in XML encryption. Then, the encrypted WSDL is available in UDDI to be accessed by users. UDDI is a repository that makes it possible to submit, register and search web services. All service providers have to register their information in UDDI and requesters use this repository to find the needed information of desired service[11]. WSDL must be encrypted before registering in UDDI. For this purpose, web service provider sends its WSDL to trusted service which is a service for hiding the complexity of cryptography steps. This web service generates a pair key and registers it in a repository such as XKMS[3] which is responsible for managing the keys, then, encrypts the WSDL using an asymmetric algorithm. By encrypting the WSDL, the security layer is generated and it is protected against threats. In this step, the registering request is sent to UDDI.

Consumers use UDDI to find the URL[4] address of WSDL when they want to call a web service. Referring to the address of WSDL, it would be encrypted and user can not access to the content of WSDL unless decrypts it. These steps have been illustrated in figure 1. Users ask XKMS to get the needed key for decryption. XKMS would give the key if the user has been authenticated. Mechanism of authentication is dependent to the infrastructure of SOA. The simplest way is using the LDAP[5] which is a server containing the information of all valid users [10]. While user asks a key from XKMS, XKMS sends an authenticating request to LDAP server.

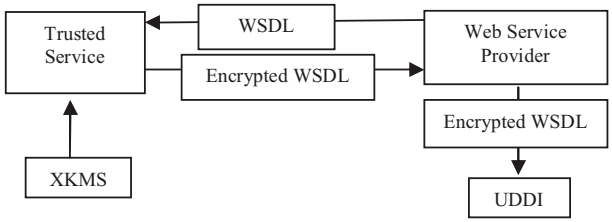


Fig. 1. The Steps of suggested approach

In continue, we describe the steps of encryption and decryption phase by details.

A. Encryption Phase

The steps of this phase are shown in figure 3. There are two main components in this stage. First, trusted web service that all of the other components trust this service and it has permission to access to WSDL of all web services. We named it as "Trust". Second, provider web service, which is a sample of a service who wants to encrypt one of its WSDL file. We called it "Provider".

Step 1: Trust generates a pair of public and private keys (label 1 in figure 2). Web services who want to encrypt their WSDL need the public key of Trust to send their request. At this step, asymmetric encryption is used because only the provider service is known as a primary service. Since this web service is not aware of the future clients or consumers, it is not possible to use symmetric encryption.

Step 2: As public key should be available to everyone, a structure such as PKI[6], which have the ability to manage keys, is needed to maintain the keys that are generated asymmetrically. The Trust's public key is stored in the PKI structure and the private key is only in his possession and is protected by special algorithms. In order to maintain the public key which is generated in the previous step, Trust sends a Soap request message to XKMS to store the public key. XKMS is an interface on PKI that hides its complexity from users.

Step 3: XKMS sends the received request from Trust to PKI. This request is a XML-based message which contains the public key of Trust. Then, PKI creates a record containing digital certificates and the basic profile for accessing its

[1] Universal Description, Discovery and integration
[2] Simple Object Access Protocol
[3] XML Key Management Specification
[4] Uniform Resource Locator
[5] Lightweight Directory Access Protocol
[6] Public Key Infrastructure

information. Afterwards, PKI extracts the public key from the received message and imports it into the certificate.

Step 4: PKI sends a message to XKMS, informing the success of storing the web service public key and the generation of Trust certificate.

Step 5: XKMS resends a SOAP response message to Trust in response to initial request. This message indicates that storing public key in PKI was successfully done. After this process, the public key will be accessible through PKI structure while the private key is available only for its owner, i.e. Trust. The private key of each web service is protected by special mechanisms.

Step 6: After storing the Trust's public key in PKI, encryption is available for the web services who want to secure their WSDL for public access. At this point, the Provider sends his own profile along with a request for public key to Trust. Then, PKI checks compares the web service`s profile with issued digital certificates to ensure the validity of Provider`s request and Trust public key. If these two actions are successfully passed, PKI provides Trust public key for XKMS, embedded in an XML message. Finally, XKMS forwards the received message containing requested public key to the web service.

Step 7: Using the Trust's public key, Provider encrypts his WSDL file and prepare it to send as a SOAP message. Before sending the message, Provider encrypts it using his own private key in order to sign it. Finally, the message containing encrypted WSDL file will be sent to Trust.

Step 8: After receiving the message from Provider, Trust will check the signature of the message by using Provider`s public key to ensure that the message containing encrypted WSDL is accurate and has not changed. It is also needed to inform web service consumers that the WSDL is encrypted. Therefore, in the proposed model, a new security tag named <wsdlx: WSDL_secured>will be added to extended WSDL to show the state of this file. True value of this tag means that the cryptography layer is applied on WSDL. Notice that this tag should not be encrypted in order to users can find its status easily. The outline of expanded file is shown in figure 3.

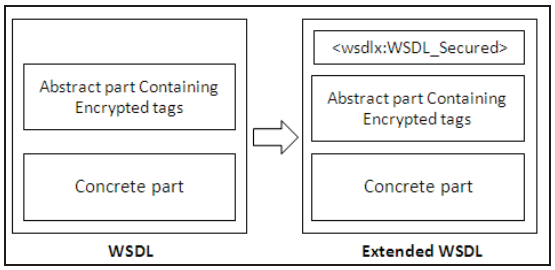


Fig. 2. Extended WSDL

Step 9: If Trust's private key is lost, none of the encrypted WSDL files would be accessible. Therefore, a mechanism is needed to store the information of web services whose WSDL have been secured. If necessary, they will be notified to encrypt and resend their WSDL file with the new Trust's public key. This information includes the business name and service name of provider's web service. This repository is named "Encrypted WS Repository" in figure 3.

Step 10 (final stage): After ensuring the safety of important information of WSDL, the provider's information should be recorded in UDDI, to be accessible.

The first five steps are related to Trust registration in PKI. This process is required only for the first time usage and there is no need to repeat it in later WSDL encryptions. In addition, the provider should register his information in PKI and receive the signed certificate to be identified as an authenticated unit for others. Because this process is similar to registering Trust in PKI, its description is avoided.

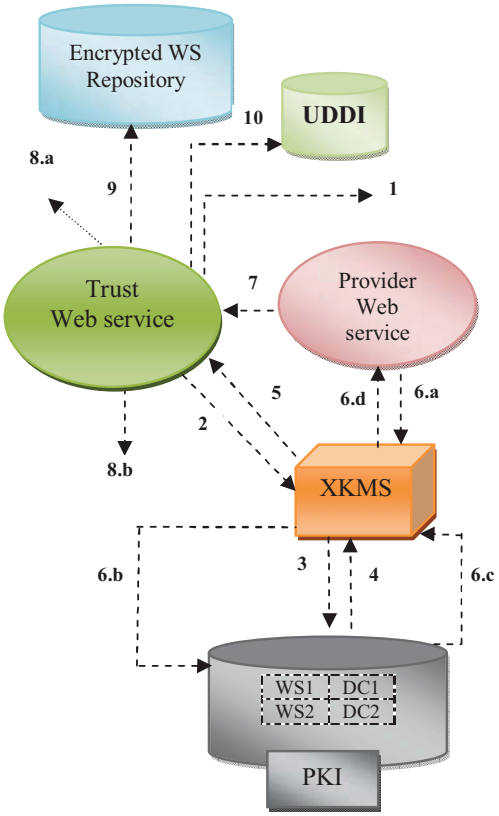


Fig. 3. Details of Encryption process

B. Decryption Phase

This phase includes two main components: consumer who wants to use a WSDL file that may be encrypted and Decryption web service that is service which acts as an interface for consumers to hide the complexity of decryption. We call this web service as "Decryption" web service.

Step 1: Consumer search his needed service in UDDI

Step 2: After accessing the requested WSDL file, consumer should find out the security status of WSDL by checking <wsdlx: WSDL_secured> tag. If the value of this tag is False, it means that the WSDL file is an ordinary file and it is used as usual. If the value of this tag is True, it shows that the WSDL

file is an encrypted file that needs to be decrypted in order to be used (figure 4, label2). Therefore, according to figure 4, the user will enter the third step.

Step 3: The consumer sends a signed message, containing his request to the Decryption service.

Step 4: After receiving the message, Decryption service opens the message signature by using the public key, which is saved in PKI, to ensure the authenticity of the message. In other words, Decryption service ensures the valid identity of the user. Decryption service accesses the required public key through XKMS interface (drawing the relation between decryption Web Service and XKMS is avoided because of its similarity to previous steps).

Step 5: After ensuring the consumer's authority and authenticity of the message, Decryption service goes through two stages, he prepares a SOAP Request message, containing the consumer's request along with the user's authentication and authorization. Then, in the second stage, he signs the message and sends it to Trust.

Step 6: After receiving the decryption message, Trust accesses the public key of Decryption service through XKMS. Then, Trust checks the message signature by this key to ensure the accuracy and validity of received data and to identify Decryption service as a valid user. If these stages carried out successfully, step 7 begins.

Step 7: Using the identity of user received in step 4, Trust checks his qualification for accessing the requested web service. This step, which is illustrated as label 7 in figure 4, takes place using LDAP server. LDAP server check the role of the requesting user to ensure that if the user is permitted to access to the demanded service or not[12]. If the requested access is determined as permissible, step 8 begins.

Step 8: At this point, two operations will be done. First, Trust decrypts the WSDL file by his private key and sends it to the Decryption service. Since the WSDL file supposed to remain secure until the last step to be delivered to the consumer without any possibility of misuse by malicious users and attackers, in this step it is necessary to re-encrypt the WSDL after decryption and before sending. To provide the best speed for encrypting WSDL, symmetric encryption is used in this step. Second, Trust exchanges the shared key between Decryption service and himself. This is done by using the STS key exchange protocol. After that, Trust encrypts the decrypted WSDL file of previous step by using the shared key and sends as a SOAP response message to the Decryption service. In addition, this message should be signed by his private key before sending.

Step 9: After receiving the SOAP message which contains encrypted WSDL, the Decryption service checks the accuracy of the message by Trust public key. Then, he decrypts the encrypted part using shared key. In order to send this file to the consumer, he creates a secure channel between consumer and himself using STS protocol just like before. Then, he encrypts the file by the shared key which is obtained from STS protocol and finally, he signs this file by his private key and sends it to the consumer.

Step 10: After receiving the signed message containing the encrypted WSDL file from Decryption service, consumer checks the message authenticity by its signature and the public key of Decryption service. After ensuring its accuracy, he opens the encrypted part of the message which was encrypted using the common key obtained from STS Protocol. Now the consumer can use this decrypted WSDL file.

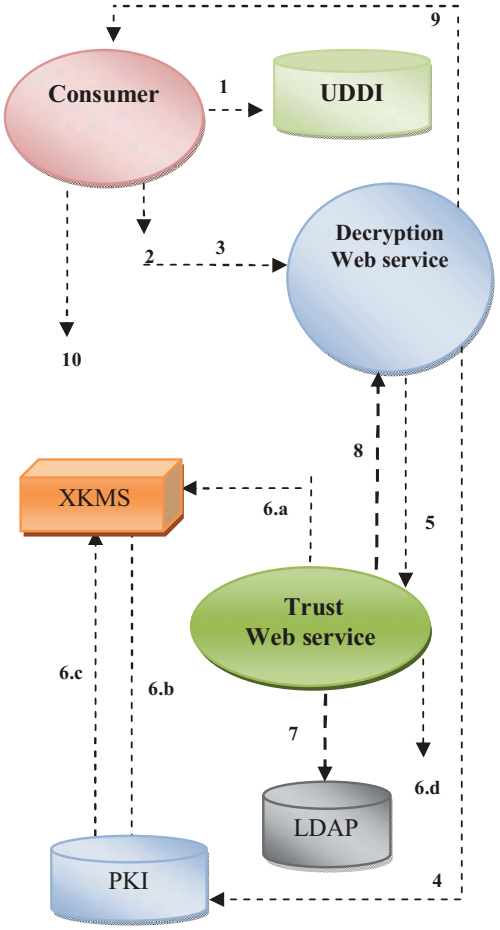


Fig. 4. Details of Decryption process

IV. Main Result

By applying the cryptography layer on WSDL file, the access time would be increased because users have to decrypt WSDL in order to can access to the needed information. So, this approach is not a common way for all web services and is suitable just for web services which heir WSDL contains valuable information for hackers according their policy, For example when the lists of methods of a web service should be hidden.

Table I. Evaluating some Metrics of Security

Scenarios	Action	Reaction	Reason	trick
Scenario 1	Encrypt WSDL by unauthenticated provider	unsuccessful	Provider is not authenticated	Detected by XKMS and cancels request
Scenario 2	Encrypt WSDL while public key of Trust service is not valid	unsuccessful	Trust service's public key is invalid	Detected by XKMS and cancels request
Scenario 3	Encrypt WSDL	successful	All conditions are available	Done all steps successfully
Scenario 4	Encrypt WSDL while a hacker attack to exchanged message	unsuccessful	An attack to transferred message from provider to Trust service	Trust service detects by checking provider's digital signature
Scenario 5	Encrypt WSDL while Trust service miss his private key	unsuccessful	The private key of Trust service is missing	Create a new public key
Scenario 6	Decryption request by unauthenticated user	unsuccessful	User is not authenticated	Detected by XKMS and cancels request
Scenario 7	Decryption request by unauthorized user	unsuccessful	User is not authorized	Detected by Trust service and cancels request
Scenario 8	Decrypt WSDL	successful	All conditions are available	Done all steps successfully

Evaluating of this approach is done in an example system including some service provider and service requester. Some scenarios in suggested model are explained in table I. This table has shown that how our approach acts in different situations.

In this paper, a method was presented to improve security of WSDL using WS-Security standard. Those who use WSDL must be able to successfully pass this security stage and access their needed information. In comparison, this proposed model has the following summarized benefits over the previous models:

- Providing WSDL privacy.
- Hiding a web service policy through securing WSDL information only for authorized consumers.
- Authenticating requesters by PKI and exchange protocols.
- Authorizing of requesters by LDAP server.
- The performance of this security solution is thoroughly analyzed in applicable environment based on service- oriented architecture.

V. Future Works

In this article, efficiency was not concernand while the speed was decreased due to cryptography including encryption and decryption processes. The speed of service-oriented environments is partly lower than other architectures because of the high amounts of message communications and this security layer intensifies it. Using other encryption algorithms will have a considerable effect on this model's performance or use EaaS[1] in clouds to improve the response time of the encryption algorithms[13].

[1] T. Erl and T. Boubez, "SOA Security," ed: Pearson Education, Limited, 2011.

[2] M. P. Cristescu, E. A. Stoica, and L. V. Ciovică, "Web Services Specific Security Standards," *Procedia Economics and Finance,* vol. 16, pp. 597-602, 2014.

[3] M. I. P. Salas and E. Martins, "Security Testing Methodology for Vulnerabilities Detection of XSS in Web Services and WS-Security," *Electronic Notes in Theoretical Computer Science,* vol. 302, pp. 133-154, 2014.

[4] A. Mirtalebi and M. R. KhayyamBashi, "Enhancing Security of Web Services against WSDL Threats," in *IEEE 2nd International Conference on Emergency Management and Management Sciences*, 2011, pp. 920-923.

[5] R. Kraus, B. Barber, M. Borkin, and N. J. Alpern, "CHAPTER 6 - Internet Information Services – Web Service Attacks," in *Seven Deadliest Microsoft Attacks*, R. Kraus, B. Barber, M. Borkin, and N. J. Alpern, Eds., ed Boston: Syngress, 2010, pp. 109-128.

[6] H. Yue and X. Tao, "Web Services Security Problem in Service-oriented Architecture," *Physics Procedia,* vol. 24, Part C, pp. 1635-1641, 2012.

[7] R. A. Oliveira, N. Laranjeiro, and M. Vieira, "Assessing the security of web service frameworks against Denial of Service attacks," *Journal of Systems and Software,* vol. 109, pp. 18-31, 2015.

[8] E. Moradian and A. Håkansson, "Possible attacks on XML Web Services," *IJCSNS International Journal of Computer Science and Network Security,* vol. 6, pp. 154-170, 2006.

[9] H. Krawczyk and M. Wielgus, "Security of Web services," in *International conference on dependability of computer system* 2006.

[10] N. Sidharth and J. Liu, "IAPF:A Framework for Enhancing Web Services Security " in *COMPSAC. 31st Annual International of Computer Software and Applications Conference*, Beijing 2007, pp. 23 - 30.

[11] M. B. Juric, A. Sasa, B. Brumen, and I. Rozman, "WSDL and UDDI extensions for version support in web services," *Journal of Systems and Software,* vol. 82, pp. 1326-1343, 2009.

[12] L. M. Surhone, M. T. Tennoe, and S. F. Henssonow. (2010). *LDAP*. Available: http://books.google.com/books?id=R2tJYgEACAAJ

[13] J. Salim, M. Ismail, I. Suwarno, H. Rahmani, E. Sundararajan, Z. M. Ali*, et al.*, "Encryption as a Service (EaaS) as a Solution for Cryptography in Cloud," *Procedia Technology,* vol. 11, pp. 1202-1210, 2013.

[1] Encryption as a Service

Analyzing Vulnerability Databases

Rashad Aliyev
Khazar University
Baku, Azerbaijan
rashad@aliev.info

Lourdes Peñalver
Universitat Politècnica de València
Valencia, Spain
lourdes@disca.upv.es

***Abstract*—There are more and more vulnerabilities in database of computer system about weakness of software technology. Not all of them are major or most important. Our goal in this paper is to analyze a selected four major vulnerability databases that are the most important and their collect more than 230000 vulnerabilities. In this analysis we compare databases by the number of the vulnerabilities and the CVEs implemented, to find the best vulnerability databases. From those selected vulnerability databases we have got information about the vulnerabilities and applying those details to our work in penetration tests for doing reports.**

***Keywords*—Vulnerability databases, CVE, NVD, BID SYM.**

I. Introduction

Starting 1989 [1] and after of the Morris worm virus, DARPA establish the CERT -the Computer Emergency Response Team - at the Software Engineering Institute at Carnegie Mellon University for finding, collect and publish all information about vulnerabilities.

After CERT displayed information about vulnerabilities publicly and started publishing it, all of these vulnerabilities appeared in many new vulnerability databases with published it in different information format about these weaknesses. For which vulnerability databases is early published information, or which of this have more details (Solution, software and so.), or which database has more countable vulnerabilities for all of that is difficult to choose. One of our first objective is to analyses that.

Firstly we explained the problem in section two, that in our Penteston project for show detailed information about vulnerability for choose best vulnerability database when we analyze databases found more databases. Also when we read paper about other researchers we found more of that used CVE, NVD and OSVDB databases and have interesting results in different times, but for our research these three databases is not enough and need choose more databases.

In section third our idea is to explain about why we have planned to use of the selected vulnerability database and why we use it in our project and what will be the working structure of project.

In details section four we used ten vulnerability databases that Common Vulnerabilities and Exposures, U.S. National Vulnerability Database, Open Sourced Vulnerability Database, Secunia database, Security Focus. BugTraq, Symantec's Threat Database, Microsoft Security bulletin, Mozilla Foundation Security Advisories, Exploit Database, National Computer Network Intrusion Prevention Center to select for our research.

In next step We selected the four major vulnerabilities databases from the list: Common Vulnerabilities and Exposures (CVE) [2], National Vulnerability Database (NVD) [3], Open Sourced Vulnerability Database(OSVDB) [4], Security Focus. BugTraq (BID) [5], Symantec's Threat Database (SYM) [6], Secunia database [7], Microsoft Security bulletin [8], Mozilla Foundation Security Advisories [9], Exploit Database [10], National Computer Network Intrusion Prevention Center [11].

To choose the four major databases we get those details; CVE have data from 1999. Until 2015 December CVE and has 86494 vulnerabilities. National Vulnerability database (NVD) work from 2002 to 2015 has information about 74163 CVEs. Securityfocus BugTraq has 71366 vulnerabilities and all data is in HTML format. Symantec's Threat Database (SYM) has information about 2639 vulnerabilities from 122986 Technologies (Software)

In our research work we have collected information about vulnerabilities from selected four vulnerability databases to our structured database. When we analyze a vulnerability database we get interesting information as all vulnerability databases have relations between CVE vulnerability databases by CVE-ID. And in next steps we started work for comparing those vulnerability databases by CVE-ID and by counting all the vulnerabilities in those databases.

And finally in section five we presented as conclusion statistics for four databases, comparing these databases by count of vulnerabilities, relations near these databases and by update times. For future, it is planned to get detailed information from those databases and we will apply these details to our Penteston Project (PP).

II. The problem

There are lots of types of vulnerabilities in several databases, but they have not statistical information about the best database. For example, MS database have information about vulnerabilities in Microsoft products, BugTraq requesting solutions about vulnerabilities how to solve it, NVD have score level for vulnerability as explained in [12], also as explained in [13], [14] OSVDB is large database containing more reports about vulnerabilities.

In the paper [15] appears 70078 records, obtained information about CVE from 1999 until May 5, 2014 , when we analyzed databases until December of 2015 there are 86494 records. As a result in the last year, they prepared 16416

records, and in [16] show a published daily rate was 20, but last year rating is showing it's increased from 20 to 30.

In paper [17], Mikhail Buinevich, and others write that one of actual problems is find software vulnerabilities, because vulnerability databases have information in different formats, in this paper were analyzed NVD database information about software displayed in common platform enumeration (CPE) [18] format like "cpe:/o:freebsd:freebsd:1.2", but in another Symantec database it's displayed for software name "FreeBSD 4.0.0" and It's a really problem why can we compare software vendors and vulnerability databases, and also which vulnerability is actual for a given software or what not.

In [19], they show information about NVD where from April of 2009 collect 37031 vulnerabilities listed in the National Vulnerability Database, but now in 2015 NVD have 74163 vulnerabilities near of the record number in 2014 with 7365 vulnerabilities, as shown in figure 1

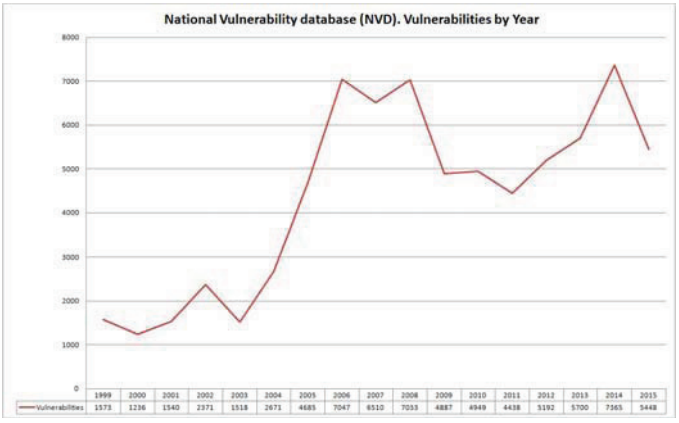


Figure 1: National Vulnerability database (NVD). Vulnerabilities by Year

In paper [13] they work with information about OSVDB (Open Sourced Vulnerability Database) where they collect over 57000 vulnerabilities, but currently OSVDB have information about 120963 vulnerabilities, also in paper [20], they given information about 21% of OSVDB that have not CVEs. From achieved results and information about vulnerability databases for choose best vulnerability database we need research.

III. Our Idea

In penetration testing for getting information about weakness and generate detailed reports, It is necessary analyze all vulnerabilities of the system. These Vulnerabilities should be collecting in different databases to bring information to the public. Moreover, these vulnerability databases have information referred about weakness, detected date, level of security, results and how to solve the it.

In our Penteston Project (PP) as the beginning we have decided to analyze and find what will be the optimized database always following our criteria and after that integrate it in our system. Done that, we will need to collect vulnerabilities information as white hats or with automated scanners that is the final objective of PP.

After that, we can obtain generated reports in different times and formats, then to analyze and compare they will cost much time. To improve that we designed the PP platform, that can quickly create a new report summarizing from the obtained ones: collecting all vulnerabilities about system in one source, getting detailed reports for security experts, site owners or top managers of company getting security information about tested system by security audit or penetration testers.

To use PP, the members can register and sending their automated scan report results manually or gather automated scan by the system. After that, it is possible to obtain a report for managers or a detailed one for developers. Moreover, could be obtained a compare different time created reports.

For get results on PP we need follow four steps as in [21]:

1. Access: In PP members for create targets and gain security tests, and in future for access to their systems it is necessary register and open accounts. Currently when we researched for first step we see more members have Google accounts and depending on this we used Google OAuth 2.0 protocol for registering and authorization to PP and registration by email. In next steps have plan is use another methodologies also such as Facebook, Twitter, LinkedIn for authentication and authorization.
2. Planning: In this section the member can add information about their own project as information about the project Scope of Work, NDA, team information and how to conduct penetration testing (Ex.: Whitebox, Greybox or Blackbox), Types (Ex.: Internal, external, application-layer or network-layer), etc.
3. Footprinting and Exploits steps: In order to carry out measures on this stage, the team has to gather information on target on which they will do penetration tests for the system and send collected alerts to exploits and collecting information about weaknesses.
4. Report: They have to download a report and sending it to management. Also, if in the system they are more than two reports they will have the option to obtain a compared and summarizing report changes.

In generated reports for getting detailed information about vulnerabilities, solutions or getting security level of weakness we started to research vulnerability databases for implementing it to PP.

IV. Details about our work

The main goal for the beginning is collecting information about vulnerability databases. Different databases publish different information related to vulnerabilities. Each database has their personal prefix [22]. For example: Security Focus of Bugtraq prefix is BID, Microsoft Security Bulletin have prefix MS, Common Vulnerabilities and Exposures is CVE.

These sources for collect information about vulnerabilities and to publish it have different rules and guides, and for this they have different formatted information. For relation with vulnerability information in two or more sources, they register

vulnerability in CVE database. For how to get CVE number let's see the sample: BID is a prefix supported by Security Focus Bugtraq (Example: BID 8829) and have information about vulnerabilities by their criteria. From those sources or from security experts sending information about vulnerabilities to CVE database, if CVE is confirming these weakness then it giving CVE prefix to this vulnerability, and CVE database is increasing with these information. But there are situations where some weakness have not that information in the CVE database but that is in the database source.

In some cases one vulnerability data is registered in different databases by another ID, in this way for find information about all vulnerabilities from other sources using CVE ID as primary key for communication different vulnerability databases. For example: CVE-2003-1523 - In another sources registered by different ID's SECUNIA:10001 (SA10001), BID:8829 (Bugtraq ID:8829), XF:dbmail-multiple-sql-injection (13416).

We have decided to use CVE ID for comparison of databases. In order to compare them we choose some databases and started for first selection step as:

- Common Vulnerabilities and Exposures (CVE) [2]. CVE is started from 1999 and from the years publishing information about vulnerabilities public. Every vulnerability have Unique ID and registering buy CVE standards [23].

- U.S. National Vulnerability Database (NVD) [3]. Supporting by US government and working together with CVE vulnerability database.

- Open Sourced Vulnerability Database (OSVDB) [4]. One of big vulnerability databases have more than 100,000 vulnerabilities. After contacted to OSVDB we get informed that it is owned by Risk Based Security (RBS) and changed its name to VulnDB and RBS does not need or want research in an academic thesis and for this reasons we do not work on this database.

- Secunia database [7]. Is a commercial database, then you needed to pay to obtain information.

- Security Focus. BugTraq [5]. This vulnerability database operates since 1999 and established by SecurityFocus company. There is detailed information about every vulnerability, in which software is detected, if has exploit information about it and detailed information to solve the problem.

- Symantec's Threat Database (SYM) [6]. There is detailed information about vulnerabilities, recommendations for solve problem, also supported software list.

- Microsoft Security bulletin [8]. Information about vulnerabilities in Microsoft products. Also if security experts found some weakness on these products, in Microsoft have Bug Bounty program for reward from 500 USD until 100,000+ USD [24].

In this bulletin is only used Microsoft products, have few information and all vulnerabilities have CVE identification, for this reason we do not use this vulnerability database in our research.

- MFSA: Mozilla Foundation Security Advisories [9]. Placed information about vulnerabilities in Mozilla products. Also if members found on system some weakness then using Bug Bounty program can rewards from 500 USD till 10,000+USD [25].

In our research we do not use this database, because these vulnerabilities are only for Mozilla products and not cover all services.

- Exploit Database [10]. This database has information about vulnerabilities if this vulnerability have exploit. So that information about vulnerability in this database, you need only copy paste code and apply this to lots of systems where have this weakness and make total automated attack to systems. In our work we do not use this database also, because here have information about vulnerabilities where only have exploits.

- National Computer Network Intrusion Prevention Center (NIPC) [11]. Vulnerability database were represented in Chinese language. In this vulnerability database information about vulnerabilities in Chinese language and for current analyze we cannot use this database in our work because language barriers.

After some work and searches we started to analyze databases and we have decided to choose some of them. Our selection criteria about vulnerability databases is database not commercial, covering all software, have detailed and enough information about vulnerabilities. After that we have select four vulnerability databases. In a first stage we get information vulnerabilities from those databases and structured that in a one format. As a middle point of those vulnerabilities we found CVE ID, that represent year and vulnerabilities counter.

First: Common Vulnerabilities and Exposures (CVE) format at MITRE [26]. CVE Identifiers (also referred to by the community as "CVE names", "CVE numbers", "CVE entries", "CVE-IDs" and "CVEs") are unique, common identifiers for publicly known cyber security vulnerabilities [27].

CVE Initiative first began in 1999. Until 2014 CVE-YEAR-NNNN. After 1 January 2014, CVE-YEAR-NNNN...N (min 4N). CVE-ID numbers using the new syntax were issued beginning on January 13, 2015 [28], [29].

For downloading and working with CVE database, MITRE presents data in many different formats [30] as XML, HTML, CSV and text, we have chosen HTML. The HTML formatted file CVE gives CVE ID and information about vulnerabilities from 1999 until our last download date, 18 December 2015. Our Downloaded 86494 vulnerabilities has 787072 references, from here we found all CVE information updating daily monitored by CERIAS/Purdue University using a special developed free tool [31].

Second: National Vulnerability database (NVD) works together with CVE database, but NVD has a Common Vulnerability Scoring System (CVSS) [12] that calculates values from 0.0 to 10.0 depending of the vulnerability security level. NVD use XML format, as a result presented a XML file ordered by the vulnerability date. nvdcve-2.0-[YEAR].xml 2002<=[YEAR]<=Current year. Where “Current year” is a variable that is incremented every year. Sample “current year” is 2016.

In NVD database from period: 2002 to 2015 December 17 there are recollected information about 74163 vulnerabilities,

these vulnerabilities has 363606 references and vulnerable in 1915880 Software. In both CVE and NVD the last update date is similar to date when we get information (17 December 2015).

Third: Security Focus. BugTraq (BID) has identification numbering for vulnerabilities from one up incremental sequence. In our data fetching time (25 November 2015) system had 77651 ID's. In Security Focus for represent information about vulnerabilities has not any standard formats, and for all that we extracted information by scraping data from the site. From the 77651 ID's obtained, 6285 have not any information about vulnerabilities, and we started working on the rest:71366 BID. From these vulnerabilities, 42337 BID have CVE ID and 29029 BID haven't CVE ID's. Despite of time of getting information is 25 November 2015, last update of information about vulnerabilities shows 19 November 2015. From here in order BID is updating temporary.

Fourth: Symantec's Threat Database (SYM). For presenting information about vulnerabilities SYM is using numbering format from Security Focus BID. But information about vulnerabilities in SYM differ from data about vulnerabilities on BID; Moreover information about vulnerability in BID is discussion, exploit, solution and references in different pages, but SYM have all information in one page, and have not information about exploit. When we analyzed SYM database we get 2639 ID and this ID has 4946 reference, 122986 affected technologies (Software), 1400 CVE ID's. Also on fetching date (25 November 2015) we got information about last update time of vulnerabilities is 10 November 2015.

From here we have information about four databases. There are 86494 vulnerabilities in CVE, 74163 vulnerabilities in NVD, 71366 vulnerabilities in BID and 2639 vulnerabilities in SYM database, also there are references and software information were these vulnerabilities exist. In Figure 2 we have designed this database structure and their relations.

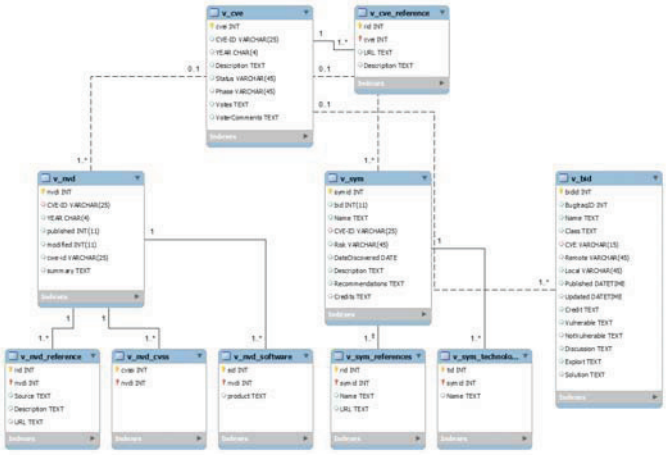


Figure 2: Database design of integrated vulnerability databases

In that figure it is presented design structures of four vulnerability databases that we used in our research. Here v_cve, v_nvd, v_sym, v_bid are main structures for vulnerability databases, and v_cve_reference, v_nvd_reference, v_nvd_software, v_nvd_cvss, v_sym_references, v_sym_technolgies are additional tables of the databases. From these figures we get interesting information as all vulnerability databases there are relations between cve_id table and CVE-ID field. In our research work we have collected information about vulnerabilities from vulnerability databases to our structured database. And in next steps we started work for compare those vulnerability databases by CVE-ID and by count of vulnerabilities in those databases.

V. Conclusions and Further Work

In our work we analyzed 234662 vulnerabilities in four databases. In figure 3 showed chart of databases.

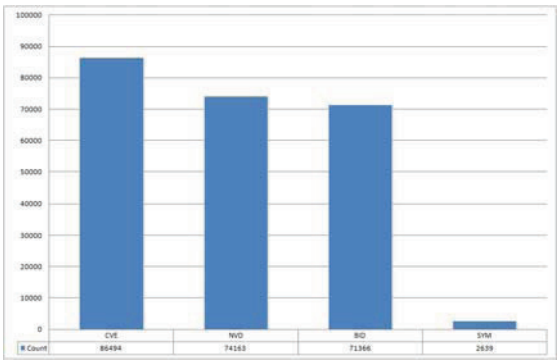


Figure 3: Vulnerability databases

When we analyze this vulnerabilities by CVE Identifiers we found information showed on Figure 4.

-	CVE	NVD	BID	SYM
Vulnerabilities	86494	74163	71366	2639
Have CVE	86494	74163	42337	1400
	100%	100%	59.32%	53.05%

Figure 4: CVE statistics by database vulnerabilities

From these figures we get interesting information. NVD is integrated database to CVE and for all these NVD vulnerabilities have CVE Identifiers. All NVD vulnerabilities have their CVE identification numbers and as we write on the details section NVD has their additions like CVSS, from here we can use these database as extended format for CVE database. But BID and SYM have their own numbers and not all vulnerabilities have CVE Identifiers. From this we get result where 30%-35% vulnerabilities have not CVE Identifiers.

Finally, when we analyzed databases we found some rules for compare. One of this by number of vulnerabilities, is best vulnerability databases are CVE, BID and NVD.

By updates best vulnerability databases are CVE, NVD and BID. By percent of CVE if we make rating CVE, NVD and BID is good for choose.

As an continuation of the research, we are going to implement the selected best vulnerability database to PP. In this report if alert has CVE-ID or BID then at first we find CVE-ID from NVD database get from there severity information, from BID we get solutions and exploits and other

useful information and add all those information for generate detailed report.

ACKNOWLEDGMENT

This work was partially supported by the ”Programa Estatal de Investigación, Desarrollo e Innovación Orientada aRetos de la Sociedad, Proyecto I+D+I TEC2014-52690-R”.

REFERENCES

[1] E. H. Spafford, “Crisis and aftermath,” Commun. ACM, vol. 32, no. 6, pp. 678–687, Jun. 1989. [Online]. Available: http://doi.acm.org/10.1145/63526.63527
[2] “Common vulnerabilities and exposures (cve),” http://cve.mitre.org, accessed: 2016-04-28.
[3] “National vulnerability database,” https://nvd.nist.gov, accessed: 2016-04-28.
[4] “Open sourced vulnerability database,” http://osvdb.org, https://osvdb.org/, accessed: 2015-10-29.
[5] SecurityFocus, “Bugtraq,” http://www.securityfocus.com/vulnerabilities, accessed: 2016-04-28.
[6] “Symantec’s threat database,” https://www.symantec.com/security_response/landing/vulnerabilities.jsp, accessed: 2015-10-29.
[7] “Secunia database,” https://secunia.com/community/advisories/historic/, accessed: 2015-10-29.
[8] “Microsoft security bulletin,” https://technet.microsoft.com/security/bulletin/, accessed: 2015-10-29.
[9] “Mozilla foundation security advisories,” https://www.mozilla.org/en-US/security/advisories/, accessed: 2015-10-29.
[10] “Exploits database by offensive security,” https://www.exploit-db.com/, accessed: 2015-10-29.
[11] “National computer network intrusion prevention center,” http://www.nipc.org.cn/, accessed: 2015-10-29.
[12] P. Mell and K. Scarfone, “Common vulnerability scoring system,” Security Privacy, IEEE, vol. 4, no. 6, pp. 85–89, Nov 2006.
[13] M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker, “Beyond heuristics: Learning to classify vulnerabilities and predict exploits,” in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’10. New York, NY, USA: ACM, 2010, pp. 105–114. [Online]. Available: http://doi.acm.org/10.1145/1835804.1835821
[14] M. Gawron, A. Amirkhanyan, F. Cheng, and C. Meinel, “Automatic vulnerability detection for weakness visualization and advisory creation,” in Proceedings of the 8th International Conference on Security of Information and Networks, ser. SIN ’15. New York, NY, USA: ACM, 2015, pp. 229–236. [Online]. Available: http://doi.acm.org/10.1145/2799979.2799986
[15] A. Fedorchenko, I. Kotenko, and A. Chechulin, “Design of integrated vulnerabilities database for computer networks security analysis,” in Parallel, Distributed and Network-Based Processing (PDP), 2015 23rd Euromicro International Conference on, March 2015, pp. 559–566.
[16] V. Okun, W. F. Guthrie, R. Gaucher, and P. E. Black, “Effect of static analysis tools on software security: Preliminary investigation,” in Proceedings of the 2007 ACM Workshop on Quality of Protection, ser. QoP ’07. New York, NY, USA: ACM, 2007, pp. 1–5. [Online]. Available: http://doi.acm.org/10.1145/1314257.1314260
[17] M. Buinevich, K. Izrailov, and A. Vladyko, “The life cycle of vulnerabilities in the representations of software for telecommunication devices,” in 2016 18th International Conference on Advanced Communication Technology (ICACT), Jan 2016, pp. 430–435.
[18] “CPE - common platform enumeration,” http://cpe.mitre.org, accessed: 2016-05-23.
[19] J. A. Wang, L. Zhou, M. Guo, H. Wang, and J. Camargo, “Measuring similarity for security vulnerabilities,” in System Sciences (HICSS), 2010 43rd Hawaii International Conference on, Jan 2010, pp. 1–10.
[20] C. Zheng, Y. Zhang, Y. Sun, and Q. Liu, “Ivda: International vulnerability database alliance,” in Cybersecurity Summit (WCS), 2011 Second Worldwide, June 2011, pp. 1–6.
[21] L. P. Rashad Aliyev, “Penetration testing reporting y methodology,” in XXVI Jornadas de Paralelismo (JP2015), Cordoba, Spain, 2015, pp. 540–543.
[22] “Cve reference key/maps,” http://cve.mitre.org/data/refs/index.html.
[23] “About cve,” https://cve.mitre.org/about/index.html.
[24] “Microsoft bounty programs,” https://technet.microsoft.com/en-us/security/dn425036.
[25] “Mozilla client security bug bounty program,” https://www.mozilla.org/en-US/security/client-bug-bounty/.
[26] J. A. Wang and M. Guo, “Ovm: An ontology for vulnerability management,” in Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, ser. CSIIRW ’09. New York, NY, USA: ACM, 2009, pp. 34:1–34:4. [Online]. Available: http://doi.acm.org/10.1145/1558607.1558646
[27] “About cve identifiers,” https://cve.mitre.org/cve/identifiers/index.html, accessed: 2016-04-29.
[28] “Cve-id syntax change,” https://cve.mitre.org/cve/identifiers/syntaxchange.html, accessed: 2016-04-29.
[29] “Cve-id syntax,” https://cve.mitre.org/cve/identifiers/cve-ids.html, accessed: 2016-04-29.
[30] “Cve downloads,” https://cve.mitre.org/data/downloads/index.html, accessed: 2016-04-29.
[31] “Index of cve changes,” https://cassandra.cerias.purdue.edu/CVE_changes, accessed: 2016-04-29.

A light weight dynamic attribute based access control module integrated with business rules

Vali Tawosi
Electrical & Computer Engineering Department
Tarbiat Modares University
Tehran, Iran
vali.tawosi@modares.ac.ir

***Abstract*—User authorization in software systems is and has been a serious security concern for a long time. Attribute based Access Control (ABAC), as a new model of user authorization, makes it possible to restrict user access based on rules against different attributes. In the context of service access control in enterprise systems, it seems necessary to separate business rules from service logic and user authorization mechanism. This paper is an experimental report on the implementation of an ABAC module in which business rules are used to restrict user access to the services. The ever changing nature of the business rules in an enterprise system made a necessity to the proposal of such a light weight dynamic attribute based access control module, in which end user is able to change access policies and business rules in run time. Challenges of building this module are revealed and plausible solutions which have been put in place are reported.**

***Index Terms*—Authorization, Attribute Based Access Control, Business Rule, Banking System.**

I. Introduction

User authorization in software systems is and has been a serious security concern for a long time [1]. Through the time some approaches are introduced and practiced, each one with its advantages and draw backs. Any resource that could be accessed by a user should be protected from unauthorized access to protect confidentiality, user privacy and system integrity. Among all the resources accessed by user, system services are of special concern. Restricting user authority on the system service-calls is a major point of concern in enterprise systems.

On the other hand business rules defines some restrictions on the way a service is used. In any business there are always a set of restrictions that determines the business policies. Combining these policies with user authorities gives the system administration a new tool to manage how system could be used by users. Strictly speaking a user may have authority to call a system service, but he/she may use it in a rule restricted environment. It means user can use the service, but with the inputs that are restricted by business rules.

In a business context that rules and regulations are frequently changing and there is an urgency in the affection of the changes, we need a system that lets users to change the rules on the fly (while the application is running) and see the results in real time. This paper explains a real experiment of developing an ABAC module with specific requirements and discusses the challenges and solutions.

The paper organized in six sections. Section 2 provides with a background of the access control models. Section 3 explains the requirements that realized in the presented method. Section 4 describes the design, implementation and working policy of the method and section 5 points out the challenges which the implementation was faced with and the solutions worked out that challenges. And finally section 6 is a conclusion to this paper.

II. Background

The concept of ABAC has been existed long before it is attracted researchers' attentions in authorization systems. Necessity for an authorization system that considers not only the subject that demands access, but also the demanded object's attributes, and the environmental conditions that the access is initiated in, made a way for ABAC come into existence. These considerations that are addressed before but separately in different models (DAC, MAC, IBAC and RBAC) now became united in ABAC. A formal unification of these access control models is discussed in [2].

ABAC is a logical access control model in which access to an object (that can be a service) is controlled by evaluating rules against the attributes of the subject, object, action and environment relevant to the request [3]. For example when a user requests 'read' access to a file, access control policy using ABAC can impose restrictions over a variety of attributes like:

1) User (requester) information like: username, user role, user age, user machine IP address, last request from this user and etc.

2) File (something that is requested) information like: file name, file size, file format, file owner, file directory and etc.

3) Action (what the requester wants to do with the object) type like: read, write or read-write.

4) Request environment like: the time that the request has been issued, number of requests to this object in last [time period], number of requests the user had in last [time period] and etc.

Extensible Access Control Markup Language (XACML) is an example of a framework consistent with the ABAC [3]. This framework provides syntax for a policy language and semantics for processing this language [4]. Figure 1 shows the architecture of the XACML. Every request is intercepted by

Policy Enforcement Point (PEP) that decides whether the request is authorized or not. Decision making happens in Policy Decision Point (PDP) which gathers and evaluates rules stated in the access policy. Policy Administration Point (PAP) is where policy admin (developer or in some cases end user) can manage policies. Policy Information Point (PIP) is a point in which additional attributes that are not embedded in request are fetched, calculated or prepared to use in PDP [5]. Based on the decision that PDP makes, PEP allows user to access or deny it.

III. Requirements

In the Private Banking System (PBS) there could be a variety of user specific rules; meanwhile there should be a way to put exceptions in place in the favor of some customers. For example take the case of creating a new contract. Following rules could be placed on the "create new contract" service:

- Customer Contributing Cash should be *more than* 15,000$ **AND** Interest Rate should be *less than* 10% **AND** Contract Duration should be *less than or equal to* 12 months **OR** Customer is a VIP.
- Request comes from a terminal in the same branch that the requester (user) works in.

In the aforementioned rule the values like 15,000$, 10% and 12 may change frequently. The rule set itself, may be subjected to changes. For example adding a new rule likes "Interest Payment Period should not be *less than* 3 months" to the set, or removing one, or temporary disable/enable one.

What we are interested in here is that an ordinary user be able to perform this changes in runtime using a user friendly interface and the changes be applied on the system immediately. Beside this, PBS needs to have a flexible system. This means that any rule could be violated with an approval from a second user with super user privileges. Since a super user is not present in every branch of bank, we had to provide a method in which ordinary users requests a rule violation on a service, and once a super user approved the request, service with violating inputs should be executed and results be send to the requester.

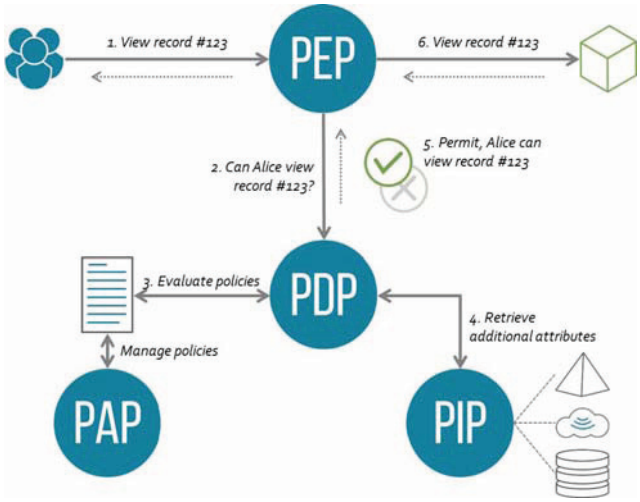


Fig. 1. XACML architecture and flow [6].

Therefor requirements can be expressed as follows:

1) Particular services called by user should be restricted by business rules.

2) User with specific privileges (call them super user) should be able to change business rules, using a user friendly interface, any time, while the application is live and running.

3) Super users should be able to perform ruled actions even when rules are violated. But in this case user should be informed of the rules he/she is violating.

4) When a user is banned from performing an action due to a rule violation, he/she should be able to send a request to super user to approve the execution of the action.

With having these requirements in mind, the situation hints us to use an ABAC to restrict ordinary users from calling specific services with unauthorized inputs. Since we need to give the super user a graphical user interface to alter ABAC restrictions runtime, we have to design and implement it with specific features. The basic design is borrowed from XACML and common ABACs, and required enhancements and specializations are done. Requirement number 4 alludes to the flexibility that PBS needs to have. A mini workflow like system helps us to send rule violation request from user to super user, and proceed to execute service if the super user confirms that. This is done asynchronously. Integrating this with aforementioned ABAC, brings up some new challenges that we had faced.

IV. Method

A light weight ABAC system is implemented. Its design that is depicted in Fig. 2 resembles that of XACML's. In order to provide end user with the ability of changing policies in runtime, it is decided to store policies in the database. There is no doubt that any other storage could be used instead of database.

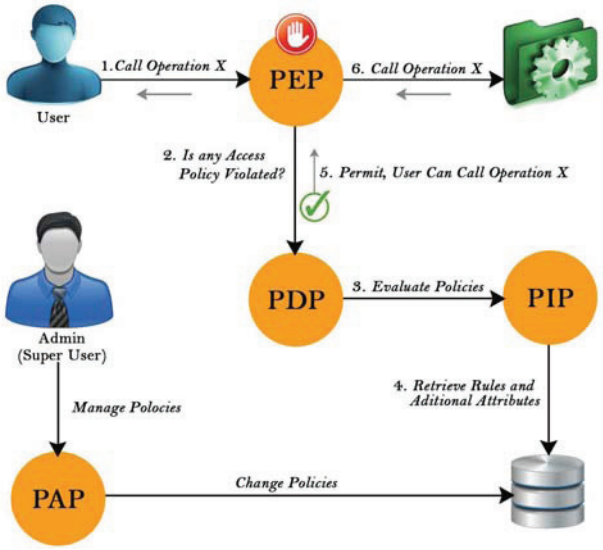


Fig. 2. ABAC architecture for Private Banking System.

A. Policy and Rule Representation

First we need to come up with a representation for rules and policies to store them in database and evaluate them later. A rule is a predicate that its evaluation could produce a True or False value. A sample predicate can be "Interest Rate should be less than 10%". Here 'Interest Rate' and '10%' are operands, and 'less than' is operator. We call this rule predicate as 'Rule Element'. Interest Rate is an attribute and its value is unknown when the rule element is defined (i.e. this is a variable). When the 'create contract' service is called, the contract object will be given as input. Interest rate is one of the attributes of the contract object. Let's assume that contract's interest rate is set by user to 9%. PDP that wants to evaluate the rule will ask PIP to resolve the interest rate variable using the service input parameter (i.e. the contract). In this example, PIP will acquire 9% as the value of the interest rate. Now, PDP substitutes variable with the value and can evaluate the predicate that has become like this: "Is 9% less than 10%?". This predicate evaluates to True and thus the rule is passed. PDP returns True to the PEP, and PEP permits user to do the action at the moment.

A policy is a combination of rule predicates. In fact a policy is a logical formula in which logical predicates are combined together using "OR"s and "AND"s. A standard form of representing logical formulas is Disjunctive Normal Form (DNF) [7] which is a disjunction of conjunctive clauses. In this form predicates (or their negative form) are combined to each other using AND operators and produce the conjunctive clauses. Then conjunctive clauses are combined together using OR operators. It has been proven that any logical formula can be represented by a DNF [7]. For example the policy provided for "create contract" service from section 2 is shown in a DNF here:

$$DNF: (A \wedge B \wedge C \wedge E) \vee (D \wedge E)$$

Where *A* is a place holder for predicate *"Customer Contributing Cash should be more than 15,000$"*, *B* is a place holder for predicate "*Interest Rate should be less than 10%"*, *C* is a place holder for predicate *"Contract Duration should be less than or equal to 12 months"*, *D* is a place holder for predicate *"Customer is a VIP"* and finally *E* is a place holder for predicate *"Request comes from a terminal in the same branch that the requester (user) works in"*.

Building a DNF form for a policy may requires user to have a level of logic skill that could be considered as a difficulty for user to working with the module. But using DNF as a base for representing policies helped us to overcome with policy representation and evaluation complexities.

Rule Element is an entity that holds data to represent a single rule predicate. Every rule predicate has a variable operand, an operator (i.e. a logical operator like equals, greater than, less than, between, and etc.), two values (i.e. constant operands), and a Boolean attribute which shows that the predicate is to be evaluated as positive or negative. A negative predicate will negate its result after evaluation and it means that the operator is negated (i.e. equals evaluated as not equals and greater than evaluated as less than).

Evaluating the rules, rule elements are combined using logical AND to compose an AND clause. We call every AND clause a Rule. Thus every rule has a list of rule elements. Rules are combined using logical OR to compose a DNF (i.e. a policy). Any service under attribute based authorization control is called an Operation. Each operation can hold a policy of access (i.e. a list of Rules). This policy can be altered at runtime by adding, removing or changing any rule or rule element through a graphical user interface. Policy data model is shown in Fig. 3.

B. How PIP works?

PIP needs to be able to resolve the variable attributes of every rule element. To this respect, we categories the source of variable attributes into three categories. Type one are attributes that belongs to an entity object; type two are primitive values provided in operations' input arguments; and type three are environmental attributes. We assume that for every environmental attribute, there is a method implemented in a class that if it is invoked, computes and returns a value for the environmental variable. Now in this taxonomy, if we consider that every entity is a Simple Java Bean (having setter and getter methods for its attributes), we can be sure that always there is a method (getter method) that will return the attribute value of the entity object. As it is shown in Fig. 3, Rule Method class is responsible for holding the method information that retrieves an attribute value for the rule element. If the rule method is responsible to retrieve a type one attribute, it is a getter method and receives no input arguments. But if the rule method is responsible to retrieve a type three attribute, it may need some input arguments. In this case, inputs should be retrieved from somewhere. This somewhere can be input arguments of the operation, or other methods. Since input arguments for a rule method are defined as rule methods, the retrieval process could be called recursively till the attribute value is returned. Type two attributes are retrieved from operation input arguments and hence their information is held by another entity named as Parameter in Fig. 3.

Rule References is a class that holds rule methods. On the other side rule references are also used as input types to the operations. Information held in rule reference like rule method, parameter, and operation entities will help PIP to use reflection to call appropriate methods runtime to resolve the attribute values. This information are gathered from source code and stored in database every time the application is built and run (application start time).

C. How PEP works?

PEP is a simple interceptor that is implemented by aspect oriented programming (AOP). Ruled operations are those operations that their authorization is controlled by ABAC. To mark a ruled operation, an annotation is defined (i.e. @CheckRule). Any call to an operation that is marked by @CheckRule annotation, will be intercepted by PEP. In the absence of requirement number 4 from section 2, PEP will simply call PDP to fetch policies for the current operation, evaluate them, and decide to proceed or otherwise throw an access denied exception.

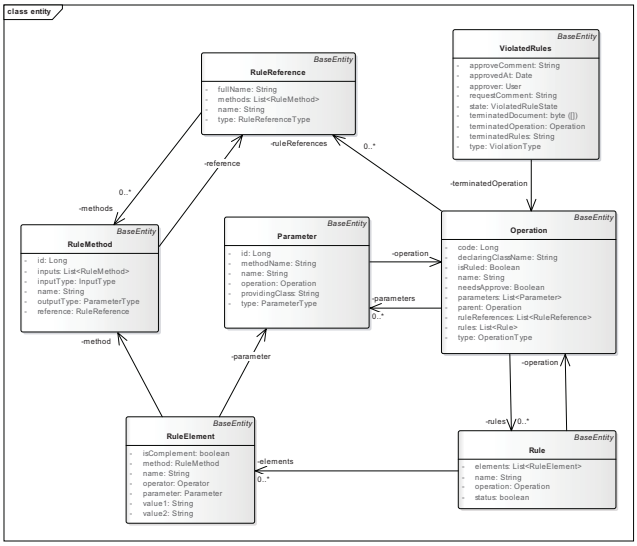


Fig. 3. ABAC data model for Private Banking System.

In order to realize requirement number 4, in the case that PDP reported that a rule is violated, and if user decided to issue an approval request to the super user, the operation input arguments should be serialized in a byte array output stream, and alongside other information like a request comment, operation name, and an explanation of violated rules (in natural language) is persisted in the database. Meanwhile a notification is send to the super user to inform him/her that "a violated rule is awaiting your decision to proceed". Super user checks the violated rule and in the case that approves the operation execution, adds an approvement comment and issues the execution command by clicking a button that calls violated rule execution method. This method deserializes the input arguments, calls the operation by reflection and provides it with the arguments. The operation executes and the requested user is informed by the result. This process faces a number of challenges that will be discussed in the next section.

V. CHALLENGES

***Challenge No 1*:** Firs challenge rises when some operations alter their input arguments before persistence. Technically rule evaluation should be done after this alterations happened.

***Solution*:** To solve the issue we proceeded the operation in PEP before calling PDP to evaluate the rules. After rule evaluation, if all the rules are passed PEP needs to do nothing. But in the case that there was a violation, any changes that has been done by proceeding the operation should be reversed. Hence a Rule Violation Exception is thrown to rollback the transaction and hence to rollback any changes that the operation proceeding had on the database.

This pre-proceed has another advantage: the operation is executed once and controls over input arguments are done once. This way if the input arguments are not valid to let the operation execution, user will get errors and correct them. If this control of input arguments is not checked in this point, it might produce errors when super user approves the rule violation request. Since super user should not be engaged with this kind of errors, this pre-proceed helps to avoid this situation.

***Challenge No 2*:** If rules are violated in PEP, how process could be hold until user decides to request an approval from super user or just ignore the execution? The information of violated rule (serialized arguments and other information) somehow should be passed to the request issuing process and PEP should not be concerned with approval request process at all.

***Solution*:** The information is wrapped up in an object and passed out within the Rule Violation Exception. Front-end beans are alert to catch this exception and handle it properly. To handle this exception an approval-request-issue window appears and user can add a comment and issue the request or simply ignore it.

***Challenge No 3*:** If the rules are violated and user requested a super user approval, the operation arguments should be serialized and stored to be fetched later and provided to the operation called by reflection. Since we has proceeded the operation once in the PEP, before evaluating rules, arguments became impure by possible alterations and cannot be used in second call.

***Solution*:** serialization of the input arguments is done before the operation proceeds in the PEP, and are held till the end and if a rule violation is reported by PDP, this clean serialized information are thrown out within the Rule Violation Exception, instead of the impure arguments.

***Challenge No 4*:** If an update operation happened to be ruled. Think of a rule over an attribute (let's call it X) that once is violated and approved by super user. Take for example a rule element like "X should be greater than 15". Then a rule violating value (let's consider the value as 11) is stored in attribute X in the database. When the update operation is called PEP will intercept it, and even if user didn't change X's value, PDP will report a rule violation. This should not happen since the super user approved value 11 for attribute X once before.

***Solution*:** Before rules are evaluated, PEP checks the X's old value, if it isn't changed by user, there is no need to check rules on this attribute. Rules will be checked for attributes of type one and two only when the attribute values are new or changed by the user in the current request.

The solutions provided to the challenges in this section are the best at hand solutions that are taken in place considering the PBS requirements and constraints.

VI. CONCLUSION

In this paper an experimental implementation of an ABAC module in a java application is presented. The architecture of the model is similar to a renowned ABAC framework called XACML, but this model is much more flexible and friendly in rule and policy changing by the end user. Because of the ever changing nature of the business rules in enterprise systems, rules and policies are separated from system business logic and stored in database. This needed an efficient representation for the rules and policies. The data model presented in this paper supported the rule and policy consistent storage as well as their evaluation. This implementation faced with some challenges

that are discussed in section 5 and solutions are presented. Integrating this module in a system is as simple as putting some predefined java annotations on the methods and classes. The end user can use a friendly graphical interface to define and change policies runtime.

ACKNOWLEDGMENT

This module is implemented as part of an enterprise private banking system at Rayan Eghtesad Novin Company, Tehran, Iran. The author thanks the Private Banking team members for their contribution in testing, reviewing and commenting on the project: Hossein Akbari, Kourosh Mozafari, Amir Nazarzadeh, Danial Jalal Nouri, Hojjat Gohargazi, Mohammad Behrouzifar, Nazanin Sharifi and Mahin Jahani.

REFERENCES

[1] G. N. H. Irani and V. Tawosi, "AAMA: A new Authentication and Authorization architecture for modular information systems, a robust object oriented approach," 2011, pp. 1–5.

[2] X. Jin, R. Krishnan, and R. Sandhu, "A Unified Attribute-Based Access Control Model Covering DAC, MAC and RBAC," in Data and Applications Security and Privacy XXVI, N. Cuppens-Boulahia, F. Cuppens, and J. Garcia-Alfaro, Eds. Springer Berlin Heidelberg, 2012, pp. 41–55.

[3] V. C. Hu, D. Ferraiolo, R. Kuhn, A. R. Friedman, A. J. Lang, M. M. Cogdell, A. Schnitzer, K. Sandlin, R. Miller, K. Scarfone, and others, "Guide to attribute based access control (ABAC) definition and considerations (draft)," NIST Special Publication, vol. 800, p. 162, 2013.

[4] M. Lorch, S. Proctor, R. Lepro, D. Kafura, and S. Shah, "First experiences using XACML for access control in distributed systems," in Proceedings of the 2003 ACM workshop on XML security, 2003, pp. 25–37.

[5] R. Yavatkar, D. Pendarakis, R. Guerin, and others, A framework for policy-based admission control. RFC 2753, January, 2000.

[6] "XACML - Wikipedia, the free encyclopedia." [Online]. Available: https://en.wikipedia.org/wiki/XACML. [Accessed: 15-Aug-2016].

[7] M. Hazewinkel, Ed., Encyclopaedia of mathematics: an updated and annotated translation of the Soviet "Mathematical encyclopaedia." Dordrecht ; Boston : Norwell, MA, U.S.A: Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, 1988.

Results of implementing WPA2-Enterprise in Educational Institution

Konstantin Yanson
SPO Faculty
ITMO University
Saint-Petersburg, Russia
k-yanson@corp.ifmo.ru

Abstract — **This article devoted to practical results of implementing wireless access to an educational network for students end employees in the SPO department of the ITMO University during 4 educational years. Experimental result shows how users deal with security whilst using their BYOD. Security and simplicity are two desirable but incompatible features that need to be balanced. Paper can help to find trade-off between security and convenience when we implement WPA2-Enterprise authentication. Results show what happens when we put security on users and at the same time cutback IT supplying.**

Index Terms **— BYOD, WPA2-Enreprise, MSCHAPv2, educational network.**

I. Introduction

This paper discusses the findings of the survey results pertaining access to the wireless network in the SPO department of the ITMO University during 2012-2016 years. ITMO University is an acknowledged leader in IT and quantum technologies in Russian Federation and in the world. It enters the top-100 QS BRICS rankings. The SPO department provides secondary special education in areas: “Programming in computers systems” “Information systems in economic”, “Computer networks” and “Automatic control systems”. The graduates are highly demanded in different areas of IT. The SPO department provides program of uninterrupted continuous professional education. It means that graduates continuing study on the 2nd or 3nd level of higher education (bachelor). The department works with students in different ages, providing uninterrupted educational program. [1][12]

The use of mobile devices in education and work environment is constantly increasing. It had brought new challenges for IT departments since BYOD (Bring You Own Devices) students come to educational institutions. Intel coined the BYOD term in 2009 and now it refers to the policy of permitting employees to bring personally owned mobile devices (laptops, tablets, and smartphones) to their workplace, and to use them to access privileged company information and applications. Increased processing power and memory of smartphones and tablet computers have made affordable ability to use BYOD in educational process on a par with regular computer.

BYOD is in use at 59% of organizations with another 13% planning to allow it [2]. BYOD is most common in the manufacturing and education sectors, and there are more small companies, with 50 or fewer employees, allowing it when compared to larger organizations [2].

There are certain ways to organize access to a corporate or an educational network but nowadays most popular and affordable way appears IEEE 802.11 standards called Wi-Fi [3]. On the one hand it is convenient and easiest way to use your own mobile device in a classroom (or workspace), but on the other hand it brings more constrains for IT personal because of inherit technological drawbacks [4]. Most critical drawback is security issue. It is difficult to find balance between security and simplicity for users. Most algorithms were compromised recently. WPA and WPA2-PSK is no longer safe, with the appearance of high-speed brute forcers [5]. Moreover, the use of WEP and TKIP for confidentiality, authentication, or access control was deprecated and that algorithms are unsuitable for the purposes of standard 802.11 [6].

Analyzing challenges and variety of frameworks, it is obvious that BYOD security needs advance research and development. Scholars recommend implementing a multi layered approach when devising BYOD security policies [7][8][9].

This article devoted to practical results of implementing wireless access to educational network for students with BYOD in the SPO department of the ITMO University. This research has attempted to find a tradeoff between security and simplicity. It is two desirable but incompatible features that need to compromise.

II. Related work

BYOD was investigated and described in variety of papers [7][10][11]. Combining different frameworks and mitigating constraints could lead to unpredictable results. In order to paper of Kathlin Dower and Maumita Bhattacharya [7], implementing BYOD brings new challenges for IT department:

- Deployment challenge:
  - Determining how to implement BYOD security measures into existing networks.
  - Determining who in the organization need BYOD.

- Technical challenge:
  - Access control.
  - Providing ongoing support.
  - Maintaining secure and stable connection.
  - Protecting company data stored on cloud facility.
- Policy and regulation challenge:
  - Local government regulations and laws.
  - Ethical and privacy issues.
- Human Aspect Challenge:
  - Training and education users.
  - User reactions, emotions and compliance of BYOD policy.

Wireless network standard and security techniques used by the higher educational institutes were analyzed in paper of Ranjana Shukla, Samad S. Kolahi, Robert freeth and Avikash Kumar [3]. It shows that 64% of the survived educational institutes use highly secure protocol. They have shown that current price of the new technologies is holding back some institutions from staring installing in the future.

There are a lot of researches pertaining security in Wi-Fi networks in itself. Obsolete algorithms, such as TKIP and WEP are not considered safe by the standard [12]. WPA/WPA2 algorithms are no longer safe because of possibility Men-in-the-middle attacks and Rogue AP [13].

Cisco and Microsoft has own frameworks for secure and seamless implementation wireless networks. [14]

Survey has founded that WPA2-Enterprise with RADIUS server standard, mostly popular among educational institutions [3]. In order to the research, about 64% educational institutions are using WPA2-Enterprise.

Inside the WPA2-Enterprise framework, a user must authenticate to the authentication server (RADIUS) using a secure and reliable EAP method. Although many EAP methods there are only two the most commonly used EAP methods: EAP-TLS and PEAP-MSCHAPv2 [14].

A. EAP-TLS

EAP-TLS, defined in RFC 5216 and well supported among wireless vendors, is one of the challenging methods in consequence of the certificate requirements. It provides authentication through the exchange and verification of X.509 certificates. Therefore, installing the correct certificates on BYOD and the authentication server is essential. Every end-user and computer, including the authentication server must possess at least two certificates: a client certificate, signed by the certificate authority (CA) and a copy of the CA root certificate.

EAP-TLS is the most secure method since the client's certificate cannot be forged. A user should make a personal client certificate as proof of identity. Afterwards, the client's certificate should be signed by the CA that issued it. It requires personal contact between IT-support and user. Ultimately, the authentication server must have a copy of the root certificate for the CA that signed the user's certificate.

B. PEAP-MSCHAPv2

The inner authentication protocol is Microsoft's Challenge Handshake Authentication Protocol (MSCHAP), meaning it allows authentication to databases that support the MS-CHAPv2 format, including Microsoft NT and Microsoft Active Directory. Behind EAP-TLS, PEAP-MSCHAPv2 is the second most widely supported EAP standard in the world. There are client and server implementations of it from various vendors, including support in all recent releases from Microsoft, Apple Computer and Cisco.

PEAP-MSCHAPv2 authentication relies on a password, not certificate as EAP-TLS and traffic is encrypted using TLS. A CA certificate must be used at each client to authenticate the server before the client submits authentication credentials.

For secure connection every user that participates in PEAP-MSCHAPv2 must possess the following credentials: Root CA certificate for the CA that signed the certificate of the authentication server and MSCHAPv2 username and password.

If the CA certificate is not validated, in general it is trivial to introduce a fake Wireless Access Point, which then allows gathering of MSCHAPv2 handshakes [15].

III. IMPLEMENTATION

The SPO department constantly strives to improve strategies of IT management. It makes the best to provide access to the network for each student and at the same time always pursuits to minimize cost of supplying customers. There are certain challenges for the department through the framework provided by Kathlin Dower and Maumita Bhattacharya [7]:

A. Deployment challenge

Determining how to implement BYOD security measures into existing networks

To provide high level of security for an educational institution, wireless network for BYOD should be separated from backbone network. Access controlling, queuing, shaping and policing are very desirable. Under no circumstances BYOD should interrupt the backbone network's work.

Determining who in the organization need BYOD

It is important to provide access to the educational network to each member involving in an educational process. Largest group is students. 73.8% of the SPO department's students use BYOD for education each day. 36.1% have one BYOD and 50.8% carry 2 personal devices for educational purpose. Tutors (teachers) and other employees group are smaller but need to be prioritized higher. Using CIS is mandatory so employee need stable connection on a regular basis.

B. Technical challenge

Access control

Access control could be organized on a various level. First and most important level is Level-2, where 802.11x authentication take place. The survived network is using PEAP-MSCHAPv2 authentication. MAC control is used as

addition measure. Only a valid MAC addresses can enter the network.

Providing ongoing support

A current price of the new technologies is holding back some institutions from staring installing in the future [3]. Having specialists on a regular basis could be costly. The SPO department always strives to decrease expenses on IT-supplying by implementing high level of automatization.

Maintaining secure and stable connection

There are a lot of devices that have to work in an 802.1x networks but some of them inconsistent and behave unpredictably. Drivers sometimes make things worse. Some users have obsolete devices that have no support of this feature at all.

Protecting company data stored on cloud facility

This paper doesn't cover Layer 3 and upper lever security. Other protection techniques should be implemented to secure educational services inside network. Furthermore, most corporate cloud services are presented in the Internet, so have already had inherit security.

C. Policy and regulation challenge

Local government regulations and laws.

In order to the Russian Federation's law it is illegal to give someone unauthorized and anonymous access to a public network. Any network operator providing public access to a Wi-Fi network must collect user's personal data, such as name, surname or any ID. As long as there is no liability for this infringement a lot of organization still use open Wi-Fi without registration but it is not completely legal. So, Russian educational institutions are limited with using Wi-Fi without password and registration.

Ethical and privacy issues

The main issue with wireless security is the control on the communication medium. Open system can't get security for anyone but presents easiest and uncontrolled way of connection. EAP-TLS is the most secure and reliable for IT suppliers but difficult and undesirable for users. PEAP-MSCHAPv2 puts burden of security on users but more controllable and considered in this paper as tradeoff between easiness and security.

D. Human Aspect Challenge

Training and education users

The hardest part about implementing secure Wi-Fi appeared training the users. Users today have incredibly high expectations for ease of use. They also have more options than ever before to work around official access. If the network is too hard to use, they'll use data. If the certificate is bad, they will ignore it. If they can't access something they want, they will use a proxy.

User reactions, emotions and compliance of BYOD policy

This survey has assessed user's feedback. Measurements of overall level of our customer satisfaction found that easy access to the network is very important for students. Most student and employees tend to sacrifices security to the simplicity.

IV. RESULTS

In the 2013-2014 our students and employee had to apply for a PEAP-MSCHAPv2 login (or EAP-TLS certificate) personally. Dedicated specialist on a regular basis was responsible for registering new users, and consulting them. We had about 56 and 96 requests for a 2012-13 and 2013-15 year respectively (Figure 1).

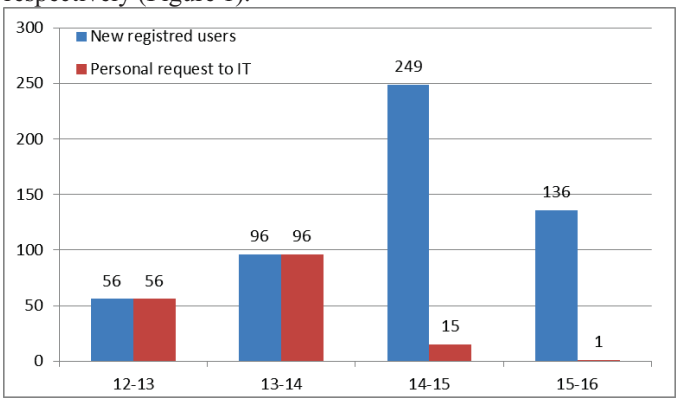


Fig. 1. Amount of new registered users per educational year

Measurements of overall customer satisfaction level found that easy access to the network is very important for students. In the 2014 the department implemented fully automated registration for students and employees through CIS (Corporate Information System) [15][12]. All students and employees registered and authenticated in the CIS is able get personal login PEAP-MSCHAPv2 in "one click". It caused increase of new users in the network, and, at the same time, it caused cutback of requests to the IT-support. In the 2015 we implemented some tutorial and advertising campaign and then discarded IT specialist relating to Wi-Fi on an every-day basis.

Total amount of student and employees in 2012-13, 2013-14, 2014-15 and 2015-16 is 532, 446, 480 and 509 persons each year respectively.

Easy access to registration and absence of every-day IT support resulted in slight increase of "never connected users", who got login but never managed to connect (Figure 2).

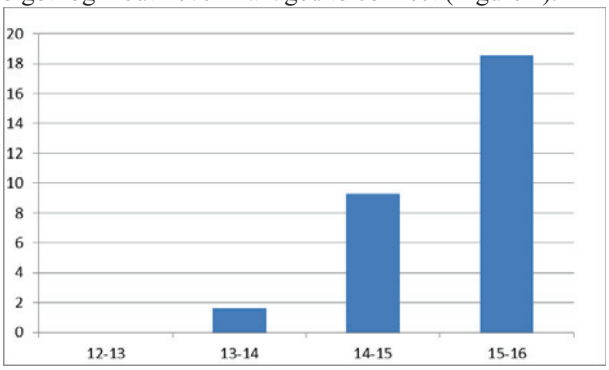


Fig. 2. Percent of registered but never used network users per year

The department asked students and employees to estimate importance of security and simplicity from 0 to 10. 0 – means no security at all (open Wi-Fi), 10 – means EAP-TLS and necessity installing certificates. 42% of the students and employees don't want to use any reliable security protocols and

consider it as disturbing. Only 17% are ready and will use certificate-based EAP-TLS instead of password-based PEAP-MSCHAPv2.

Finally, survey found that 43% of our students don't use CA certificates to authenticate the server during PEAP-MASCHPv2 based authentication. Consequently, "Rouge AP" and "Men-in-the-middle" attacks are the biggest threat against an educational wireless network. While WPA2 sets up a very secure connection, we also have to be sure that the users will only connect to the official network. A secure connection is meaningless if it's to a honeypot or imposter signal. Institutions often sweep for and detect rogue access points, including Man-in-the-Middle attacks, but users can still be vulnerable off-site. Even if the server has a certificate properly configured, there's no guarantee that users won't connect to a Rouge SSID and accept any certificates presented to them. Some operating systems make things worse: Android doesn't use CA certificate by default and more tend to compromise PEAP-MSCAHPv2 login and password connecting to Rogue AP. Windows, iOS and Macintosh show warning messages then certificate changes but users don't take it seriously most of the times. Android OS are very popular among students end employees (Table 1) and 26.24% may come across Rogue AP without warning messages.

TABLE I. USAGE OF OPERATING SYSTEM IN THE DEPATRMENT

Operating system	*Percent of users*
Windows	51,68%
Android	26,24%
iOS	12,10%
Macintosh	6,40%
Linux	1,89%
Windows Phone	1,65%
Unknown	0,02%
BlackBerry	0,01%

V. CONCLUSION

It is a vital to allow access for BYOD users to an educational network. Most educational institutions are using authentication. Security and simplicity are two desirable but incompatible features that need to compromise. But since security is burden of users they tend to use simple ways of connection compromising network even in case we make deal with experienced users. Therefore, level 2 cannot be the single level of security, but needs to be covered by security on upper levels.

REFERENCES

[1] V. Korolyov, D. Grinshpun, "The technology of surpassing studying in the system of continuous qualified education", IACEE World Conference on Continuing Engineering Education, WCCEE 2012.

[2] Teena Maddox, "BYOD, IoT and wearables thriving in the enterprise", 2016, [Online] Available: http://www.techproresearch.com/article/byod-iot-and-wearables-thriving-in-the-enterprise/

[3] Ranjana Shukla, Samad S. Kolahi, Robert freeth and Avikash Kumar, "Educational Institutes: Wireless Network Standards, Security and Future", Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference.

[4] Amirali Sanatinia, Sashank Narain, Guevara Noubir, "Wireless Spreading of WiFi APs Infections using WPS Flaws: an Epidemiological and Experimental Study", 6th Symposium on Security Analytics and Automation 2013

[5] Lui Yong-lei, "Defense of WPA/WPA2-PSK Brute Forcer", 2015, 2nd international Conference on Information Science and Control Engineering.

[6] IEEE 802.11-2012 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification (2012)

[7] Kathleen Downer, Maumita Bhattacharya, "BYOD Security: A New Business Challenge", 2015 IEEE International Conference on Smart City/SocialCom/

[8] Simon Denman, "Why multi-layered security is still the best defence", Network Security, Vol 2012. Issue 3.Pp. 5-7

[9] Rhodes J (2013), "Building Security Around BYOD. Managing Mobility, Rough Notes". Vol 156. Pp 104,114.

[10] V. Shakhalov, L. Zinchenko, E. Rezchikova, A. Glushko, An Opportunity in Engineering Education Russian BYOD Tendencies, 2015 International Conference on Interactive Collaborative Learning (ICL).

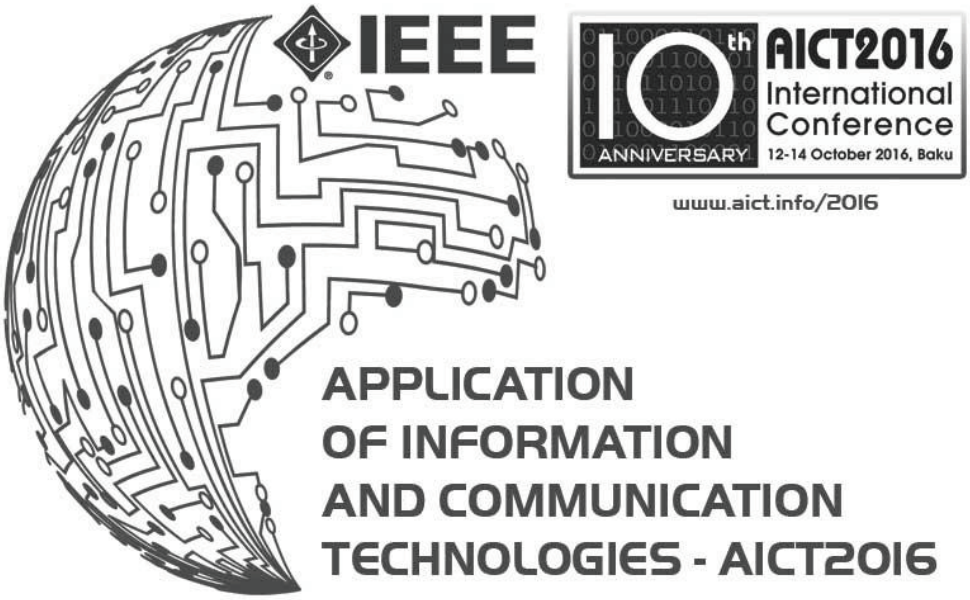
[11] Sara Ali, Muhammad Nauman Qureshi, Abdul Ghafoor Abbasi, "Analysis of BYOD Security Frameworks", 2015 Conference on Information Assurance and Cyber Security (CIACS)

[12] V.Korolyov, D. Grinshpun, R. Nuretdinov, "Change management in continuing engineering education in IT (on the example of secondary vocational education faculty of ITMO University)", 15th world conference of international association of continuing engineering education, IACEE 2016.

[13] N. Asokan, Valtteri Niemi, Kaisa Nyberg, "Man-in-the-Middle in Tunneled Authentication Protocols" Nokia Research Center, 2002.

[14] Cisco, "Wired 802.1X Deployment Guide", [Online] Available http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Security/TrustSec_1-99/Dot1X_Deployment/Dot1x_Dep_Guide.html

[15] A. Lukin, R. Nuretdinov, D. Grinspun, "Electronic system of controlling educational process with elements of results predictions", Sociosphere: Modern technologies in system of additional and professional education, 2014.



CONFERENCE PROCEEDINGS

SESSION 4.

LATEST TRENDS IN ICT APPLICATION

12-14 October 2016, Baku, Azerbaijan

Beacon AuthPath

Augmented Human Path Authentication

Emin Huseynov
Faculté des Sciences de la Société
University of Geneva
Geneva, Switzerland
emin@huseynov.com

Jean-Marc Seigneur
CUI, ISS & Medi@LAB, Faculté des Sciences de la Société
University of Geneva
Geneva, Switzerland
seigneuj@gmail.com

***Abstract*— BLE (Bluetooth Low Energy) beacons are being deployed in smart cities, especially to augment the shopping experience of customers in real outlets. Thus, humans as they walk in cities pass by different beacons and the sequence of beacons form a path. In this paper, we present how those augmented paths can authenticate a user in a secure way whereby the users can prove they have passed by a path, even with current unsecure beacons. We have built a prototype to validate this new authentication scheme with unsecure Estimote beacons. In addition, as an alternative to BLE, a similar system utilizing Wi-Fi to detect user proximity is reviewed as well.**

***Index Terms*— Authentication, Physical security, Augmented Reality.**

I. Introduction

As per Stajano [1] and Dey [13], ubiquitous computing with all its sensors embedded in the environment and carried out by humans has open the door for novel context-aware authentication schemes. Determining exact geographic location of a person is always seen as a way of proving identity. While there are obvious ways of implementing this, such as determining coordinates using a space-based system, such as Global Positioning System (GPS) or getting location information based on Internet connection used, they are not accurate and trustworthy: IP address based information's accuracy is generally limited to determining a city/village/neighborhood, GPS's can only be used outdoors and in areas with direct satellite coverage. Furthermore, both methods are relatively easy to forge, jam or spoof.

In this paper, we present how augmented paths can authenticate a user in a secure way whereby the users can prove they have passed by a path, even with current unsecure beacons. In Section II, we survey related work. Section III details our new scheme: its model, use-cases and alternative implementations. In Section IV, we present how we have validated this new authentication scheme with unsecure Estimote beacons.

As the technology is based on the new Bluetooth standard (4.0), older devices will not be supported [3]. As an alternative, in one of the use-cases, we will replace BLE based broadcast with Wi-Fi SSID based broadcasts. This will help overcome the limitations of Bluetooth 4.0.

II. Related Work

Electronic geo-fencing [2] is a technique that has been proposed to ensure that people, devices and machinery are accessed in or from authorized physical locations only. In contrast, our scheme focuses on secure paths with currently deployed BLE technology.

Authentication using a virtual iBeacon as a second factor has been used in the product offered by SAASPASS [15]- the application installed on a user's smartphone automatically transmits the generated onetime password to a special connector (a BLE listener daemon- currently only available for MacOSX). This connector searches for BLE packets transmitted in the near or immediate range and if the packet parameters match, passes the values gathered to the application (e.g. a browser) emulating keyboard keys.

The same idea, but using hardware BLE Beacons has been researched by van Rijswijk-Deij [16]. The main concept is to replace the static BLE device identifiers by dynamically changing attributes that can serve as one-time passwords to be used for authentication.

IT security is not the only area where BLE beacon technology may appear; Bluesmart [12], marketed as a "smart suitcase" relies on BLE beacon technology for physical security: the owner can open the suitcase by just approaching or tapping on an app installed.

Although the use Wi-Fi as a proximity base was already researched [17], transmitting OTPs as a part of Wi-Fi SSID seems to be a new idea; we were not able to identify any related or similar work in this area.

III. Beacon AUTHPATH Model

In this section, we describe our new Beacon AuthPath model including two use-cases. The model is based on authenticating a user's physical path by checking proximity to a set of geographic locations (checkpoints) in a predefined order with proximity distance varying depending on technology used. When the user with a mobile device running the AuthPath application approaches a checkpoint, the application records various parameters, such as current time and the data broadcasted by the beacon device (beacon data) the checkpoint is equipped with. The same procedure repeats on with other checkpoints the path consists of. This data then is

transferred to a validation server, which checks the submitted data using predefined algorithms and accepts or denies the person as an authenticated user.

The first case is based on a system with standard beacons; other two use-cases will utilize enhanced beacons, called "smart beacons", based on BLE and Wi-Fi SSID broadcast to transmit the security context information to the mobile application.

The security level of the first use-case is low as the beacon data for all checkpoints are static and therefore the use-case is vulnerable to replay attacks: the beacon data may be passed from a user physically present in the checkpoint to any remote user, allowing the latter to forge the authentication path. In the same time, the low cost and availability of standard beacon devices make this use case very easy to implement, therefore this use-case may still be of an interest in scenarios where low security level is tolerable (e.g. if authentication path is used as a second factor or in augmented sports etc.)

The other two use cases use dynamic beacon data, periodically changing similar to one-time password mechanisms, therefore the attack window per every checkpoint is very narrow. With a large number of checkpoints used in an authentication path and very short beacon data rotation period (even 1 second is possible if system times of all devices are correctly synchronized), the replay attacks become practically impossible.

A. "Beacon AuthPath" with standard beacons

1) Standard BLE-based beacons

One of the possible use-cases can be a system for verifying a path passed by a person within a wider geographical area, such as a smart city. Standard BLE beacons transmit radio packets consisting of four main pieces of information: "UUID" – a 16-byte unique identifier of the device, "Major" and "Minor" - 2-byte string and "TX Power" – value used to determine the distance between devices [3].

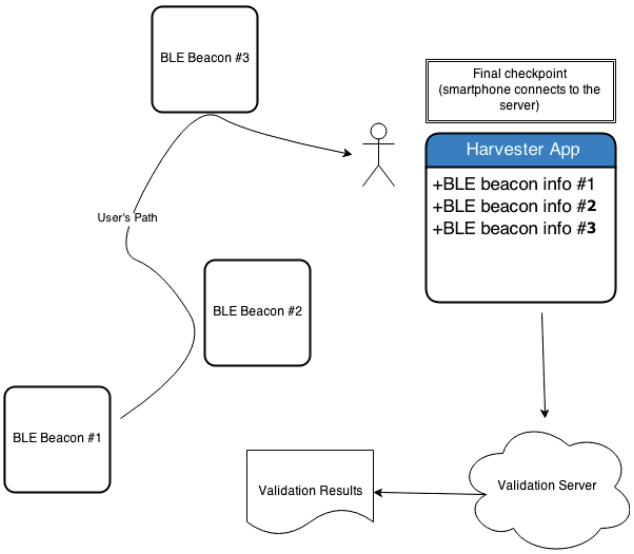


Fig. 1. AuthPath sequence diagram

The mobile application used in this use case will detect this information, save locally together with current timestamp, and transfer the harvested data to the server at the final checkpoint. To protect from the replay attack (e.g. generating a virtual beacon with the same parameters) the mobile device then initiates changing the value of "Major" variable of the BLE device to a random value using a special SDK built in the application. The app sends the new value to the server after it assigning it to the beacon device. However, this is only possible if the SDK of BLE device allows such manipulation; the availability and mechanism of this feature is purely on manufacturer's' discretion and is not defined in iBeacon or any other specifications.

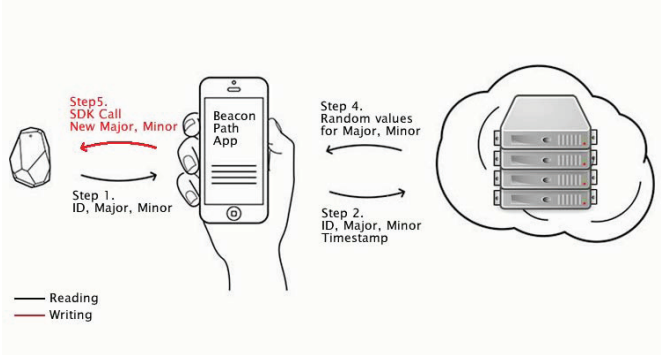


Fig. 2. Beacon AuthPath data flow

This use case makes use of standalone BLE beacons that can be managed via Bluetooth using an SDK call, and with no security mechanisms in place. Estimote beacon [4] is an example of such device and is planned to be used in the proof-of-concept implementation.

2) Security analysis

The model with standard beacons provides limited security, as the beacon broadcasts are publicly visible. With the scenario where the app uses SDK calls to modify beacon's parameters, the risks of "replay" attacks are minimal (as the server generates and sends new values only to the last user passing the checkpoint). However, as the Estimote's SDK provides no protection, the authentication sequence is insecure. This is a major obstacle to using the given scenario in systems requiring higher level of security.

B. "Beacon AuthPath" with "smart" beacons

1) BLE-based "Smart Beacons"

This is a slight modification of the previous use case where standard beacon devices are replaced with compact devices ("smart" beacons) running an operating system capable to periodically run simple scripts. This will allow avoiding modifying beacon values when passing the checkpoints; instead, the major value will be regenerated automatically every N seconds. This also removes the requirement of the mobile device to be online, the collected beacon information can be stored locally and uploaded to the server in bulk at any convenient time or place, e.g. when reaching the final destination point. Each "smart" beacon will have a secret hash key (not visible, stored in device's internal memory). The Major value broadcasted by smart beacon device is generated

using TOTP [5] algorithm based on current time and the secret hash key. When user approaches the beacon, the broadcasted packet information (UUID and Major) is saved to mobile devices' memory together with current timestamp. After the path is completed, the array of information gathered at all checkpoints and afterwards uploaded to the server. The server will have a copy of hash keys of all smart beacons and is able to analyse the submitted matrix and verify whether it contains valid "Major" values, by rerunning TOTP algorithms with the submitted timestamps and stored hash keys, and comparing with the submitted "Major" values.

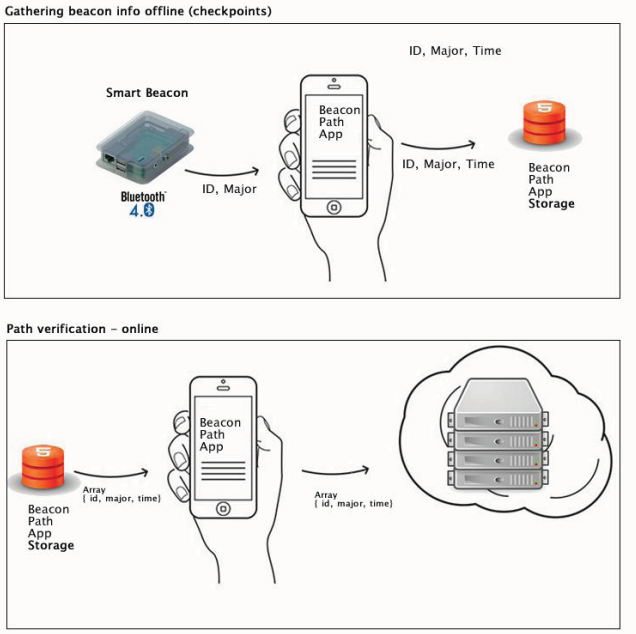


Fig. 3. Beacon AuthPath with smart beacons data flow (online and offline)

2) Wi-Fi SSID based "Smart Beacons"

This use-case is similar to BLE Based "Smart Beacons" but uses Wi-Fi SSID as the mean of transmitting the broadcasts from the smart beacon to the mobile application. Using Wi-Fi instead of BLE will allow using older hardware not supporting BLE and/or provide redundancy in the event that Bluetooth has been switched off on the client device.

The Wi-Fi smart beacon will emulate an access point and broadcast an SSID in a special format that will change every 30 seconds as per TOTP specifications. The format will contain a constant system id, location id and a one-time password; having a constant system and location id will allow distinguishing the SSID amongst others, whereas one-time password will be used to authenticate the user and validate a path. By analogy with BLE smart beacons, the location ID will serve as device's UUID.

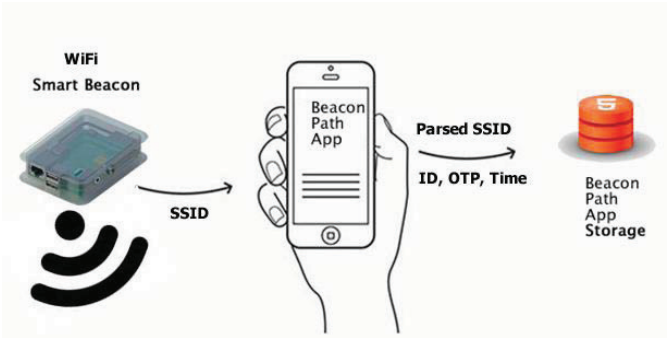


Fig. 4. Wi-Fi Smart Beacon data flow

Implementing SSID broadcast is possible on any commodity hardware including miniature systems like Raspberry PI or even SOHO Wi-Fi routers. The routers running any version of Linux based operating system, such as DD-WRT are relatively easy to be reprogrammed in order to transmit periodically changing SSID broadcasts [19].

Only SSID scan functionality of the mobile device is utilized in this use-case, the mobile application only needs to list down the available SSIDs and there is no need to connect to the Wi-Fi beacon's network. The possibility of searching for other SSIDs broadcasted while connected to one is confirmed and easily reproducible and this means that Wi-Fi smart beacons can be used even while connected to a Wi-Fi network. Once the list of SSIDs have been obtained, the application should search the list for a specific prefix and then parse it to obtain the location ID and the one-time password. E.g. if the format of SSID is "WIFIBEACON_XXX_YYY_ZZZ " , the prefix to look for will be WIFIBEACON, and further on, XXX- is the system ID, YYY- the location ID, and ZZZ is the current one-time password.

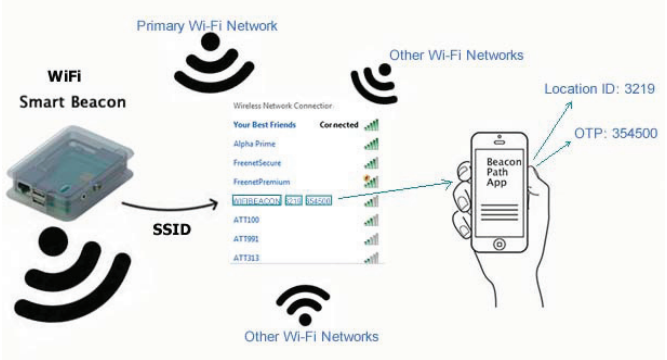


Fig. 5. Wi-Fi Beacon SSID Parsing

3) Security analysis

In classic implementation, BLE beacons provide limited level of security, primarily due to the possibility of detecting and cloning beacon IDs. However, the model with "smart" beacons provide higher security in comparison with previous model. Although, the smart beacons are relying on the same BLE technology and attackers can still intercept broadcast packets (the same applies to Wi-Fi beacons, as the SSID is broadcasted in plain-text as well), the risks of "replay" attacks are minimal due to the limited time of validity of the data broadcasted: in majority of TOTP implementations, the period

is not more than 30 seconds [5]. In addition, the app should also determine device's GPS coordinates and submit together with harvested broadcast information in order to verify the exact geographic location of the user at the time when the point validation occurred. Additional environmental parameters, such as temperature can be used to avoid wormhole attacks [14]. Additionally to safely identify a device, the parameters transferred to the validation server should be encrypted with device specific key.

IV. Beacon AUTHPATH Validation

In this section, we first detail our model implementation with Estimote devices. Then we discuss what the validation of this prototype has underlined including areas for extensions and current limitations.

A. Prototype Implementation with Estimote

We have developed a prototype path verification app using Estimote Android SDK using Apache Cordova [9]. The app is a combination of an HTML5 interface connecting to android library using Cordova. The object passes nearby (within BLE's allowable distances of 10 cm – 70 meters) certain points equipped with BLE packet emitter or harvester devices. Prototype app has a sample path containing four checkpoints. Two checkpoints were equipped with physical Estimote beacons, 2 remaining checkpoints used virtual beacon apps running on iPhone 5s devices. The interface consists of a graphical path showing the current location of the user and a list of checkpoints. Once a user approaches a checkpoint, the relevant checkbox becomes checked, and the user icon animation of moving to the checkpoint appears. The algorithm checks the order of the beacons ranging, e.g. if checkpoint #3 has been visited before #2, the checkpoint is ignored. For the prototype, any proximity is being considered valid (the TX power value is not checked). The Minor value of beacons are used as their identifications.

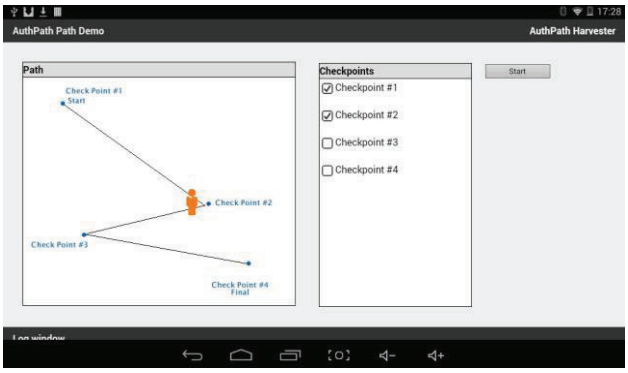


Fig. 6. AuthPath application interface

B. Prototype Validation Outcome Discussion

A simple implementation of an Estimote SDK based app has demonstrated how a physical path can be validated using a set of BLE beacons. The exact approach can be used for a real-life application with some improvements: the identification of BLE beacons should be based on UUID with Estimote security enabled, and the validation should be done with integration of Estimote Cloud API into the application. The accuracy of the proximity authentication can be improved by checking the transmission power of checkpoints' BLE beacons (RSSI value), e.g. if we want to validate a checkpoint as passed only if the distance was below 20 meters. Although this prototype is vulnerable to replay attacks as described in 3.1.1, it provides a solid base for creating a more secure implementation, where standard beacons can be replaced with smart devices and the path validation data needs to be sent to the sever at the final checkpoint. It is also important to include GPS coordinates in the array of checkpoints and implement device key encryption before submitting to the validation servers.

While the security level of the presented demo application is disputable due to technology limitation, the app can be used in less strict environments. A possible use area might be Augmented Sports; namely, a smaller scale version of Amateur radio direction finding (also known as radio orienteering) [10] where the only equipment required would be a smartphone and a path of BLE beacons need to be followed showing the best speed and accuracy possible.

An alternative to Apple's iBeacon communication protocol, Eddystone by Google [21] should also be researched in this context. In addition to standard BLE beacon features, Eddystone provides a new type of packet: Eddystone-TLM, which provides telemetry data (TLM stands for "telemetry"). Telemetry data can help to verify the authenticity of a beacon with further accuracy. Different from iBeacon that only operates with numeric IDs, Eddystone can transmit data packets in URL format, which makes it much easier to use for application development.

As already described above, older hardware and mobile operating systems do not fully support Bluetooth 4.0. In addition, many users still perceive Bluetooth as a battery hog. Therefore, for a number of real-life applications BLE might be a less favorable option.

We are proposing the following as alternative solutions to avoid using BLE in the use-cases researched:

1) Near field communication (NFC) [18]– a technology similar to iBeacon working on shorter distances (~ 10 cm)

2) Signals of opportunity (SoOP) [7] based geo-location. This solution uses no beacon devices at all and is more accurate outdoors. However, some companies are promising to provide better accuracy [8]

For Wi-Fi smart beacons, we have done a surface research and it seems to be quite easy to be accomplished, however it is not supported out of the box by Cordova/PhoneGap and a custom plugin will need to be developed [20].

V. Conclusion

In this paper, we have proposed a new authentication model based on physical location of a person wearing or passing by BLE beacons. We have validated a part of the overall concept - BLE Beacons based path tracking - with a real implementation and critically evaluated its weak points and proposed potential improvements. Future work will involve larger scale deployment and tests in the shops of a real smart city in the

mountain, i.e., a smart ski resort. Creating prototype Smart Beacons, both BLE and Wi-Fi based should be included as a part of further research.

REFERENCES

[1] F. Stajano, "Security for Ubiquitous Computing." John Wiley & Sons, 2002

[2] Chen R., Guinness R. Geospatial Computing in Mobile Devices. Artech house, 2014

[3] What is iBeacon? A Guide to Beacons | iBeacon.com Insider [WWW Document], n.d. URL http://www.ibeacon.com/what-is-ibeacon-a-guide-to-beacons/ (accessed 11.6.14)

[4] Estimote Beacons — real world context for your apps [WWW Document], n.d. URL http://estimote.com/ (accessed 11.6,14)

[5] RFC 6238 - TOTP: Time-Based One-Time Password Algorithm [WWW Document], n.d. URL https://tools.ietf.org/html/rfc6238 (accessed 11.6.14)

[6] Bring your own device - Wikipedia, the free encyclopedia [WWW Document], n.d. URL http://en.wikipedia.org/wiki/Bring_your_own_device (accessed 11.6.14)

[7] Chen, R. Ubiquitous positioning and mobile location-based services in smart phones. Hershey, PA: Information Science Reference. 2012

[8] Merlin [WWW Document], n.d. URL http://merlintek.com/ (accessed 11.7.14)

[9] Apache Cordova [WWW Document], n.d. URL http://cordova.apache.org/#about (accessed 11.26.14)

[10] Amateur radio direction finding - Wikipedia, the free encyclopedia [WWW Document], n.d. URL https://en.wikipedia.org/wiki/Amateur_radio_direction_finding (accessed 11.17.14)

[11] What is Secure UUID and how does it work? – Estimote Community Portal [WWW Document], n.d. URL https://community.estimote.com/hc/en-us/articles/204233603-What-is-Secure-UUID-and-how-does-it-work- (accessed 11.26.14)

[12] Bluesmart: World's First Smart, Connected Carry-On | Indiegogo [WWW Document], n.d. URL https://www.indiegogo.com/projects/bluesmart-world-s-first-smart-connected-carry-on (accessed 11.26.14)

[13] Anind K. Dey. Understanding and Using Context. Personal and Ubiquitous Computing. 2001

[14] YounSun Cho, Lichun Bao, and Michael T. Goodrich. LAAC: A Location-Aware Access Control Protocol. In 2006 Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services. IEEE, 2006

[15] Computer Connector - Two-factor Authentication SAASPASS [WWW Document], n.d. URL https://www.saaspass.com/download/two-factor-authenticat-computer-connector-apple-mac.html (accessed 11.17.14)

[16] Van Rijswijk-Deij, Roland. Simple Location-based One-time Passwords - bringing location to the cloud in a secure and simple way. Radboud University Nijmegen. 2013

[17] D. Namiot and M. Sneps-Sneppe "Wi-Fi proximity as a Service", Think Mind SMART 2012, The First International Conference on Smart Systems, Devices and Technologies. 2012

[18] About the Technology. NFC Forum. [Online] Available at: http://nfc-forum.org/what-is-nfc/about-the-technology/. (accessed 23.02.15)

[19] Heldenbrand, David, and Christopher Carey. "The linux router: an inexpensive alternative to commercial routers in the lab." Journal of Computing Sciences in Colleges 23.1 (2007): 127-133

[20] Cordova - Can Phonegap query the wifi system and return network names? - Stack Overflow. 2015 [ONLINE] Available at: http://stackoverflow.com/questions/4824841/can-phonegap-query-the-wifi-system-and-return-network-names/6769269#6769269. (accessed 23.02.15)

[21] "Beacons | Google Developers". 2015. [ONLINE] Available at: https://developers.google.com/beacons/?hl=en (accessed 23.08.15)

Improving Collaborative Recommender Systems via Emotional Features

Soghra Lazemi
Department of Computer Eng., Faculty of Electrical & Computer Eng
The University of Kashan
Kashan, Iran
Soghra.lazemi@gmail.com

Hossein Ebrahimpour-komleh*
Department of Computer Eng., Faculty of Electrical & Computer Eng
The University of Kashan
Kashan, Iran
*Corresponding author: Ebrahimpour.kashanu@gmail.com

Abstract— **Nowadays, by communication networks expansion, recommender systems play an important role in our daily life. Recommender system tries to recommend items that are attractive and pleasant for user. At the same time, collaborative filtering approaches is one of the most successful approaches which recommend to the customer other users pleasant items which have similar interests with him/her in the past. In this paper, we present a novel approach of these systems by entering user's emotions in recommender system. Then, by emotion matrix definition, the users who are similar emotionally are searched to improve the performance of traditional user-based collaborative filtering (UBCF). The proposed algorithm offers items to him/her by considering user's current emotion which users with similar emotion like them. This algorithm identifies suitable number of neighbors automatically and chooses high-quality neighbors to do predictions by applying pre-processing phase. Our proposed algorithm has a solution for user cold-start problem. Evolution results in Jester database show that proposed method has better prediction accuracy and makes high quality recommendations.**

Keywords- Recommender System, Collaborative Filtering, Emotional Category, Social Recommender Systems.

I. INTRODUCTION

Increasing communication networks make access to information easier but, on the other hand, it develops a huge amount of information and makes people confused. So, users need systems that provides useful information through a massive amount of information according to their interest. The system that provides attractive recommendations for users by using their personal information, called recommender system. Generally, these systems are divided in two categories: traditional recommender systems and recommender systems based on social networks. Different approaches had presented for traditional recommender systems that contain: 1- collaborative filtering approaches (CF) 2- content-based recommendation techniques (CB) 3- knowledge-based recommendation techniques (KB) 4- hybrid-based approaches (HB) [1].

Collaborative filtering approaches main idea is that similar users have similar interests, in other words, if two users had similar interest in the past, they will have similar taste in the future. In this method, by any user's feedback, which can be identify explicitly (such as assigning a rank to an item by user) or implicitly (such as user clicks on a link, buying an item) is used to calculate the similarity. We can find similar users and offer items among users when a large number of them and their feedbacks are available. CF methods do not need any information about the items [2].

CB method is based on additional information about the items and user preferences (user profile) and it doesn't need the existence of great community of users or rating history. Even if a user is present, the recommendation list would be prepared. The effort of this method is learning user preference and offering items that is similar to user preference. The challenge subject is qualitative features achievement that shows people imagination from item's exterior design.

In both CF and CB methods, people interest is assumed stable over the time, whereas this is not the case in reality. KB method presents recommendations based on user needs. In the other words, it presents the items to him/her that is proportionate with user needs and by this way it also considers probable changes during time.

Hybrid recommender systems combine several algorithms or recommender system, or they use several knowledge sources as input. Different hybridization designs had presented that include: 1- monolithic exploiting different features 2- parallel use of several systems 3- pipelined invocation of different systems [3].

Variety algorithms are suggested about CF that we now discuss them. The first suggested approach is PureCF method that gets user-item rating matrix as input and produce output in the form of: 1- numerical prediction that shows current user interest (like or dislike) to a specific item. 2- The list contains recommended top-n items that current user didn't purchase them yet. One of oldest model of this method is user-based nearest-neighbor (UBNN) that seeks peer users which have similar preferences with current user in the past, and then,

calculates prediction based on rating for p, with peer users for each p items that current user had not ever seeing it yet [4]. Because rated matrix is very large and sparse, the methods using matrix factorization techniques such as singular value decomposition or principal component analysis (for extracting a set of latent factors of ranking patterns), for create offline complex model that could produce immediate online predict are presented [5, 6].

Another method to predict which item with which rate will be liked by a user is using of probabilities theory. These methods break initial CFs assumption and believe that it can be possible two users have similar interests toward a subset of items and have completely different behavior with each other toward other subset of items. Therefore, it tries to group same items and users in subsets. Bayesian belief networks, clustering models and probabilistic latent semantic analysis are example of this category [7, 8]. The initial and simplest method to perform CF with probabilistic method is seeing prediction problem as a classification problem [9]. The methods based on clustering put users into distinct groups. The number of unknown clusters and parameters could be estimated with EM algorithm [10]. Variety algorithms are presented for clustering. In [11] a new algorithm is presented for clustering based on user-item subsets to improve CFs performance. Association rule mining methods that usually are used for purchase behavior analysis emerged about RSs with rules automatic discovery targeting such as: if the user likes item 1 and item 2, in a very probable manner he/she will like item 5[12].

The reviewed methods perform prediction act by using user-item rating matrix. In [13] different approach, slope one predictors, is recommended that has better performance and unlike simplicity had recommendations with good quality. The main idea of this method is simple and is based on subtraction between items for users. Subtraction average is used as prediction.

Available information in user's social network utilization was proposed to improve CF approach accuracy. In the real world, people often ask about opinion and inquiry from their friends to purchase products and so on in social networks. Present findings in the field of sociology and psychology show that, people tend to think of people similar themselves [14]. So, people share their personal believes with each other in social networks and receive safer recommendations from their friends. Famous online social networks like Facebook [15], Twitter [16] and YouTube [17] have created new method for communication among users. Systems based on social networks increase accuracy of traditional recommender systems by using communication between users as additional inputs [18]. One of the most widely used approaches, is CF-based social RSs approach. These approaches are divided into two categories: 1- Matrix factorization (MF) based social recommendation approaches 2- Neighborhood based social recommendation approaches.

In the first approach, social information about users and user-item feedback history (such as rating, clicks, purchases) are collected to improve the accuracy of traditional recommender systems based on MF. The MF basic idea is user-item interaction modeling with the factors that shows users hidden features and items (such as user-friendly class or category of items) in system. This model is trained by using existing data and then used to predict user's rating for new items [19, 20]. The second approach includes Social Network Traversal (SNT) and nearest neighborhood approaches. Approaches based on nearest neighborhood directly use stored rating for prediction. These approaches navigate user neighbors in social networks and then ask them target item rate [18]. Most of exiting methods for social RSs explore entire network. [21] proposes circle-based recommendations for online social networks. In [21] the author has offered a set of algorithms to get user followers in a particular field. The main idea is that, a user trusts all of his/her friend just by attaining to a context without considering other contexts. They divided social networks into several sub-networks.

In this paper, we have extended traditional user-based CF approaches by entering information about user's feelings and have tried to improve the performance of these systems by considering user's current feeling. Also, our proposed method can solve the problem of determine appropriate number of neighbors by using of emotion matrix. Most previous works have focused on the accuracy of predictions but, our method improves the accuracy of predictions, and by selecting the appropriate neighbors, improved the quality of recommendations too.

The remaining of this paper is structured as follows. Section 2 describes our proposed algorithm in detail. Experimental results and discussion are given in Section 3. Finally, conclusion and future work are described in Section 4.

II. THE PROPOSED METHOD

Emotion is one of the important elements of human nature that has a significant impact on his/her behavior and choices. Much researches in the area of psychology, neurology, and behavioral sciences shown that individuals choices are related to their feeling and mental moods [22, 23, 24]. The main objective of this paper is presentation of a recommender system that considers users' emotion and mental moods. To achieve this goal, we try to find users with similar feelings and examine its effect in improving UBNN approaches performance. There are two main questions:

1. How can obtain the users feelings?
2. How we can enter feelings in UBNN approaches?

Our proposed method answers both questions- we attain feelings explicitly by using emotional category approach and by defining emotion matrix, we present a method to insert feelings in UBNN approaches. Emotional category approaches are based on discrete emotional classes that assume users have a limited set of discrete emotions. In this paper is used Ekman base model [25], that categories user's mental moods in six

	I1	I2	I3	I4	I5
U1	HAPPINESS	SADNESS	HAPPINESS	FEAR	HAPPINESS
U2	FEAR	SADNESS	HAPPINESS	HAPPINESS	HAPPINESS
U3	SURPRISE	DISGUST	ANGER	SURPRISE	SADNESS
U4	SADNESS	SADNESS	HAPPINESS	FEAR	FEAR

Fig. 1. Emotion matrix

sets, include anger, disgust, fear, happiness, sadness and surprise. When the $user_i$ rates $item_j$, he is asked to choose one of the categories for display their feeling according to his/her own internal feelings. To enter emotions, we define the new matrix called emotion matrix. TABLEI shows an example of rating matrix for four users and 5 items. When user rates an item, her/his mental moods store in emotion matrix. For example, in Fig.1 when user1 rates item1, he/she has "happiness" mental mood.

Our recommended method is added to traditional UBNN approach by two different ways. In the first case, we just use emotion matrix and we don't use rating matrix in our calculations. First, we find the most similar people (neighbors) in terms of the emotion to the current user. To calculate similarity, we use combination of Jaccard and mean squared difference similarity measures. We can obtain the final formalization by combining formula (1) and (2), is defined as formula (3).

$$sim(u,u')^{MSD} = 1 - \frac{\sum_{i \in I} (r_{u,i} - r_{u',i})^2}{|I|} \quad (1)$$

$$sim(u,v)^{Jaccard} = \frac{|E_u| \cap |E_{u'}|}{|E_u| \cup |E_{u'}|} \quad (2)$$

$$sim(u,u') = \alpha sim(u,u')^{Jaccard}.(1-\alpha) sim(u,u')^{MSD} \quad (3)$$

We assume that E= {anger, disgust, fear, happiness, sadness, surprise} is the set of user's emotions and the similarity measure is weighted by α.

Then, in the next phase, we predict the current user rating for the unknown item by using formula (4).

$$\text{Pred}(u,i) = \bar{r_u} + \frac{\sum_{u' \in N} sim(u,u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} sim(u,u')} \quad (4)$$

TABLE I. RATING MATRIX (ON A 5-STAR SCALE)

	I1	I2	I3	I4	I5
U1	1	2	3	3	?
U2	3	5	4	4	3
U3	2	2	3	3	1
U4	4	4	5	5	5

To find neighbors for the current user by using emotion matrix, we obtain the similarity of user A and user B toward all items and then, unrelated neighbors are removed among neighbors according to user mental moods toward the unknown item. For example, in Fig. 1, to predict item5 rate by user1, user2 and user4 are considered as neighbors and then user4 is deleted as unrelated neighbor.

The second case is presented to solve the problem of neighbors' appropriate numbers for current user in traditional UBNN approaches. In traditional UBNN approach, if all users be considered as neighbor for current user, not only increases computational time but also has negative effect on system performance (accuracy). The common method for reducing the number of neighbors is determining minimum threshold for users' similarity or limiting and specifying a fixed number of neighbors. The problem with this method is that, if the threshold is too large, the size of the neighbors will be low for some users, so, reducing the convergence occurred and then we wouldn't have prediction for the major of items and if the threshold value is too low, the number of selected neighbors wouldn't decrease dramatically. We can solve this problem by using the emotion matrix and obtain the appropriate number of neighbors automatically. The method is as follows:

1. By using emotion matrix, we select the user that has similar mental moods with current user and add to neighbors' vector.
2. We calculate similarity of selected candidate neighbor in the first step to current user by using RM. If similarity is greater than the threshold, it means that the candidate neighbors, is appropriate neighbor and remains in neighbors list, otherwise, it removes from neighbors list and we go back to step 1.
3. The predicted value for unknown item i for current user is calculated by using existent neighbors in neighbors.

User cold-start problem occurs when a new user has joined the system but her/his preferences are not yet known. This problem can be addressed with the help of hybrid approaches – that is, with the help of additional, external information, other strategies are also possible. One option could be to ask the user for a minimum number of ratings before the service can be used. To deal with this problem, we use emotion features and identify other users that have similar emotion with current user, then recommend those items enjoyed by them.

III. EXPERIMENTAL EVALUATION

A. Data set

Our experiences have been done on Jester Dataset1 and Jester Dataset2. Jester Dataset1 contains joke rating data that have been collected between 1999 and 2003. In this dataset, 73421 users rated 100 joke in 20 degree. As, in this dataset, sparsity level is 0.447. The Jester Dataset2 has ratings of 150

jokes from 63974 users between 2006 and 2009. To examine the emotion influence, we choose 5000 users randomly and divided new dataset by using 5-fold cross validation methods to training and testing set. According to rating matrix, we create emotion matrix randomly with this assumption that users with similar rating will have same mental moods.

B. Evaluation metrics & Results

The fundamental assumption in RSs is that, more accurately predict most preferred by users. So, many of researches are going to find algorithms that provide better predictions. In this paper, we examine accuracy of prediction in three different categories, so that, not only the system error have been evaluated but also the quality of recommended list have been examined by using recommended items rate. We consider the following three categories for measurement accuracy: 1- Measuring Ratings Prediction Accuracy (MRPA) 2- Measuring Usage Prediction 3- Ranking Measures.

We use MAE (Mean Absolute Error) metric to evaluate MRPA that is defined as formula 5:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \quad (5)$$

P_i is the calculated proposed score and r_i is the real score of $item_i$ and n is the number of existence items in test dataset.

In many applications, RSs don't offer user-friendly item but recommends items that the user may use. To evaluate this metric, we use a precision that is defined as formula 6:

$$\text{Precision} = \frac{|\text{the number of correctlyrecommendad relevant items for user u}|}{|\text{total number of recommendad items}|} \quad (6)$$

When there is an item whit rate3 in test dataset, a system that predict rate2 for this item has same absolute Error with a system that predict rate4, while theses predict will appear in different places on rated list. Furthermore, the pleasant item that is in the first place of list and the pleasant item that is in 10th place of list have the same precision. In many cases, the application presents a vertical or horizontal list of recommendations for user. These lists are long and need user search. So, items have ranked. Then the ranked list needs evaluating. To do this, we have used RankScore and NDCG (Normalized discounted cumulative gain) that are defined as formulas 7 and 8:

$$rankscore = \frac{rankscore_p}{rankscore_{\max}} \quad (7)$$

$$rankscore_p = \sum_{i \in h} 2^{-\frac{rank(i)-1}{\alpha}}$$

$$rankscore_{\max} = \sum_{i=1}^{|T|} 2^{-\frac{i-1}{\alpha}}$$

$$nDCG_{pos} = \frac{DCG_{pos}}{IDCG_{pos}} \quad (8)$$

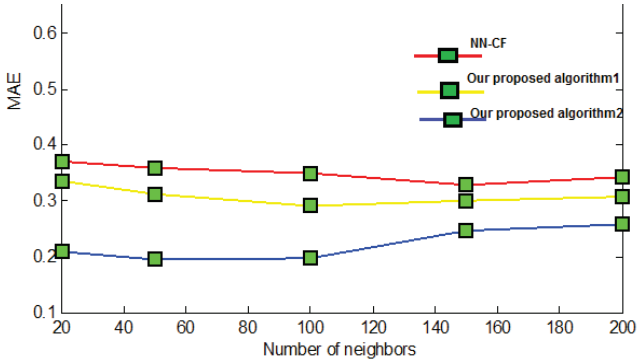


Fig. 2. Comparison of prediction quality.

$$DCG_{pos} = rel_1 + \sum_{i=2}^{pos}\frac{rel_i}{\log_2 i}$$

$$IDCG_{pos} = rel_1 + \sum_{i=2}^{|h|-1}\frac{rel_i}{\log_2 i}$$

Where h is the set of correctly recommended items, *rank* returns the position (rank) of an item, T is the set of all items of interest, α is the *ranking half life*, *pos* denotes the position up to which relevance is accumulated, rel_i returns the relevance of recommendation at position i.

To do comparison, we run traditional UBNN method with different value of k and we examined entering emotions effect on this system by implementing our proposed method.

To obtain the MAE, we ran both algorithms with same K. At each phase, k is determined automatically by our proposed method. Figure 2 shows the effect of neighbors' number on the quality of predictions. In traditional UBNN method, when a few neighbors have been used for prediction, MAE is high, and if neighbors number is too large, we have accuracy loss again.

In the first case, our proposed method only uses emotion matrix for obtaining users similarity, because of high-quality neighbors choice (neighbors very similar to the current user) has good prediction accuracy, and as shown in Fig. 2, it has reduced MAE. In the second case, few neighbors number with very high similarity with current user have remained for calculation by filtering neighbors that have less similarity with current user and this is reduced prediction error dramatically.

In TABLE II, we recorded other evaluated metrics: P, NDCG and RankScore. For comparison, we considered our proposed method performance with traditional UBCF method based on the nearest neighbor by taking k=100. The results show that the proposed method presents us good recommendations because of considering user current mental mood. This is because when the user has a mental mood like "sadness", our proposed method offers items to him that chooses by users who had "sadness" mental mood.

IV. Conclusion

In this paper, our goal is evaluating effectiveness of additional information or using another type of information

TABLE II. COMPARISON OF PERFORMANCE

	Jester dataset 1			Jester datase t2		
	TUBNN	Our proposed method1	Our proposed method2	TUBNN	Our proposed method1	Our proposed method2
P@10	0.331	0.402	0.382	0.273	0.332	0.397
NDCG@10	0.271	0.262	0.294	0.156	0.291	0.302
F1@10	0.107	0.117	0.120	0.038	0.104	0.166
R@10, α=2	0.1578	0.166	0.257	0.105	0.016	0.207

instead of information about users rating. Experiences results show that by considering users mental mood, we can use a different type of information to improve CF approaches performance. For future works, we can use this information in other methods of CF to solve presented scalability and reducing computation time problem such as MF or Cluster method and examine performance improving in these approaches.

ACKNOWLEDGMENT

Authors are grateful to Research Council of the University of Kashan for providing financial support to undertake this work.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. on Knowledge and Data Engineering*, vol. 6, pp. 734-749, June 2005.

[2] *Recommender Systems*, first ed., Francesco Ricci, LLC, 2011, pp. 44-60.

[3] R. Burke, "Hybrid recommender systems: Survey and experiments," User modeling and user-adapted interaction 12, no. 4, pp. 331-370, 2002.

[4] K. Goldberg, R. Theresa, G. Dhruv, P. Chris, "Eigentaste: A constant time collaborative filtering algorithm," Information Retrieval 4, no. 2, pp. 133-151, 2001.

[5] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Application of dimensionality reduction in recommender system," Dept. Computer Science no. TR-00-043, Technical Report Minnesota Univ., Minneapolis, USA, 2000.

[6] T. Hofmann, "Latent semantic models for collaborative filtering," ACM Transactions on Information Systems (TOIS) 22, no. 1, pp. 89-115, 2004.

[7] C.H.Q Ding, "A similarity-based probability model for latent semantic indexing*," in Proc. of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 1999, pp. 58-65.

[8] T. Georg and S. Merugu, "A scalable collaborative filtering framework based on co-clustering," *IEEE Trans. on Data Mining*, vol. 5, pp. 4-pp, 2005.

[9] C. Basu, H. Hirsh, C. William, "Recommendation as Classification," *in Proc. of the Fifteenth National Conference on Artificial Intelligence*, 1998, pp. 714-720.

[10] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Y. Yu, Z. Chen, "Scalable collaborative filtering using cluster-based smoothing," *in Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information*, pp. 114–121, 2005.

[11] B. Xu, B. Jiajun, C. Chun, C. Deng, "An exploration of improving collaborative recommender systems via user-item subgroups," *in Proc. of the 21st international conference on World Wide Web*, 2012, pp. 21-30.

[12] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," *in Proc. of the 2nd ACM Conference on Electronic Commerce*, pp. 158–1672000.

[13] D. Lemire, A. Maclachlan, "Slope One Predictors for Online Rating-Based Collaborative Filtering," *In SDM*, vol. 5, pp. 1-5. 2005.

[14] M. Jamali, M. Ester, "Using a trust network to improve top-n recommendation," *in Proc. of the third ACM conference on Recommender systems*, pp. 181-188, 2009.

[15] Twitter, available at <http://twitter.com>.

[16] YouTube, available at <http://www.youtube.com>.

[17] Facebook, <http://www.facebook.com>.

[18] X. Yang, Y. Guo, Y. Liu, H. Steck. (Mar 2014). A survey of collaborative filtering based social recommender systems. *Computer Communications*. 41. pp.1-10. Available: http://www.elsevier.com/locate/comcom.

[19] M. Jamali, M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," *in Proc. of the fourth ACM conference on Recommender systems, 2010*, pp. 135-142.

[20] H. Ma, H. Yang, M.R. Lyu, I. King, "Sorec: social recommendation using probabilistic matrix factorization," *in Proc. of the 17th ACM conference on Information and knowledge management, 2008*, pp. 931-940.

[21] X. Yang, H. Steck, Y. Liu, "Circle-based recommendation in online social networks," *in Proc. of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012*, pp. 1267-1275.

[22] L. Canales,P. Martínez-Barco, "Emotion Detection from text: A Survey," *in Proc. of the 5th Information Systems Research, 2014*, pp. 37-43.

[23] P. Rodriguez, A. Ortigos, R.M. Carro, "Extracting emotions from texts in e-learning environments," *Sixth International Conference on Complex, Intelligent and Software Intensive Systems*, pp. 887-892, 2012.

[24] R. Cowie, R.R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech communication*, no. 1, pp.5-32, 2003.

[25] *Basic emotions in Handbook of Cognition and Emotions*, first ed., Paul Ekman, 1999, pp. 45–60.

Information interaction of decentralized group control of unmanned vehicles

S V Tikhonov, R T Sirazetdinov, N R Surkin
Department of Dynamics of Processes and Control
Federal State Government-Funded Educational Institution
of Higher Professional Education
Kazan National Research Technical University named after A. N. Tupolev – KAI
Kazan, Russia
Rif-kat@inbox.ru

***Abstract*— In this article, we research decentralized group (swarm) of unmanned vehicles two-level control, as a set of material points. We provide methods for information exchange in the group and algorithms of distributed control based on imitation of different aggregate states of matter by a group of unmanned vehicles. In the article, we review three swarm "aggregate states": solid, liquid and gas, examine issues of sustainability of the system in different variants of the aggregation state simulating algorithms.**

***Index Terms*—Swarm, Group of unmanned vehicles, Two-level control, Decentralized group control, Information interaction, Aggregate states of matter, Distributed control**

I. Introduction

Robotic systems are used in very different fields of science, technology and industry for a long time. A large number of studies have been devoted to the problem of creating them. Recently, many technical, technological and algorithmic problems faced by developers of such systems are solved.

Currently, became possible to create a design small enough sizes, for example, quadrocopters which can autonomously move in space, fulfill their tasks and have a small cost. Instead of one large unmanned vehicle (UV), it allows using several small-size and low-cost UVs which jointly could perform the same tasks and would have the same advantages. These advantages include low cost and less vulnerable UVs group as a system, related to the fact that in case of losing few drones, the group is able to perform the task, invisibility for people and radars, related to small size of each drone and opportunity to disperse in space in case of group movement, ability to make decisions and transmit information more quickly, as well as greater operating range. In addition, UV group as a system demonstrates new properties that allow you to perform other kinds of tasks unfulfillable by one UV, e.g., large-scale scanning and surface research, assembly of complex structures et.

First of all, centralized control is considered as principles of control, which becomes a problem of control multilevel hierarchical system in case of increasing the number of UVs in the group. In the same time, there are some approaches of decentralized control by an object group, which now communicate with problems of multiagent systems more and more tightly [1].

Centralized control of UV group meets more and more challenges with increasing amounts of UVs. Firstly, presence of a central control UV makes the system more vulnerable and less robust, because any faulty operation of the central UV is enough for losing control of the entire group. In addition, the requirements for the CPU, RAM, central controlling UV performance, reliability and power of central UV communication system, are increasing. In this way, with a sufficiently large amount of drones group, use of centralized control principles, when every UV directly "obeys" to more "senior", becomes impractical, expensive, and even unrealizable.

Hence, it follows problem of actuality of development of decentralized control for a large group, or swarm, of UVs, when a single command is sent to all swarm, and each UV executes this command on their own [2]. All units in group shall be completely identical. They completely should repeat each other in construction and algorithms, and their actions should depend both on total command for the group and on the mutual position.

Most often, when designing the system, which is decentralized controlled group of UVs, the researchers are inspired by the creations of nature. E.g., the UVs group can simulate the movement of animals flock or insects swarm [9-15]. In this case it is possible to use analogs of quorum sensing as in bacteria [14], trophallaxis as bees, ants or termites [13], as well as hormones [15] for information exchange. Each UV realizes the so-called situational behavior [16], in accordance with the general mission and information that is received by the selected method. On the hardware level, the connection between the UVs can be implemented on the basis of radio, Wi-Fi and Bluetooth, light, sound, and even chemical signals.

II. Task Definition

We have a certain group of UVs, which consists of a large number of units

$$X_i \ (i = 1,n). \quad (1)$$

Each unit is identical to the others and its state at time point t can be described by the vector-function of state

$$\boldsymbol{x}_i(t) = [x_{i,1}(t), x_{i,2}(t), \ ..., \ x_{i,m}(t)]^T. \quad (2)$$

Then the state of the entire group will be described by the vector

$$\boldsymbol{X}(t) = [\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \ ..., \ \boldsymbol{x}_n(t)]^T. \quad (3)$$

We need to develop methods of control UVs group and of interaction between the UVs. This means we need to organize the timely exchange of actual information between the UVs and to design algorithms, which will translate the system (3) in the state required in the general task, based on the information received.

For each UV the restrictions are imposed by the environment, e.g., obstacles that he has to fly around. Also, all neighboring UVs impose restrictions on each UV. This is the minimum distance between the UVs, that is necessary to prevent collisions, and "visibility" of UV by at least one other UV, for the possibility of information transmission and prevention of losing the UV by group, et. Additional constraints may be transferred to the group in a task.

III. Control of Group Motion of Unmanned Vehicles

We will use a mobile ad-hoc Wi-Fi network to ensure communication and information transfer between the UVs. I.e. each UV or the operator will be able to communicate with any neighboring UV and send him a request to provide any sort of information and resources within the group. Also, each UV should process requests of the operator and the other UV and send the information requested.

Such a network would allow adding and removing units from the system at any given time. And at the same time, resources which are required for system operation, are shared among all UVs.

We offer a two-level principle of organization of UVs group control. The upper level implies centralized command by the operator, or the forming of target designation by automated system for a group of UVs as a whole. On the lower level, each UV fulfills own motion algorithms independently, basing on a predetermined common goal and mutual relative position to the nearest UVs.

The basic idea of group control of UVs, offered by us, is an imitation of aggregate state of matter by a group of UVs. On the lower level control, control algorithms of each UV are compiled in such a way that the group as a whole is presented to the operator either as a solid or as a liquid or as a gas. Then the high-level commands are becoming intuitively understandable for the operator. He has a finite set of high-level commands for the group control as a "substance", that is certain body, and a group of UVs independently perform these commands, based on their inherent low-level standard algorithms.

High-level commands for controlling a group of UVs as a substance can be divided into the following groups:

- Commands for conversion to a new aggregate state (AS) with the specified parameters. Conversions can be carried out from a previous AS, and from an arbitrary initial state, not corresponding to any AS.
- Commands for moving a group of UVs in a preassigned direction for the current AS. For example, moving the center of mass of "solid", the movement relative to the center of mass, etc.
- Commands related to the direct fulfillment of the goal task. For example, aerial photography of location, etc.

When group control of UVs is being implemented, not real properties of matter simulation in the appropriate AS is the aim, but the imitation of the key properties of the AS of matter that may be useful in the group of UVs control. Let us consider these key properties.

At the "solid" AS a group of UVs has a constant structure. In this case, it is convenient to use the "crystal" structure. All UVs rigidly positioned at predetermined nodes in relation to each other. Each UV can slightly oscillate a relatively a predetermined position, by analogy with the motion of atoms in the crystal lattice. The structure of the "crystal lattice" and the degree of permissible oscillations is set at the high level of control. The control system of each UV executes its own set position.

There is a second possible method of constructing a model of "solid" AS, by using a unique non-recursive scheme, which is defined in an array of nodes, in which UVs should be situated. This method assumes that there is certain finite location scheme of UVs relative to each other.

Group of UVs in "liquid" AS has the following key features:

- Maintaining a constant density, i.e. the concentration of UVs in space. The consequence of this property is a volume, constantly occupied by UV group.
- Tendency of a group of UVs to occupy structure with a minimum square of the shell, which covers a group of UVs. This requirement reproduces such property of the liquid as a surface tension.

Group of UVs as "fluid" moves compactly, while UVs don’t have a fixed position in the group and are interchanged freely. Group of UVs avoids obstacles by flowing around or seeping. Minimum distance between UVs is one of limiting conditions for the "liquid" AS. It is needed to avoid collisions of UVs.

In the "gas" AS, group of UVs has such key property, as a tendency to occupy maximal domain evenly. It is assumed that at the high level of control the domain configuration is set, i.e. the shell, which should occupy the "gas". If the shell is not limit a group of UVs in the "gas" AS, then the whole group occupies the maximum possible domain, at which it remains controllable. Visibility range of UVs each other is one of the limiting conditions on the maximum domain. UVs cannot dispose to a greater distance.

IV. The Relative Motion of Unmanned Vehicles in the Group

Let's consider the relative movement of vehicles under the influence of low-level algorithms. Group of UVs imitates the movement of molecules of the substance under the influence of internal forces. The resultant of these forces will be required thrust for each UV. At the same time, UV calculates forces, which would affect on it if it were a particle of matter. UV finds the resultant force and implements its action by its own traction. Values of the acting forces are computed in accordance with a predetermined law of motion.

In each of the described AS', every UV mimics forces, acting on it from the rest of UVs, which can be divided into the forces of repulsion and attraction. Repulsive forces are necessary to prevent the collision of UVs. Forces of attraction are necessary to prevent loss of communication between the UVs.

We can set the interaction between UVs ourselves, thereby changing the properties of the UV group. Let us imagine UVs as material points.

Compare the effect of various laws of motion. They can be divided into two classes: linear (such as Hooke's law) and nonlinear (such as Coulomb's law, the law of the hyperbolic, quadratic laws and others).

The basic advantage of non-linear laws, in contrast to the linear: with decreasing distance between vehicles, force acting between them increases rapidly up to infinity. Thereby we ensure that the UVs never collide. But this is only possible in an ideal system. In fact, we cannot attach to the UV endless traction, and fulfilment of prescribed program is not instantaneous, this imposes restrictions on the initial conditions of the system.

V. Conclusions

In this article, we propose an approach to the organization of the control of large groups of unmanned vehicles. It is based on an imitation the aggregation states of matter by a group of unmanned vehicles and its decentralized control. It allows you managing a large group of UVs using the intuitive high-level commands. Thus, the algorithms for conduct and control of each vehicle are quite autonomous and simple. Control is carried out in accordance of the information about the relative position of the nearest vehicles based on the given law of interactions. The choice of the laws of UVs interactions is in the designer hands. Currently, the authors investigated several types of interaction of unmanned vehicles, including Hooke's law, the law of Coulomb, hyperbolic law and other. During the carrying out the work the software module was developed that allows you to simulate the control of vehicles group in three aggregation states of matter. In addition, the sustainability of the group as a system of material points with different laws of their interaction has been investigated.

Acknowledgment

Review has been made at the expense of subsidies under support of Kazan (Volga Region) Federal University.

In Kazan National Research Technical University named after A N Tupolev - KAI in the department of Processes and Control Dynamics, researches related to the problem of optimization and stability of systems with distributed parameters, mathematical modeling and management of complex, transforming systems, are carried out for many years. [3-8, 17]. This paper is a continuation of these studies, and is associated with mathematical modeling of such a complex spatially distributed transforming system, like a large UV group.

References

[1] Tarasov V B 2002 From multi-agent systems to intellectual organizations (Moscow: Editorial URSS) p 352.

[2] Kalyaev I A, Kapustyan S G and Gayduk A R 2010 Managing large systems. Special Edition "Network models in control" Self-organizing distributed systems of group management of intelligent robots that are based on the network model30.1 (Moscow: Institute of Control Sciences of RAS) pp 605-639

[3] Sirazetdinov T K 1977 Optimization of systems with distributed parameters(Moscow: Akadem izdat center "Nauka" RAS) p 480

[4] Sirazetdinov T K 1987 Sustainability of systems with distributed parameters. (Novosibirsk: Akadem izdat center "Nauka" RAS) p 232

[5] Degtyarev G L,Malikov A I, Sirazetdinov T K, Sirazetdinov R T 2007 Vestnik KGTU im. A N Tupoleva Studies on the sustainability and control in KNRTU. named after A N Tupolev 3 pp 90-97

[6] Sirazetdinov R T 1994 Izv. VUZ. Aviatsionnaya Tekhnika Mathematical modeling of the system development similar facilities with regard to the intensity them use (The example of airplane-helicopter fleet) 1 (Kazan) pp 63-68

[7] Sirazetdinov R T 1998 Izvestia Rossiiskoi Akademii Nauk. Teoriya i Systemy Upravleniya Mathematical modeling of infrastructure capacity of complex systems 3 pp96-104

[8] Mironov A B 2013"XXII Tupolev reading (School for Young Scientists)": International Youth Scientific Conference Control of large groups of unmanned vehicles on the basis of imitation physical state of matter Scientific adviser Sirazetdinov R T (Kazan: Kazan A N Tupolev State Technical University Press) pp 341-434

[9] Duarte, M., Oliveira, S. M., and Christensen, A. L. (2014b). Evolution of hybrid robotic controllers for complex tasks. Journal of Intelligent and Robotic Systems. In press.

[10] Duarte, M., Oliveira, S. M., and Christensen, A. L. (2014b). Hybrid control for large swarm of aquatic drones. In 14th International Conference on the Synthesis & Simulation of Living Systems (ALIFE), pages 785–792. MIT Press, Cambridge, MA.

[11] Rodrigues, T., Duarte, M., Oliveira, S. M., and Christensen, A. L. (2015). Beyond onboard sensors in robotic swarms: Local collective sensing through situated communication. In 7th International Conference on Agents and Artificial Intelligence (ICAART). SciTePress, Lisbon, Portugal. In press.

[12] Duarte, M., Oliveira, S., and Christensen, A. L. (2011). Towards artificial evolution of complex behavior observed in insect colonies. In 15th Portuguese Conference on Artificial Intelligence (EPIA), pages 153–167. Springer, Berlin, Germany.

[13] Schmickl, T. and Crailsheim, K. (2008). Trophallaxis within a robotic swarm: bio-inspired communication among robots in a swarm. Autonomous Robots, 25(1-2):171–188.

[14] Bassler, B. L. (1999). How bacteria talk to each other: regulation of gene expression by quorum sensing. Current Opinion in Microbiology, 2(6):582–587.

[15] Stamatis, P. N., Zaharakis, I. D., and Kameas, A. D. (2009). A study of bio-inspired communication scheme in swarm robotics. In 7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS), pages 383–391. Springer, Berlin, Germany.

[16] Pospelov D.A. Situational control: the theory and practice.– M.: Science, 1986.

[17] Tikhonov S V and Sirazetdinov R T 2015 VIII International Scientific Conference "Problems and prospects of aviation, land transport and energy" (ANTE 2015)Research of unmanned vehicles large groups motion dynamics on the imitation basis of substance aggregate state (Kazan: Brig) pp 419-423

[18] Dias M.B., Stentz A. A Free Market Architecture for Distributed Control of a Multirobot System. Proceedings of the 6th International Conference on Intelligent Autonomous Systems (IAS), Venice, Italy, July, 2000.

Research on the Structural Characteristic of Global Terrorist Organization Cooperation Network

Zihan Lin, Duoyong Sun, Min Tang
College of Information System and Management,
National University of Defense Technology, Changsha, China
godlove.lin@163.com

***Abstract*—Terrorist organizations are not completely independent, however, they frequently interact with each other. Emerging researches have studied the consequence and influential factors of these relationships. But there have rarely been given precise empirical analysis on its interactive strength and effectiveness. Social Network Analysis (SNA) is considered as an effective method to quantify and analyze relationship between different social actors. In this paper, our focus is the cooperation relations between terrorist organizations. Based on the terrorist incident dataset from 1970 to 2014, this paper establishes the cooperation network of terrorist organizations and studies its structural characteristic. The results have shown that both the states of cohesion and separated can be found in the cooperation network. The relationship between organizations is more closely in the central part of the network, and the organizations within the same subgroups have obvious similarities in political goals and activity regions. Additionally, some important organizations in the operational cooperation have been identified. The removing of these organizations will have a significant impact on the cooperation network structure.**

Keywords—Terrorist Organization; Operational Cooperation; Social Network Analysis; Structural Characteristic; Cohesive Subgroup

I. INTRODUCTION

After the end of the Cold War, terrorism has become an increasingly important issue among decision makers and security experts. Terrorist organizations are groups use terrorism in order to obtain a political goal through intimidation of a wider audience beyond their immediate victims [1]. These organizations are not completely independent, however, they frequently interact with each other. Some terrorist organizations share with each other and jointly conduct attacks, while other organizations are in a state of opposition and even fight with each other. The cooperative ties between terrorist organizations are quite essential for its survival, as these connections can help the organizations meet their needs relating to intelligence, personnel, training and weapons by sharing resources and increase their attraction in recruitment [2]. To understanding the interaction of terrorist organizations is quite helpful for the security department to monitor and predict the covert activities as well as develop strategy for the countering international terrorism.

Emerging researches have studied the consequence of these relations and showed that the inter-organizational ties have important impact on the terrorist organization [3-4]. Some scholars theorized the formation of relations, and use case study to explain why some terrorist organizations cooperation occurred in a specific situation [5]. With the increasing attention to relationships between actors in the study of political violence, social network analysis also has been used to study the relations between states, groups or individuals. Scholars increasingly use the social network framework to understand terrorism, especially the structural characteristics of covert networks [6]. These works have studied the terrorist organization from different perspectives and provided useful results in how the terrorists act and its influence. However, there are two limitations in current reported studies: The first limitation is the shortage of quantitative research on the relations between terrorist organizations. Secondly, most of the studies using social network approach have focus on the examining of relations between individual terrorists, not groups.

In this paper, we address these shortcoming in the literature by focus the cooperation relations between terrorist organization. A framework is proposed for the quantitative analysis of the terrorist organization cooperation network. Based on the terrorist incident dataset from 1970 to 2014, we construct the cooperation network model of terrorist organization, and analyze its network-level properties, cohesive subgroup and important organizations. The rest of the paper is organized as follows. Section 2 describe the design of the research, which include the method for model construction and structural characteristic analysis. Section 3 present the quantitative empirical analysis for the cooperation network of terrorist organizations. Section 4 concludes the paper.

II. RESEARCH DESIGN

Fig.1 demonstrate the research framework for the cooperative relations, which consists of the following three components: (1) The data collection and preprocess, including the data extraction based on terrorist incident dataset. (2) The construction of cooperation network model, involving the definition and visualization of the cooperative interaction between terrorist organizations. (3) The characteristic analysis of the cooperation network, including the network analysis approach of the cooperation network in structural properties and nodes position, also providing the statistical results and discussion of the cooperative behavior of terrorist organizations.

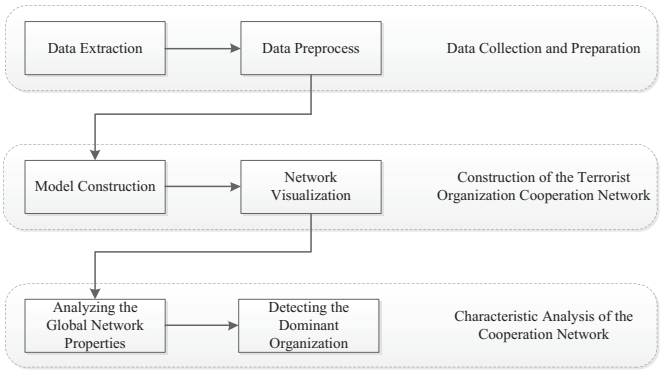


Fig. 1. The Research Framework for Cooperative Relations

A. Construction of the Cooperation Network Model

Cooperative connections between terrorist organizations involve substantive logistical coordination and operational coordination [7]. In this paper, we attempt to address the operational cooperative behavior of terrorist organizations, which means that two groups jointly conduct an terrorist attack. Each terrorist organization is taken as a node in the cooperation network. We have reason to regard that there is a definite relations between two terrorist organizations if it is recorded in the database that they both claimed to responsibility for a certain terrorist attack. Thus, we use an edge between the two nodes to presents the operational relations between the two terrorist organizations.

In the cooperation network model, the set of terrorist organization is denoted by set $A=\{A_1, A_2, \cdots\cdots, A_n\}$, the relation on A is defined mathematically as a subset of the $n \times n$, and the cooperation relation between each organization is represented by the $n \times n$ adjacency matrix $AA=(a_{ij})$. The a_{ij} represents the number of terrorism attacks that have jointly committed by terrorist organization i and j ($i, j=1,...,n$).

B. Structural Characteristic Analysis

Plenty of measures and method are used to calculate and map the structural properties of the network in SNA [8]. The measurements we used in the framework can be classified as network level and node level.

Measurements in network-level can help us better understand the global structure and characteristics of its activities of operational cooperation network, which including the density, clustering coefficient, average distance, centralization and so on. Network density is used to describe the levels of intimacy relationship among terrorist organization (1). Clustering Coefficient measures the degree of clustering in a network by averaging the clustering coefficient of each node, which is defined as the density of the node's ego network. Here we use the definition of Barrat [9] to calculate it (2). Network Centralization is based on the centrality of each node in a square network. It makes greater emphasis on the situation of the whole network (3) (4) (5).

$$D(A)=\frac{2l}{n(n-1)} \tag{1}$$

$$CC(A)=\frac{1}{s_i(k_i-1)}\sum_{j,k}\frac{w_{ij}+w_{jk}}{2}a_{ij}a_{jk}a_{ki} \tag{2}$$

$$C_D(A)=\frac{\sum_{i=1}^{n}[C_D(A^*)-C_D(A_i)]}{[(n-1)(n-2)]} \tag{3}$$

$$C_B(A)=\frac{2\sum_{i=1}^{n}[C_B(A^*)-C_B(A_i)]}{[(n-1)^2(n-2)]} \tag{4}$$

$$C_C(A)=\frac{\sum_{i=1}^{n}[C_C(A^*)-C_C(A_i)]}{[(n-1)(n-2)]/(2n-3)} \tag{5}$$

Many researches have been done in the detecting of key nodes. Scholars have used centrality analysis combined with attribute measures to evaluate the key person in the terrorist organization [10]. Centrality Analysis is often used to describe the social status of a node in the network, which is a kind of quantitative research of prestige and power [11]. The organization occupy cenral position in the cooperation network is believed to be more influential in terms of the flow of resource and the spread of behaivors. Degree Centrality is used to describe the importance of a node in the network (6). Organizations with high degree centrlity is more active in the operational cooperation. The Betweenness Centrality is used to measure the connections of one given organization with others in the cooperation network (7). Organiations with high betweenness could be thought of as a gatekeeper of information flow. The Closeness Centrality reveals the closeness of an organization to the other organizations in the network (8). High scoring organization in closeness have the shortest paths to all others in the cooepration network, which has advantage in the transfering of information.

$$C_D(A_i)=\frac{\sum_{j=1}^{n}l(A_i,A_j)}{n-1} \tag{6}$$

$$C_B(A_i)=\frac{2\sum_{j<k}g_{jk}(A_i)/g_{jk}}{(n-1)(n-2)} \tag{7}$$

$$C_C(A_i)=\frac{n-1}{\sum_{j=1}^{n}d(A_i,A_j)} \tag{8}$$

Clustering analysis is used to identify the cohesive subgroups of actors within a network [12]. Cohesive subgroups are subsets of actors among whom there are relatively strong, intense and frequent ties. The concept of cohesion subgroup includes the N-Cliques, N-Clan, K-Plex, K-Core, and Lambda Set. Many grouping method are used in social network analysis in order to explain how many cohesive subgroups exist and how the relations between terrorist organizations within the same groups.

III. Emprical Analysis of the Cooperation Network

The data used to describe the terrorist organization cooperation is from the Global Terrorism Database (GTD) terrorist incident dataset [13]. The GTD recorded the brief information on publicly reported terrorist events around the world from 1970 to 2014, including the incident date and location, attack type, target and perpetrators, weapon

information, casualties and so on. The results of data preprocess show that there were 141,966 terrorist attacks by 2014. Terrorist organizations claimed to be responsible for 76,850 attacks (about 54.13% of the total). 828 attacks among them were jointly conducted by more than one terrorist organization.

Based on the definition of operational cooperation, we have built the global cooperation network of terrorist organizations from 1970 to 2014. Here we use ORA [14] to visualize and explore the structural characteristic of the cooperation network. The analysis is from two levels: In the network level, we focus on the global structure and cohesive subgroup of cooperation network. In the node level, centrality analysis is used to explore the important terrorist organizations.

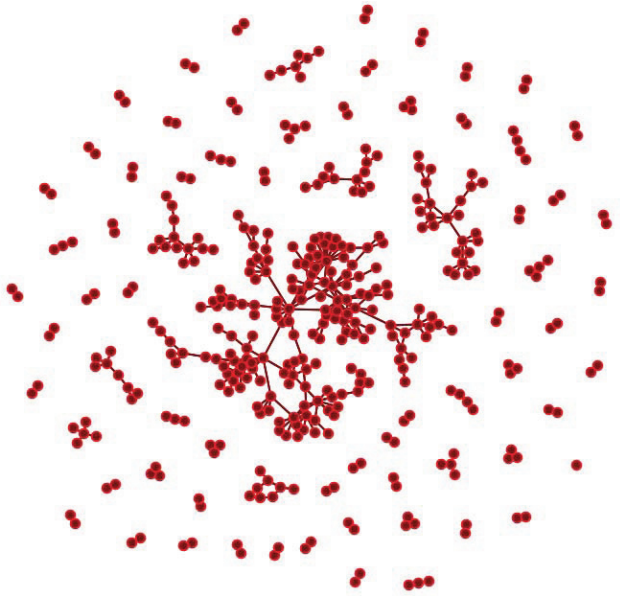


Fig. 2. The Cooperation Network by 2014

Fig. 2 shows a snapshot of the cooperation relations by 2014. Both the states of cohesion and separated can be found in the cooperation network. The relationship between organizations is more closely in the central part of the network, while many isolated node pairs can be found on the edge of the network.

Table 1 demonstrates the statistical results of the network-level properties. We found that 403 terrorist organizations have formed 443 cooperative relations with other organizations. Compared with other social networks, the cooperation network of terrorist organization is loosely connected, and more tend to hierarchical model. The strength of the nodes are uneven distributed, heterogeneity in the strength of nodes is shown in Fig. 3. The strength of the majority nodes is very small, while very fewer nodes with greater strength. The local efficiency of the cooperation network (0.2168) is much larger than the global efficiency (0.0286), as there exist many cohesive subgroups in the network.

TABLE I. THE NETWORK-LEVEL PROPERTIES OF COOPERATION NETWORK

Measure	Value	Measure	Value
Node Count	403	Link Count	443
Average Distance	8.987	Average Speed	0.1113
Degree Centralization	0.0061	Betweenness Centralization	0.1116
Density	0.0054	Transitivity	0.228
Density (weighted)	0.0005	Clustering Coefficient	0.1134
Diffusion	0.1798	Connectedness	0.1817
Efficiency (Global)	0.0286	Efficiency (Local)	0.2168

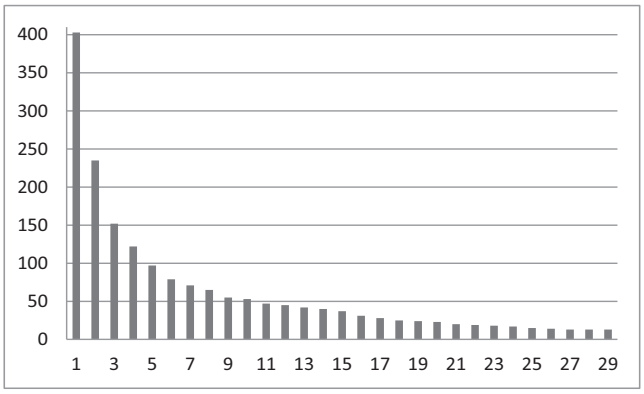


Fig. 3. The Cumulative Distribution of Node Strength

In this paper, Newman clustering algorithm was used to analyze the cohesive subgroups. 83 cohesive subgroups have been identified in the cooperation network (Newman modularity: 0.889). Table II shows the top 10 cohesive subgroups in the cooperation network, including both the structural and organizational properties of these groups. It is founded that measures for the extent to which ties are concentrated within a subgroup, rather than between subgroups. This is because the forming of relations are quite related to the organizational properties. There exist some preferences in the selection of cooperation partners for the terrorist organizations, which is confirmed that terrorist organizations within the same subgroup have obvious similarities in political goals and activity regions. These organizations tend to form cooperative ties with ones in similar properties. Take the largest subgroup for example, 41 terrorist organizations can be found in this subgroup, most of the them have goals related to religious and located in Afghanistan. Additionally, these subgroups are not isolated, they connected to each other and have formed their own union.

TABLE II. TOP 10 COHESIVE SUBGROUPS IN THE COOPERATION NETWORK

R	Structural Properties			Organizational Properties	
	Size	Density	Clustering Coefficient	Ideology	Bases of Operation
1	41	0.003	0.369	Religious	Afghanistan
2	32	0.004	0.267	Nationalist	Israel, Syria, Lebanon
3	26	0.017	0.453	Anarchist	Africa
4	24	0.013	0.419	Nationalist	India, Kashmir, Pakistan
5	21	0.009	0.383	Nationalist	Israel, Palestine, West Bank/Gaza
6	15	0.023	0.169	Religious	Indonesia, Malaysia, Thailand
7	13	0.09	0.254	Environmental	Canada, United Kingdom, United States
8	12	0.011	0	Nationalist	Ethiopia, Somalia
9	11	0.01	0.309	Communist	Colombia
10	10	0.042	0.54	Religious	Indonesia, Malaysia, Philippines

In the node-level, we have calculated the centrality of organizations. Degree centrality is used to identify the dominant organizations in the operational cooperation network (shown in Table III). It is founded that Lashkar-e-Taiba (LeT), Al-Aqsa Martyrs Brigade and Tehrik-i-Taliban Pakistan (TTP) are more likely to conduct attacks with others. The LeT is the armed wing of the Pakistan-based religious organization, and one of the three largest and best-trained groups fighting in Kashmir against India. LeT attacks have increased tensions in the already contentious relationship between India and Pakistan.

TABLE III. TOP 10 ORGANIZATIONS IN DEGREE CENTRALITY

Rank	Terrorist Organization	Value
1	Lashkar-e-Taiba (LeT)	0.0066
2	Al-Aqsa Martyrs Brigade	0.006
3	Tehrik-i-Taliban Pakistan (TTP)	0.0057
4	Taliban	0.0056
5	Revolutionary Armed Forces of Colombia (FARC)	0.0052
6	Palestinian Islamic Jihad (PIJ)	0.0052
7	Hamas (Islamic Resistance Movement)	0.0051
8	Islamic State of Iraq and the Levant (ISIL)	0.0049
9	Al-Qa`ida	0.0042
10	Al-Nusrah Front	0.0037

Betweenness centrality is used to identify the bridging organizations. (shown in Table IV). Organizations with high betweenness centrality play a significant role in the coordination of various resources including the intelligence, weapon and personnel. Al-Qa`ida, Hamas and Lashkar-e-Jhangvi are identified as the bridge organizations in the cooperation network. As the bases of operation is distributed in more than 50 countries and regions around the world, Al-Qa`ida has played an important role in connecting the terrorist organizations in different countries and also the different cohesive subgroups.

TABLE IV. TOP 10 ORGANIZATIONS IN BETWEENNESS CENTRALITY

Rank	Terrorist Organization	Value
1	Al-Qa`ida	0.1139
2	Hamas (Islamic Resistance Movement)	0.0834
3	Lashkar-e-Jhangvi	0.076
4	Harkatul Jihad-e-Islami	0.0584
5	Islamic Jihad (Ideological Grouping)	0.045
6	Hizballah	0.0363
7	Al-Aqsa Martyrs Brigade	0.0322
8	United Liberation Front of Assam (ULFA)	0.0267
9	Islamic State of Iraq and the Levant (ISIL)	0.0263
10	Tehrik-i-Taliban Pakistan (TTP)	0.0254

IV. CONCLUSION

The understanding of interactive mechanism between terrorist organizations is quite important for the government to make reasonable policy and carry out effective cooperation programs in the countering of terrorism. Based on the social network analysis method, this paper presented a framework for the quantitative analysis of the cooperation relationship between different terrorist organizations. We constructed an operational cooperation network model based on the GTD dataset, characteristics of the global terrorist organization cooperation network have been analyzed from both the network-level and node-level. The results have shown that both the states of cohesion and separated can be found in the cooperation network. The relationship between organizations is more closely in the central part of the network, and the organizations within the same subgroups have obvious similarities in political goals and activity regions. Additionally, some important organizations in the operational cooperation have been identified, such as Lashkar-e-Taiba (LeT), Al-Aqsa Martyrs Brigade and Tehrik-i-Taliban Pakistan (TTP). The removing of these organizations will have a significant impact on the cooperation network structure.

The results also raise a number of questions about counter-terrorism policy, which in turn create opportunities for further analysis. The worldwide terrorist activities would be contained to a certain extent if we could destroy the dominant organizations or cut off the important relations. With this in mind, governments should generally be concerned about the cooperation trend between terrorist organization and develop specific countering strategy.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No.71473263 and the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No.20134307110020.

REFERENCES

[1] Enders, Walter and Todd Sandler, "The Political Economy of Terrorism", Cambridge University Press, 2006.

[2] Hoffman, Bruce, "Inside Terrorism", Columbia University Press, 2006.

[3] Brian J. Phillips, "Terrorist Group Cooperation and Longevity", International Studies Quarterly, vol 58，no. 2, pp. 336-347, 2014.

[4] Horowitz, Michael C., and Philip B.K. Potter, "Alling to Kill: Terrorist Intergroup Cooperation and the Consequences for Lethality", Journal of Conflict Resolution, vol 58, no. 2, pp. 199-225 , 2013.

[5] Schiller, David Th. "A Battelgroup Divided: The Palestinian Fedayeen." In Inside Terrorist Organizations, Frank Cass, 2011.

[6] Sageman, Marc. "Understanding Terror Networks", Pennsylvania University Press, 2004.

[7] Karmon, Ely, "Coalitions Between Terrorist Organizations: Revolutionaries, Nationalists and Islamists", Nijhoff Publishers, 2005.

[8] Brass, D.J. "A Social Network Perspective on Industrial/ Organizational Psychology". Industrial/ Organizational Handbook, 2009.

[9] Barrat A, Barthelemy M, Pastor-Satorras R, et al. "The architecture of complex weighted networks". Pnas, 2004, vol. 11, pp. 3747-3752.

[10] D.M.Akbar Hussain. "Terrorist Networks Analysis through Argument Driven Hypotheses Model", Proceedings of the Second International Conference on Availability, Reliability and Security, 2007.

[11] Newman M E J. "Detecting community structure in networks", European Physics Journal B.2004, vol. 38, pp. 321-330.

[12] S. Wasserman, and K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, New York and Cambridge, 1994.

[13] Global Terrorism Database, http://www.start.umd.edu/gtd. 2016-6-15.

[14] Carley K M, Dombroski M, Tsvetovat M, et al. "Destabilizing dynamic covert networks". Proceedings of the 8th international Command and Control Research and Technology Symposium, 2003.

Multilanguage natural user interface to database

Ruslan Posevkin
Chair of Computation Technologies
ITMO University
Saint Petersburg, Russia
rus_posevkin@mail.ru

Igor Bessmertny
Chair of Computation Technologies
ITMO University
Saint Petersburg, Russia
bia@cs.ifmo.ru

***Abstract*—Nowadays most of people of developed economy countries interact with software every day. As a result of computer systems expansion to all scope of people's activity the problem of transition from visual and command interfaces to natural language user interfaces is thrown into the sharp relief. Computational linguistics and natural language processing methods are described in this article. Represent a formal definition of a dialogue system as a Markov decision process. Methods of natural language manipulation are applied in machine translation software systems, systems of search and exchange the data, text annotation and expert systems. Prototype of natural language user interface to structured data source is developed. As a result it is convert natural language user's query to SQL query to database. Also showed application of Levenshtein distance for processing natural language query to SQL query to database. Natural language user interface is created to predefined subject field. User interface interacts to database that contain information about existent program libraries and frameworks. Consequently, using natural language processing methods it is possible to develop user natural language user interface providing capability to interact with machine.**

***Index Terms*—Natural language processing, database, dialogue system, user interface, Markov decision process, Levenshtein distance.**

I. Introduction

There are many computer software systems in the world. Each of them has a different principle of interaction. As a result it leads to some problems of their free using.

Using natural language to communicate with computer software systems is a way to solve the problem. Natural language user interface may help people to interact with computer systems. And this method has such advantages as

- Minimal user preparation for working with computer system
- Simple and fast way to put an arbitrary question to software system.

Simplicity of natural language interface interaction is achieved by everyday using natural language.

Limited set of vocabulary and grammar units can be used for natural language user interface of software systems. It doesn't lead to deep functionality degrading of question-answering system.

Limited language is a subset of natural language. A native speaker easily understands the text in a limited language.

There is no need to lengthy learning skills of limited language text composition because it has a limited set of vocabulary and grammar.

It allows avoiding linguistic uncertainty and decreases natural language elements processing time [1].

II. Workflow

Natural language user interface is considered to be a specific intelligence system providing a dialogue between the user and the software system within a definite subject field.

Intelligence system of natural language user interface consists of:

- User interface whereby the user inputs the message and gets an answer from the system.
- Translator of natural language requests to internal language of queries.
- Translator of internal language units to natural language.

The operation cycle of natural language user interfaces starts with the message input in natural language by entering the text. The next step is creating a formal description of the text. All previous analysis results are used to analyse the following queries. It makes it possible to resolve the issues connected with using the same terms in different subject fields.

All components of natural language user interface knowledge machine may be classified into translators and analyzers. Translators translate knowledge from one language to another. For example, they translate the descriptive knowledge of some subject field to a natural language text.

The analyzers analyze knowledge units and develop previously unknown facts. For example, it can be the analysis of a user's question to find an answer.

III. Markov decision process

Translation system is the part of dialogue system. This section represents dialogue system as a Markov decision process. It will be illustrated by the example of dialogue where the goal of the system is to get the list of libraries to user through the shortest possible interaction. In this example user should make queries using values of "programming language" and "license type" as a filter.

Markov decision process is described in terms of a state space, an action set and a strategy. Dialogue system's state represents all resources it interacts with. For our example, the

state includes two entries: libraries list filter by "programming language" and libraries list filter by "license type". The total number of states including one initial state, different combinations of libraries list depends on input values and filter, and a special final empty state.

The action set of the dialogue system includes all possible actions it can perform, such as interactions with the user (e.g. asking the user for input, providing a user some output, confirmations, etc.), interactions with other external resources (e.g. querying a database), and internal processing.

For our example, the action set include such actions as:

1) A question where user asking libraries list for specific programming language.

2) A question where user asking libraries list with specific license type.

3) Step-by-step questions where user asking libraries list with specific programming language and license type or vice versa.

4) A final action, closing the dialogue.

When an action a is taken at state s, the system's state changes to be s'. The state transitions are modeled by transition probabilities as in Eq. 1.

$$P_T(s(t+1) = s' \mid s(t) = s, a(t) = a) \quad (1)$$

Path from initial to final state is a dialogue session. A dialogue strategy shows what is the next action will be invoked for each reached state.

It is possible to measure the system performance by an objective function C, where the costs C_i measure the distance to the achievement of the application goal, and the efficiency of the interactions as in Eq. 2.

$$C = \sum C_i \quad (2)$$

In general, the costs in Markov decision process are described by the conditional distributions as in Eq. 3.

$$P_C(c(t) = c \mid s(t) = s, a(t) = a) \quad (3)$$

Quadruple of state space, action set, transition probabilities, and cost distributions defines a Markov decision process.

For our example the objective function includes three terms as:

$$C = W_i \times \langle \#\ interactions \rangle + W_e \times \langle \#\ errors \rangle + W_f \times \langle \#\ incomplete\ values \rangle$$

The first term is the expected duration of the dialogue; the second corresponds to the expected number of errors in the obtained values; and the third measures the expected distance from achieving our application.

In order to reflect this objective function in our dialogue model, we associate a cost c to the taken action a in a state s. The cost incurred with any of the first three actions in dialogue system is $W_i + W_e \times number\ of\ errors$. If we assume that the concept error rate for recognition of filter values separately (for questions 1 and 2) is p_1 , and together (for question 3) is p_2, $p_2 > p_1$, then the expected cost accumulated when actions 1 or 2 are taken is $W_i + W_e \times p1$, while for question 3 is $W_i + 2 \times W_e \times p_2$. For action 4 (closing the dialogue) the cost depends deterministically on the state in which this action is taken and is $W_i + 2 \times W_f$ for an initial state, $W_i + W_f$ for states in which one of the filter values isn't used, and W_i for the states in which both filter values are used.

Fig. 1 shows three different strategies and their costs for the example system. Optimal strategy minimizes the objective function. In this case $C_1 = W_i + 2 \times W_f$. For example, in strategy 1 (where the system close the dialogue as the first action) is optimal when the recognition error rate is too high: $p_1 > (W_f - W_i) / W_e$.

In strategy 2, a user asking libraries list for one of the disposable filter values, and closes the session. In this case $C_2 = 2 \times W_i + 2 \times p_2 \times W_e$.

In strategy 3, a user asking library list with specific programming language and license type, and then closes the session. In this case $C_3 = 3 \times W_i + 2 \times p_1 \times W_e$. Strategy 3 is optimal when the difference in error rates justifies a longer interaction: $p_2 - p_1 > W_i / 2 \times W_e$.

There exist in literature several techniques for computing the optimal strategy given the correct model parameters (the transition probabilities and the cost distributions), including value iteration, policy iteration, etc. These techniques are based on dynamic programming that can be used due to the Markovian nature of this model. They rely on the following definition:

Optimal value $V(s)$ of a state s is the lowest expected cost incurred after the system left state s and until it reached the final state. The optimal value function is unique and can be defined as the solution to the simultaneous equations as in Eq. 4.

$$V(s) = \min_a \left(\langle C(s,a) \rangle + \sum_{s'} P_T(s' \mid s,a)\, V(s') \right), \quad (4)$$

where $\langle C(s,a) \rangle$ is the expected cost for action a in state s [2].

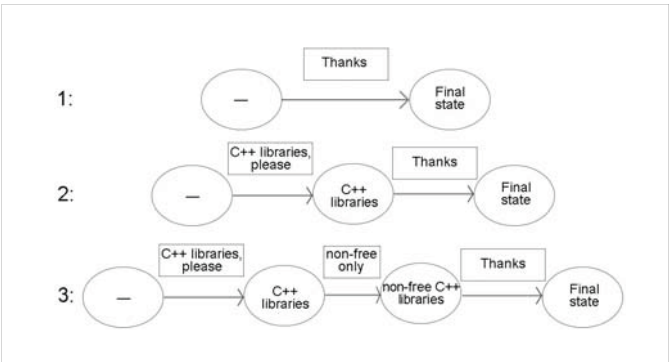


Fig 1. Different strategies of Markov decision process.

IV. Natural language processing

Natural language processing of a text consists of three phases:

- Morphological analysis.
- Syntactic analysis.
- Semantic analysis.

The first step is morphological analysis. For each allolog relations for grammatical categories like gender, case, declension and so on are formally extracted from the sentence.

The next step is syntactic analysis. Syntactic relations between words are created in the sentence, main and subordinate parts of sentence are extracted, sentence type is specified, and so on.

Syntactic analysis is executed phase-by-phase using the data obtained at the previous step. This phase uses lexical and syntactic rules for analyzing the language.

Semantic analysis phase is the most difficult phase of natural language processing. Semantic analysis is based on knowledge machine for a specific subject field and the information received within previous phases. At this stage a linguistic construction is compared with the construction stored in system memory.

Creating word semantic model is referred to as polysemanticism. Word sense is a set of possible values, each of them may be implemented in a specific subject field.

V. Translation natural language query to SQL

The first step is translation of natural language query to English language. English language is selected because column names of database scheme frequently named using exactly English language. Also in case of need translation module may be changed to another language translation module.

Then from the user's query key entities is extracted. In most cases key entities are nouns. The next step is for each pair *{key entity; table column name}* the Levenshtein distance is calculated. As a result for pairs *{key entity; table column name}* with the least value of the Levenshtein distance defined key column names. Then these key column names includes as a part of SQL query to database. Figure 2 shows the general scheme of processing natural language user query to SQL query to database.

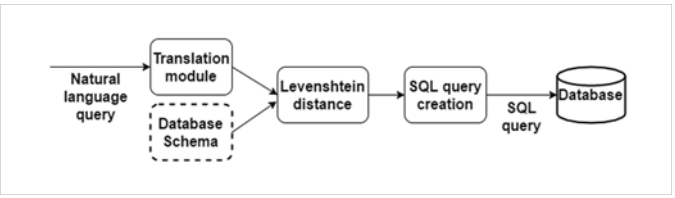


Fig 2. Processing natural language user query to SQL query to database.

Prototype of natural language user interface to database was developed in this research. Source data is relational database MySQL that contains information about existent program libraries and frameworks.

Output data of natural language user interface is SQL query to database. One of the necessary conditions is occurrence of data structure describing the database. In particular, information about table's and their fields names should be occurred.

Complex natural language sentence in user's query to database didn't used in natural language user interface. Though, it needs to create linguistic analyser as converter that consists of two text abstraction levels — morphological and syntactic [3]. Each text abstraction level should have a model component with array of rules and libraries and have custom query image such as morphological and syntactic structures.

There is no need to use the semantic analysis phase because subject field are known beforehand and linguistic analyser should be therefore named as linguistic translator.

Acceptable quality level of natural language query processing achieved without full grammatical sentence analysis. It is enough to extract the most informative parts of texts such as keywords, phrases and fragments. Next step is creation of morphological and syntactic models using morphological and syntactic analysis methods [4].

Source data is user's query to relational database that has such limits as:

- Query consists of one or more simple sentences in Russian language.
- If query contains multiple simple sentences then each next sentence should contain noun from one of previous sentences.
- Query should request information relevant to database.
- Query should has clear nature: shouldn't use demonstrative and personal pronouns ("I, you, he/she/it, they, this, that") and particles ("whether, or, which").
- It is recommended start query with special words such as "get", "show", "enumerate", "when", "which time", "what", "which" [5].

References

[1] I. Bessmertnyi, "On constructing intellectual systems in ternary logic", Programming and Computer Software, vol. 40, no. 1, pp. 43-46, 2014.

[2] Levin E., Pieraccini R., Eckert W. "Learning dialogue strategies within the Markov decision process framework", Automatic Speech Recognition and Understanding IEEE Proceedings, pp. 72-79, 1997.

[3] O. Eliseeva, "Natural language interface of intelligence systems", Minsk: BGUIR, 2009.

[4] Damljanovic, Danica, Milan Agatonovic, and Hamish Cunningham. "FREyA: An interactive way of querying Linked Data using natural language", The Semantic Web: ESWC 2011 Workshops. Springer Berlin Heidelberg, 2012.

[5] Posevkin, R., Bessmertny, I. Translation of natural language queries to structured data sources. 9th International Conference on Application of Information and Communication Technologies, AICT 2015 IEEE Proceedings, pp. 57-59, 2015.

Use of Integrated Technologies for Fire Monitoring and First Alert

A. D'Ambrogio, P. Gaudio, M. Gelfusa, M. Luglio, A. Malizia, C. Roseti, F. Zampognaro
Università degli studi di Roma Tor Vergata
Rome, Italy
dambro@uniroma2.it; gaudio@uniroma2.it; gelfusa@ing.uniroma2.it, luglio@uniroma2.it; malizia@ing.uniroma2.it; roseti@ing.uniroma2.it; zampognaro@ing.uniroma2.it

A. Giglio, A. Pieroni
Università degli studi Guglielmo Marconi
Rome, Italy
a.giglio@unimarconi.it; a.pieroni@unimarconi.it

S. Marsella
Corpo Nazionale dei Vigili del Fuoco
Rome, Italy
stefano.marsella@vigilfuoco.it

***Abstract*— The paper describes the project "Integrated technologies for fire monitoring and first alert" proposed by the authors to satisfy technical requirements of the Firefighter Bodies with particular attention to the monitoring of wood fires and release of chemical materials. The target system is composed of three sub-systems: sensors to detect target parameters, telecommunications to transfer sensor data in real time and for first alert management, fire evolution real time simulator to support decisions. The project aims also to integrate the three subsystems facing and solving technical/scientific issues to provide a significant applicative high value.**

***Keywords*—Sensors, LIDAR, hybrid networks, simulation, fire monitoring, alert, real time, support to decision.**

I. Introduction

In Italy many areas are particularly sensitive to fire and uncontrolled gas releases, such as woods close to residential blocks and industrial areas. The main issue identified by the Firefighter Body is the lack of an efficient early monitor and detect system allowing a prompt and efficient remediation. In this framework, the project aims to the design, implementation and test of an efficient monitoring and control system jointly relying on innovative sensors (LIDAR), on hybrid telecommunication networks (terrestrial and satellite) for data transmission and on advanced fire and gas propagation modelling. The target solution aims to provide a complex overall tool capable to both detect the source which causes the event and evaluate in real time the evolution/diffusion of the danger, through the use of efficient estimation models, sending a reliable, secure, continuous communication to support the organizations handling relief operations and to inform emergency operators and people.

Then, three research lines must be developed in order to address all the uncovered technical issues towards the actual implementation of the system. In the sensor field, the main objectives are the minimization of the false alarms as well as the miniaturization of the sensor hardware; in the telecommunication field, the new Software Defined Networking (SDN) paradigm will be applied to guarantee a dynamic connectivity service and to improve the end-to-end performance, the integration of different technologies, the efficiency and the optimization of quality of service (QoS); finally, as far as the event simulation is concerned, recent techniques and tools based on cloud will be applied to mainly guarantee real time assessments. Also the subsystems integration will be cared, being particularly innovative and original.

Definitively, the project aims to develop technological systems to improve the operational efficiency of the Italian Firefighter body, which will actively participate with high qualified technical and scientific contribution especially for the requirement definition and the test phases.

II. State of Art

Among the technologies for the remote sensing, the LIght-Detection And Ranging (LIDAR) techniques represent a breakthrough for the detection of fires and chemical releases. These complex solutions are widely used for research purposes but their application on field is not so common [1-4].

Although there are a lot of systems on the market, called "Range finder LIDAR", their application is limited to distance and orographic measurements. A few devices on the market are able to reproduce the wind profiles (like the "Wind Cube" produced by Leoshere) or to analyze particulate profiles at distances up to 1 km (like the systems distributed by Vasala). Both these products are not allocable on UAV (Unmmanned Aerial Vehicle) or UTV (Unmmanned Terrestrial Vehicle)

because they are too heavy and need a meaningful electrical power. In addition, all these devices are customized for specific applications [5-6].

Therefore, the main problems related to a LIDAR system development are: 1) the dimension reduction; 2) the products engineering to let the system work for real application.

The Quantum Electronics and Plasma Physics Research Group at the University of Rome Tor Vergata, coordinated by Prof. Gaudio, has demonstrated, through the development of a laser system, that LIDAR system can be used for the early detection of forest fires [5-7]. The equipment available now has dimension incompatible with an easy transportability, high cost of single components and do not provide reliable data because of the high numbers of false positive [5-8]. On the other hand, two LIDAR systems tailored to monitor the environment in case of specific dispersion in air have been developed on prototypal releases and are very promising to accomplish the target objectives:

- COLI (Compact LIdar) system - The mini-Lidar system consists of a compact and mobile system based on Nd:YAG laser as transmitter that operates in NIR wavelengths.; it is able to monitor the environment measuring particulate concentrations in urban or industrial areas (figure 1);
- SAI system - Nd-Yag Laser system based on the LIDAR technique, mounted on a little cart, used during experimental campaigns all around Italy for forest fire detection, RAMAN applications and PBL analysis. (figure 2).

Figure 1: COLI system

Figure 2: Transmitter and receiver equipment of SAI system

The SAI system has been already used in several experimental campaigns both as a tool to predict the early fire forest detecting the particulate at the beginning of the combustion process and to analyze the quantity of particulate in air in industrial areas to monitor the quality of the air.

The reference operational environment is usually characterized by the lack of cabled telecommunication systems and cellular systems (UMTS, LTE [9]), even though present, may not be satisfactory in terms of overall coverage and experienced performance.

To fulfil the specific environmental constraints, satellite systems appear particularly suitable due to the offered global coverage and the potential cost-efficient interconnection with terrestrial systems due to the minimal deployment of ground nodes. The available satellite systems [10], as well as standards and technologies [11], allow ad hoc dimensioning and set up to optimize performance and costs.

Currently, Firefighters use proprietary fixed radio links, UMTS and WiFi equipment and satellite two way systems, but each is used as stand alone. It is very difficult to find integrated terrestrial/satellite systems that guarantee a dynamic configuration and the control of QoS, optimization of resources and security, or more in general telecommunication systems tailored for first alert. To improve on this, Cloud Computing and SDN paradigms are considered suitable to achieve the telecommunication objectives [12 - 13]. The former allows a more efficient use of computational and memory resources achieving cost reduction and decreasing complexity of user terminal. The latter represents the most recent innovation in network management, particularly effective solution in case of hybrid networks to select dynamically the best path by a centralized entity or by the user.

In the field of wildfire modeling and simulation, decision support systems make use of conventional fire spreading simulators (e.g., Farsite [14], Prometheus [15], Portable Fire Growth Model [16], Phoenix [17]), which provide the fire propagation speed as a function of environmental data, such as kind of fuel, humidity, wind and terrain. The two main approaches adopted by such simulators are the vectorial one, based on the Huygens principle, and the raster one, based on cellular automata that represents the terrain as a set of cells propagating the fire according to the proximity criterion. The former provides better accuracy, the latter better efficiency. In both cases, the simulators are used as offline tools to support planning activities of future operations to be carried out.

Distributed Simulation (DS) techniques may be effectively used for an accurate and an efficient real-time decision support, by taking into account the dynamic nature of the integrated and heterogeneous system. DS techniques combine the well-known advantages of sequential simulation with a distributed approach that yields significant levels of scalability, reusability and parallelism. In order to properly address the dynamic nature of the proposed system and thus provide an effective decision support, it is necessary to introduce an appropriate middleware for dynamic DS. In this respect, the MSaaS (Modeling and Simulation as a Service) paradigm, which introduces a cloud-based architectural approach to describe, discover and use on-demand simulation services, proves to be a promising solution

to define a middleware for dynamic DS [18], and then will be followed in the development of the target system.

III. Methodology, Objectives and Expected Results

The sensor subsystem will be designed and developed leveraging on the LIDAR technique. It will be composed of several elements designed to minimize dimensions and costs and maximize performance. It will be engineered to have autonomous performance, with a data acquisition equipment able to provide an alarm when the set thresholds are exceeded.

For an efficient monitoring of fires, a large number of nodes with communication interfaces must be deployed over the area of interest. Such nodes can be fixed (very sensible points), temporary (zones interested to particular events) or mobile (moving terminals to scan large areas). Each node can be coupled with different data acquisition devices (temperature sensors, cameras, etc.), together with proposed LIDAR-based sensor, implying the need to transfer heterogeneous data for type, quantity and required service quality: video, telemetry data, alert strings. The nodes will be equipped with different interfaces: terrestrial (WiFi, UMTS, LTE, etc.) and satellite [19]. The selection of interface will be dynamically handled by a centralised SDN controller.

The objectives of the telecommunication subsystem include the optimal interconnection of different technologies guaranteeing the target end-to-end QoS, the minimization of the needed capacity and the optimization of field-collected data transfer to the headquarter in order to maximize the reachability of population using all the available technologies: cellular, TV, radio, instant messaging (Skype, WhatsApp, Viber). A dynamic routing strategy, based on the SDN paradigm, will be adopted combined with mobile network routing. In addition, the integration of a satellite segment will be exploited to transport data to a remote control center to support coordination of operations to contrast the phenomena and from which alert messages will be broadcasted to rescue teams and to people. The satellite link must be designed to reduce intrinsic inefficiencies mainly due to delay [20]. To guarantee QoS, the SDN component will be integrated with the MAC algorithms optimized for demand assignment multiple access (DAMA). Figure 3 shows the scheme of the telecommunication system.

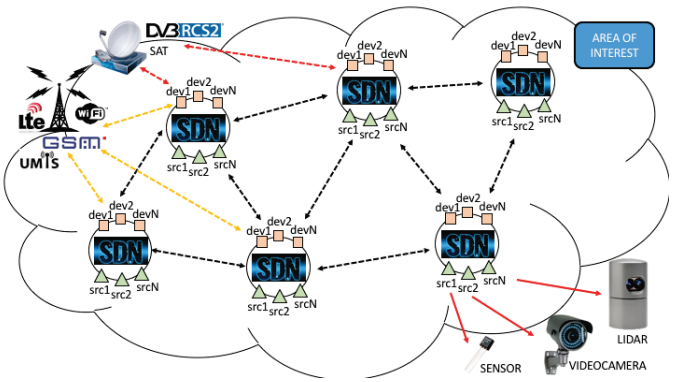


Figure 3: Telecommunication system architecture

The physical interfaces and the QoS management will be considered to guarantee data delivery satisfying specifications in terms of delay and signal degradation. The analysis will be supported by a conspicuous simulation and emulation activity using platforms, which accurately reproduce the behaviour of real networks, to evaluate architectural solutions (PEP) and innovative transport protocols both for single link and for a network [20-22]. A Satellite emulator is already available to the team and has been conceived and developed to reproduce in real-time the services (and quality of service) of a DVB-RCS satellite network, while some network impairments (e.g., end-to-end delay) are reproduced in software. The Emulator core is composed of 5 units (figure 4). Every unit-PC emulates a single component of a typical VSAT satellite environment, with DVB-RCS or DVB IP and terrestrial return channel. In particular, PCs represent: Gateway/Hub (SatGW), Satellite (SAT), Satellite Terminal (ST), Two User Terminals (UT1 and UT2). A particular emphasis is devoted in the management of the satellite return link resources that usually are allocated on demand following the ST needs, through a multiple access assignment scheme. The process regulating the resource competition and assignment has a relevant impact on the performance, so that a specific analysis aimed to identify the most suitable configuration is of paramount importance. In this regard, the available emulator implements a large set of customizable DAMA algorithms allowing an exhaustive experimentation. Further details on the emulation platform can be found at www.tlcsat.uniroma2.it and in [23-24].

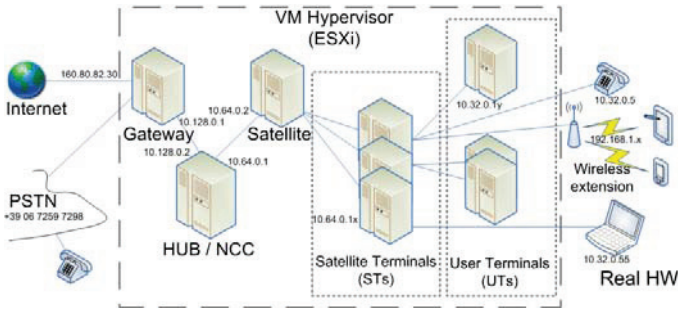


Figure 4: Satellite emulation platform

A testbed with real equipment will be developed for field trials exploitation with the active participation of Italian Firefighters Body. As expected results the network will satisfy all the requirements in terms of QoS, minimum delay and bandwidth necessary provided by the Firefighters. The overall emulation platform will allow to reproduce a pervasive and dynamic reference context, allowing to validate all the decision making activities on a real-time platform receiving as input data coming from real target devices. Thus, the main objective is to design and implement a simulation tool that overcomes the limitations of current decision support systems in terms of dynamicity, accuracy and computational efficiency.

The proposed decision support system, illustrated in figure 5, exploits distributed simulation techniques to integrate remotely-available and heterogeneous simulation components (for, e.g., wind, humidity, combustion of wood, etc.).

Currently available simulators, such as the aforementioned Farsite, Prometheus, Portable Fire Growth Model, Phoenix, as well as the simulators used by Italian Fire Corps, will be re-used and/or integrated. It will be necessary to provide semantic interoperability, in other words introduce a shift from the more conventional syntactic interoperability, which refers to the ability of exchanging data by defining an interchange format, to the ability of formally represent the meaning of exchanged data. The semantic interoperability, which will be obtained by use of domain-specific metamodels and model-driven automated approaches, is essential to provide the ability to build orchestrations of simulation components that meet specific user requirements [25-27].

Figure 5: Simulation and support to decision

The metamodels will be used to instantiate heterogeneous simulation models at various levels of granularity (multi-resolution modeling). Such models will then be implemented into simulation software components, which will be developed from scratch if not already available. The resulting components will finally be integrated by use of innovative distributed simulation (DS) techniques based on the MSaaS (Modeling and Simulation as a Service) paradigm [18].

To this purpose, it will be necessary to introduce a middleware for innovative dynamic DS, which takes into account the adaptive and reconfigurable nature of the underlying service-oriented platform. Such a DS middleware is the main expected research contribution and will allow the collection of autonomous simulators to combine actual data from sensors and data exchanged with back-end services, in order to predict the future state of the system and contribute to a more effective decision making.

In terms of system the ambitious and challenging objective is the use of three different technologies (sensors, telecommunications, simulation) integrated to provide support to decision realizing a real "system of systems". Figure 6 shows the general architecture of the system. The final goal to realize a demonstrator will be reached through analysis, simulations and field trials. To efficiently interconnect the three subsystems, interfaces will be particularly cared. Sensors will be equipped with USB or Ethernet or WiFi port that with different bit rate and flexibility will guarantee to transfer detected data to the telecommunication system. Data will travel over an IP platform and delivered to the decision support system in the right format or will be opportunely processed.

IV. Application potential and impact

The technology to be developed can be utilized for other applications: monitoring (pollution, weather, river, volcanoes, mines and glacier monitoring), hydrogeological instability monitoring, damage collection from natural or man made disasters, building and historical artwork survey.

The scientific impact concerns methods and techniques for:

- digital classification and automatic detection of alarm states and false positive minimization;
- more efficient use of tlc systems based on heterogeneous networks with end to end QoS guarantee;
- interoperability and reusability of distributed simulators based on service-oriented approaches.

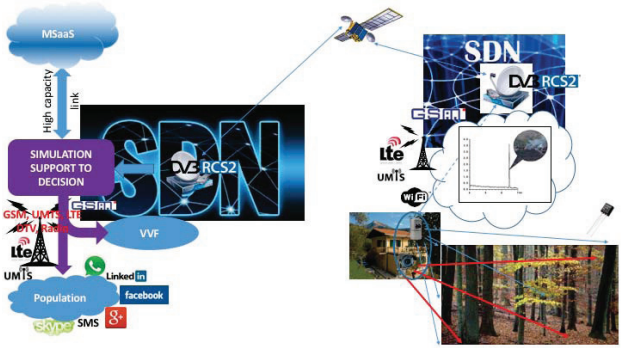


Figure 6: General system architecture

The technological impact concerns the development of a miniaturized LIDAR system that work with a quasi-real time electronic and high determinism. It will be a stand off system able to monitor huge volume, easily transportable with a tlc system based on existing components with optimized performances interconnected to guarantee high efficiency and QoS; a system able to accurately predict the evolution of a fire; the three technologies integrated to face the early fire detection.

The social impact concerns the potential applicability of the system concept to different application domains, the capability to significantly contribute to improve the security feeling of citizens, to guarantee greater safety to operators of the Fire Corps and to improve environment (less burned trees).

The economical impact can be direct or indirect. The former derives from the expected results in terms of better utilization of resources (operators and tools), due to greater automation and to availability of software tools, fed by real time data, that produce real time results that can achieve meaningful savings and drastic reduction of damages. The latter is a consequence of the social impact: with more security and safety the economy will increase.

V. Conclusions

The paper describes the main objectives and goals of a project proposed by a group of researchers, with the support of Italian Firefighter Bodies, to the Italian Ministry of Research,

at present under evaluation. The project idea can be applied also to other scenarios of emergency management and can be the basement for other projects in different contexts.

REFERENCES

[1] G. Fiocco, L.O Smullin, Detection of scattering layers in the upper atmosphere (60-140 Km) by optical radars, Nature, 199, 1275, 1963.

[2] F. Andreucci, M. Arbolino, A study on forest fire automatic detection system, Il Nuovo Cimento 1993, 16(1):35.

[3] F. Andreucci, M. Arbolino, A study on forest fire automatic detection, system 2—smoke plume detection performance. Il Nuovo Cimento 1993; 16(1):51.

[4] C. Bellecci, P. Gaudio, M. Gelfusa, T. Lo Feudo, A. Malizia, M. Richetta, P. Ventura, Raman water vapour concentration measurements for reduction of false alarms in forest fire detection, in Proceeding of SPIE, vol. 7479, 2009.

[5] G. Rustad, Technologies for standoff detection of liquid chemical warfare agents, Norwegian Defence Research Establishment (FFI-rapport 2011/01890), 2011.

[6] P. Gaudio, M. Gelfusa, A. Malizia, M. Richetta, C. Serafini, P. Ventura, C. Bellecci, L. De Leo, T. Lo Feudo, A. Murari, New frontiers of Forest Fire Protection: A portable Laser System (SAi), WSEAS Transactions on Environment and Development, 2013, vol. 9, p. 195-205.

[7] M. Gelfusa, P. Gaudio, A. Malizia, A. Murari, J. Vega, M. Richetta, S. Gonzalez, UMEL: A new regression tool to identify measurement peaks in LIDAR/DIAL systems for environmental physics applications, Review of Scientific Instruments, 2014, V. 85, 063112.

[8] J. Huang, J. Liu, B. Chen, S. L. Nasiri, Detection of anthropogenic dust using CALIPSO LIDAR measurements, Atm. Chem. Phys. Disc, 2015, 15, 10163-10198.

[9] ETSI TS 136 331 v10.7.0 (2012-11), LTE; Evolved Universal Terrestrial Radio Access (E-UTRA; Radio Resource Control (RRC); Prot. specs, 3GPP TS 36.331.

[10] H. Fenech, S. Amos, A. Tomatis, V. Soumpholphakdy, High throughput satellites: An analytic approach. 31 AIAA International Communications Satellite Systems Conference (ISCC AIAA), 2013.

[11] ETSI TR 101 545-4 V1.1.1 (2014-04) Digital Video Broadcasting (DVB); 2nd Generation DVB. Interactive Satellite System (DVB- RCS2); Part 4: Guidelines for Implementation and Use of EN 301 545-2.

[12] SDN Architecture, I 1, June 2014, www.opennetworking.org/images/stories/downloads/sdnresources/technicalreports/TR_SDN_ARCH_1.0_06062014.pdf

[13] T. Luo, H.-P. Tan, T. Q. S. Quek, Sensor OpenFlow: Enabling Software-Defined Wireless Sensor Networks, IEEE Comm. Lett Vol. 16, N. 11, 1896–1899, Nov. 2012.

[14] M. A. Finney, FARSITE: Fire area simulator – Model development and evaluation. Research Paper RMRS-RP-4 Revised. Ogden, UT: US Dep. of Agriculture, Forest Service, Rocky Mountain Research Station, Feb. 2004.

[15] C. Tymstra, R. W. Bryce, B. M. Wotton, S. W. Taylor, O. B. Armitage, Development and Structure of Prometheus: Canadian Wildland Fire Growth Simulation Model. Nat. Resource. Can., Can. For. Serv., North. For. Cent., Edmonton, Alberta. Inf. Rep. NOR-X-417, 2010.

[16] R. Shamir, "Appendix III. Portable Fire Growth Model", in J. Gould, Overview of operational fire spread simulation modelling - briefing paper to the Australasian Fire Authorities Council, Rural and Land Management Agency Sub-group. pp 30-35, 2007.

[17] K. G. Tolhurst, B. J. Shields, D. M. Chong, Phoenix: development and application of a bushfire risk management model, Australian Journal of Emergency Management 23(4): 47-54, 2008.

[18] P. Bocciarelli, A. D'Ambrogio, A. Giglio, and D. Gianni, "A SaaS-based automated framework to build and execute distributed simulations from SysML models," Proc. of the Winter Simulation Conference (WSC 2013), Washington, DC, USA, pp. 1371–1382, Dec. 8-11, 2013.

[19] M. Luglio, C. Monti, C. Roseti, A. Saitto and M. Segal, "Interworking between MANET and Satellite Systems for Emergency Applications", International Journal of Satellite Communications & Networking, vol. 25, n 5, 2007, pp.551-558, ISSN:1542-0973, DOI 10.1002/sat.890.

[20] C. Roseti, M. Luglio, F. Zampognaro, Analysis and Performance evaluation of a burst-based TCP for Satellite DVB RCS links, on IEEE/ACM Trans. on Networking, Volume 18, Issue 3, pp. 911-921, June 2010.

[21] F. Zampognaro and C. Roseti, M. Luglio, A. Detti, A. Caponi, Mobile-PEP: satellite terminal handover preserving service continuity, in the proceedings of Twelfth Intern. Symposium on Wireless Communication Systems, 2528 August 2015 — Brussels, Belgium.

[22] J. Border, "Performance enhancing proxies intended to mitigate link-related", IETF RFC 3135, June 2001.

[23] F. Belli, M. Luglio, C. Roseti, F. Zampognaro, "Evaluation of TCP performance over DVB-RCS multiple RCSTs", International Workshop on Satellite and Space Communications 2009, IWSSC 2009, 10th-11th September 2009, Siena, Italy, pp 424-428, DOI 10.1109/IWSSC.2009.5286320.

[24] M. Luglio, C. Roseti, F. Zampognaro, F. Belli, "An Emulation platform for IP-based satellite networks", 27th International Communication Satellite Systems Conference (ICSSC 2009), 14 June 2009, Edinburgh, UK.

[25] P. Bocciarelli, A. D'Ambrogio, G. Fabiani, "A Model-driven Approach to Build HLA-based Distributed Simulations from SysML Models", In: Proceedings of the 2nd International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2012), Rome, Italy, 28-31 July, 2012.

[26] P. Bocciarelli and A. D'Ambrogio, Model-driven Method to Enable Simulation-based Analysis of Complex Systems, in: D. Gianni, A. D'Ambrogio and A. Tolk (eds), Modeling and Simulation-Based Systems Engineering Handbook, CRC Press, ISBN 9781466571457, 2014.

[27] A. D'Ambrogio, D. Gianni, G. Iazeolla, A. Pieroni, "A Model Transformation Approach for the Development of HLA-based Distributed Simulation Systems", Proceedings of the 1st International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2011), Noordwijkerhout, The Netherlands, July 29-31, 2011.

Investigation of the Spatial Structure of the Analogue [LEU7] of Octarphin Molecule by Computer Modeling

L.N. Agayeva, N.A. Akhmedov
Institute for Physical Problems, Baku State University,
Baku, Azerbaijan
leylanamig@mail.ru

L.I. Ismailova, R.M. Abbasli
Institute for Physical Problems, Baku State University,
Baku, Azerbaijan
Namig.49@bk.ru

***Abstract* - The spatial structure of the analogue [Leu7] of octarphin molecule is investigated by means of the theoretical conformational analysis. The spatial structure of octarphin molecule appropriate to the fragment 12-19 of β-endorphin. The spatial structure of the molecule Thr1-Pro2-Leu3-Val4-Val5-Leu6-Leu7-Lys8-NH_2 calculated on the basis of low-energy conformations of the octarphin molecule. It is shown that the spatial structure of the analogue [Leu7] of octarphin molecule can be represented by 15 low-energy forms of the main chain. Low–energy conformations of the molecule, values of dihedral angles of main and side chains of amino acid residues are found out, the energy of intra- and interresidual interactions are evaluated.**

Keywords: β-endorphin; octarphin; theoretical conformational analysis; spatial structure; conformation

I. INTRODUCTION

The amino-acid sequence of the octarphin Thr1-Pro2-Leu3-Val4-Val5-Leu6-Phe7-Lys8-NH_2 corresponds to the fragment 12-19 of the β-endorphin. It is found that the octarphin molecule is related to macrophages of high affinity and specificity. The octarphin molecule is very interesting from the biological point view, therefore the its analogues

Leu1-Pro2-Leu3-Val4-Thr5-Leu6-Phe7-Lys8-NH_2
Thr1-Leu2-Leu3-Val4-Thr5-Leu6-Phe7-Lys8-NH_2
Thr1-Pro2-Leu3-Val4-Val5-Leu6-Phe7-Lys8-NH_2
Thr1-Pro2-Leu3-Val4-Thr5-Leu6-Leu7-Lys8-NH_2
Thr1-Pro2-Leu3-Val4-Thr5-Leu6-Phe7-Leu8-NH_2
were produced.

Amino acid sequence of octarphin conforms to the fragment 12-19 of β- endorphin. It is established that octarphin is related to macrophages of high affinity and specificity. Inhibition activity of octarphin analogues is 100 and more times as low as β- endorphin. It is shown that octarphin stimulates the activity of immunocompetent cells of the mouse in vitro and in vivo at the concentration 1-10 nM. It increases the adhesion and spreading of peritoneal macrophages, also their ability to digest the bacteria of virulent culture Salmonella typhiniurium 415 in vitro: intraperitoneal incorporation of peptide results in increasing peritoneal macrophage activity, T- B- lymphocytes of spleen as well [1].

The first analogue is not interesting for conformational analysis, because Leu1 is situated before the proline, and hence it has the situated conformation as the treonin. Therefore, use has studied the spatial structure of the analogue Thr1-Pro2-Leu3-Val4-Thr5-Leu6-Leu7-Lys8-NH_2 of the octarphin molecule, as well as, its [Leu2] and [Val5] analogues. The aim of the present article is investigation of the structure analysis of the analogue [Leu7] of the octarphin molecule.

Spatial structure of octarphin molecule Thr1-Pro2-Leu3-Val4-Thr5-Leu6-Phe7-Lys8-NH_2 has been investigated by method of theoretical conformational analysis. Calculations of conformational states of octarphin molecule are carried out regarding nonvalent, electrostatic and torsional interactions, hydrogen bonds as well. Nonvalent interactions are estimated by Lennard-Jones potential with Momany and Scheraga parameters [2]. Valence angles of amid groups as well as bond lengths and valence angles of side chains are in agreement with the values given in work [2] earlier. Electrostatic interactions have been calculated in the monopole approximation by Coulombs law using partial charges on the atoms suggested in work [2] earlier. Conformational possibilities of fragments and entire molecule of octarphin are calculated in conformity with the conditions of aqueous surroundings, in this connection the value of permittivity is taken to be 10 [3]. Hydrogen bonds are evaluated by Morze potential [4] are suggested to be weakened (bond energy in aqueous medium by optimal distance NH…OC 1,8Å is taken to be (-1,1) kcal/mol [4], torsional potentials and values of rotation barriers around the bonds $C^{\alpha} – N$ (φ), $C^{\alpha} – C$ (ψ), C – N (ω), $C^{\alpha} - C^{\beta}$ (χ) and other bonds of side chains are taken from work by Momany and coauthors [2].

Classification used for peptide structures has been plotted on the “tree” principle according to which all structural versions of peptide first fall within comparatively limited amount of types (shapes), each shape involves several certain forms of the main chain but each form is presented by a number of conformations having exact qualitative characteristics of amino acid residue geometry. The number of shapes for sequence of n residues is generally equal to 2^{n-1}, the number of forms in shape is defined by the number of combinations R, B, L, P of residue

forms, but the number of conformational states of one form is defines by the number of one form is defined by the number of rotational degree of freedom of the side chains of the residues.

To designate conformational states of the residues there has been used X_{ij}^{N} - typed identifiers, where X determines low-energy regions of conformational map φ-ψ: R (φ, ψ =-180-0°), B (φ = -180-0° ,ψ=0-180°), L (φ, ψ =0-180°), and P (φ=0-180°,ψ= -180-0°); N – the number of residue in sequence; *i, j*...=11…,12…,13…,21…, and etc. conform to the positions of the side chain (χ_1, χ_2…), in this case subscript 1 corresponds to the angle χ = 0-120°, 2 to χ = 120-(-120 °), 3 to χ = (-120)-0°.

Designation indications of dihedral angles have been measured up to the generally accepted nomenclature [5]. Given work is the extension of our investigations on structural and functional organization of peptide molecules [6-12].

II. Resuits And Discussions

Spatial structure of octarphin molecule has been studied in fragments. All the first stage conformational possibilities of N-terminal pentapeptide fragment Thr1-Pro2-Leu3-Val4-Thr5 and C- terminal tetrapeptide fragment Thr5-Leu6-Phe7-Lys8-NH_2 on the basis of low energy conformations of appropriate monopeptides have been investigated. On the base of these penta- and tetrapeptide fragments the three-dimensional structure of octarphin molecule is examined.

Initial structural versions of octarphin for energy minimization have been formed of eight forms of the main chain of N-terminal pentapeptide fragment (Thr1- Thr5) and eight forms of the main chain of C-terminal tetrapeptide fragment (Thr5-Lys8-NH_2). Among the structural versions of octarphin under the examination some of them turned out to be steric forbidden, some of them turned to be high-energy. The spatial structure of the molecule Thr1-Pro2-Leu3-Val4-Thr5-Leu6-Leu7-Lys8-NH_2 calculated using low-energy conformations of the octarphin molecule.

In Table 1 there have been presented the best optimal conformations of the analogue [Leu7] of the octarphin molecule which energy do not exceed 9,0 kcal/mol. They have 15 various forms of the main chain. There have been given energy contributions from nonvalent, electrostatic and torsional interactions for each conformation; energy of hydrogen bonds is included in value U_{nv}. Concerning geometry of N-terminal tetrapeptide fragment Thr1-Val4 given in Table 1 the low-energy conformations of octarphin fall into four groups (A-D). States of group A have the least free energy preferential both in internal energy and entropy. However, sets of conformations B and C are at disadvantage a little in relation to both factors. In Table 2 these has been shown the energy of intra- and interresidual interactions in three low-energy structures. Values of dihedral angles φ, ψ, ω of the main chain and angles χ of the side chains in these three conformations of the molecule are illustrated in Table 3.

In optimal conformations of the analogue [Leu7] of the octarphin molecule, a contribution of the nonvalent interactions changes in the energy limits (-42.6) – (33.6) kcal/mol, the electrostatical interactions in the range (7.5-10.8) kcal/mol, and the torsional interactions in the range (3.9 – 8.3) kcal/mol (Table 1). The analogue [Leu7] of the octarphin molecule has six conformations of the A type, three B type conformations, five C type conformations and only one D type conformation.

The molecule conformation $B_{12}BR_{22}R_1R_{12}B_{31}B_{21}B_{3122}$ of the eefffee shape is the most stable among various conformations. In this conformation the main chain Leu3-Val4-Thr5 has the R form, and the main chain of the remaining residues has the B form (Table 1). In such spatial fixing of the amino acid residues Thr1 residue effectively interact with Pro2; Pro2 with Thr5 and Leu6; Leu3 with Val4, Leu6 and Leu7; Thr5 with Leu6, Leu7 with Lys8 (Table 2).

TABLE 1. Relative energy (U_{REL}) and contributions of nonvalent (U_{NV}), electrostatic (U_{EL}), torsional (U_{TORS}) interactions of optimal conformations the analogue [Leu7] of octarphin molecule

Group	*№*	*Shape*	*Conformation*	$U_{nv.}$	$U_{el.}$	$U_{tors.}$	$U_{rel.}$
A	1	efeffee	$B_{12}RB_{21}R_1R_{12}B_{31}B_{32}B_{3222}$	-39.3	10.5	5.1	1.6
	2	efeeffe	$B_{12}RB_{23}B_1R_{12}R_{22}B_{21}B_{3122}$	-38.0	7.5	7.7	2.5
	3	efeefee	$B_{12}RB_{23}B_1R_{12}B_{31}B_{21}B_{3122}$	-39.2	9.2	6.0	1.3
	4	efeefff	$B_{12}RB_{23}B_1R_{12}R_{21}R_{32}R_{2122}$	-38.8	9.2	6.1	1.7
	5	efeefef	$B_{12}RB_{23}B_1R_{12}B_{21}R_{31}R_{2122}$	-37.6	9.5	6.8	4.0
	6	efeeeff	$B_{12}RB_{23}B_1B_{12}R_{32}R_{11}R_{3222}$	-40.1	10.8	7.1	3.1
B	7	eefffff	$B_{12}BR_{22}R_1R_{12}R_{21}R_{32}R_{2222}$	-40.7	8.5	7.2	0.3
	8	eefffee	$B_{12}BR_{22}R_1R_{12}B_{31}B_{21}B_{3122}$	-42.6	9.2	8.1	0
	9	eefffef	$B_{12}BR_{22}R_1R_{12}B_{21}R_{31}R_{2122}$	-36.0	10.7	8.3	8.3
C	10	effefee	$B_{12}RR_{21}B_1R_{32}B_{31}B_{31}B_{3122}$	-38.6	9.1	6.1	1.9
	11	effeffe	$B_{12}RR_{21}B_1R_{12}R_{22}B_{21}B_{3122}$	-38.1	8.3	7.7	3.2
	12	effefef	$B_{12}RR_{21}B_1R_{12}B_{21}R_{31}R_{2122}$	-36.2	9.6	5.4	4.1
	13	effefff	$B_{12}RR_{21}B_1R_{12}B_{21}R_{31}R_{2122}$	-38.4	10.4	6.3	3.6
	14	effeeee	$B_{12}RR_{21}B_1B_{12}B_{21}B_{32}B_{1222}$	-33.6	9.9	3.9	5.5
D	15	eeeeffe	$B_{12}BB_{23}B_1R_{12}R_{22}B_{21}B_{3122}$	-38.0	7.5	7.7	2.5

TABLE 2. ENERGY OF THE INTRA- AND INTER-RESIDUE INTERACTIONS (KCAL/MOL) IN CONFORMATIONS THE ANALOGUE [LEU7] OF OCTARPHIN MOLECULE B12BR22R1R12B31B21B3122 (UREL .=0 KCAL/MOL, THE FIRST LINE), B12RB23B1R12B31B21B3122 (UREL .=1.3 KCAL/MOL, THE SECOND LINE), AND B12RR21B1R32B31B31B3122 (UREL. =1.9 KCAL/MOL, THE THIRD LINE) B12BB23B1R12R22B21B3122 (UREL. =2.5 KCAL/MOL, THE FOURTH LINE)

Thr1	Pro2	Leu3	Val4	Thr5	Leu6	Leu7	Lys8	
2.8	-4.6	-0.9	0	0	-0.2	0	2.3	Thr1
2.1	-4.1	-2.4	-2.6	-0.1	-0.1	0	2.1	
4.8	-4.0	-1.2	-2.0	-3.7	-0.1	0	1.7	
2.2	-4.0	-2.3	0	0	0	0	1.7	
	0.3	-0.3	-0.4	-1.4	-3.1	-0.1	-2.2	Pro2
	0.3	-1.9	-1.3	0	0	0	-0.1	
	0.3	-1.3	-0.6	0	0	0	0	
	0.3	-1.9	-1.4	0	0	0	0	
		-0.2	-2.4	-1.4	-1.8	-2.5	-1.9	Leu3
		0.2	-2.1	-0.6	0	0	0	
		-1.2	-3.4	-0.2	0	-0.1	0.1	
		0.3	-2.1	-0.7	-1.9	1.2	-1.3	
			0.9	1.5	-0.7	-1.0	-0.1	Val4
			0.6	-1.8	-1.2	-3.7	0.1	
			0.8	-1.5	-2.2	-3.1	-0.2	
			0.6	-1.8	-1.0	-3.0	0.1	
				0.1	-3.5	-1.2	0	Thr5
				0	-3.4	-1.4	0	
				0.3	-2.7	-2.0	0.1	
				0	-0.9	-0.9	0	
					-0.5	-1.4	-3.8	Leu6
					-0.3	-1.3	-4.0	
					-0.5	-2.1	-4.1	
					0	-3.0	-3.6	
						-0.2	-4.2	Leu7
						-0.5	-4.1	
						0	-3.1	
						-0.7	-3.8	
							1.7	Lys8
							1.6	
							1.6	
							1.4	

TABLE 3. GEOMETRIC PARAMETERS (DEGREE) OF THE OPTIMAL CONFORMATIONS THE ANALOGUE [LEU7] OF OCTARPHIN MOLECULE

Residue	***Shape***			
	e e f f f e e	***e f e e f e e***	***e f f e f e e***	***e e e e f f e***
Thr1	-66 150 175 46 -175 179	-42 152 173 52 -174 180	-111 150 175 49 -179 -179	-44 151 174 52 -175 -178
Pro2	-60 132 -173	-60 -60 173	-60 -65 -176	-60 -61 173
Leu3	-110 -62 -171 -173 158 -169 180	-82 104 180 -176 -61 170 -177	-108 -69 -172 176 63 179 176	-81 104 -178 -166 -61 170 -167
Val4	-65 -36 177 73 179 -175	-144 148 180 62 177 -178	-139 152 177 55 175 177	-144 148 -175 62 178 180
Thr5	-79 -66 177 57 -177 174	-112 -64 177 57 -179 174	-101 -49 178 -51 -178 174	-112 -56 -173 60 -179 177
Leu6	-90 87 -178 -75 67 180 172	-87 79 178 -74 66 180 172	-85 95 174 -76 62 180 171	-62 -66 -173 -172 154 -171 180
Leu7	-92 147 -175 -176 65 180	-96 144 -177 -178 67 180	-88 147 -177 -72 68 180	-98 139 180 -179 68 180

	180	180	180	178
Lys8	-95 83 179 -72 70 166 -179 -178	-97 82 179 -74 68 167 179 -177	-97 80 179 -76 68 167 -179 -178	-99 103 179 -75 69 165 179 179
$U_{rel.}$	0	1.3	1.9	2.5

Note: The values of dihedral angles are given in the sequence φ, ψ, ω, χ_1, $\chi_{2, \ldots}$

The conformation $B_{12}RR_{21}B_1R_{32}B_{31}B_{31}B_{3122}$ of the effefee shape with relative energy 1.9 kcal/mol is the conformation with lowest energy in the group C. In this conformation N-ended pentapeptide generates semi-folded form of the main chain, C-ended tripeptide is in an extended form, and derivates from the N-ended part of the molecule (Table 1). In this structure the effective interactions between neighboring amino acid residues take place.

The conformation $B_{12}RB_{23}B_1R_{12}B_{31}B_{21}B_{3122}$ of the efeefee shape with relative energy 1.3 kcal/mol is the lowest energetical conformation of the B group (Table 1). In this conformation the residues Pro2 and Thr5 turn over the peptide chain of the molecule. As is seen from Table 2, in the chain appear effective interactions between the residues Pro2 with Leu3, Thr5 and Leu6; Leu7 with Lys8.

The D group has only one conformation B_{12} $BB_{23}B_1R_{12}R_{22}B_{21}B_{3122}$ of the eeeeffe shape with relative energy 2.5 kcal/mol. In this conformation N-ended tetrapeptide and C-ended dipeptide fragments are in the extended form of the main chain, and are folded relatively to each other due to the residues Thr5-Leu6.

Our computations have demonstrated, that in the most low energetically spatial structure of the analogue [Leu7] of the octarphin molecule, the side chains of the amino acid residues take forms, which are favorable to establish intermolecular interactions with environment. The theoretical conformational analysis of the analogue [Leu7] of the octarphin molecule led to such structural organization of the molecule, which do not exclude realization by the hormone numerous functions, requiring specific interactions with various receptors.It is known, that in creating synthetic analogues of the natural molecules emerge three main problems: 1) peptide should be stable in the biological media, 2) it would interact with only some specific receptors, 3) it has to reach target cells. The analogue [Leu7] of the octarphin molecule meets the aforementioned requirements. Our studies have shown, that the spatial structure of the analogue [Leu7] of the octarphin molecule has 15 low-energy structures as the octarphin molecule itself.

Would the number of the low-energy conformations of the analogue [Leu7] of the octarphin molecule less than ones of the natural molecule, would be better, because in this situation the biological functions of the octarphin molecule would reduce, and the analogue would perform only some of functions of the natural molecule. Therefore, the studied in the present work analogue is not best one.

The calculation shows that side chains Thr1, Pro2, Leu3, Val4 and Leu6 are oriented within the molecule, but the side chains Leu7 and Lys8 are oriented to the medium of the most low-energy spatial structures the analogue [Leu7] of octarphin molecule. Otherwise they take the positions which are the most suitable regarding intermolecular interactions. Theoretical conformational analysis of octarphin results in such structural organization of the molecule that do not exclude the hormone realization of the entire group of the most various functions requiring strictly specific interactions with the different receptors.

REFERENCES

[1] Yu. N. Necrasova, V. B. Sadovnikov, Yu. A. Zolotarev. E. V. Novolotskaya. “Properties and mexanizm of action of synthetic peptides octarphin” Biorg.Chem., vol. 36, pp. 638-645, 2010.

[2] F. A. Momany, R. F. MeGrure, A. W. Burgess, H. A. Scheraga. “Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bonding interaction occurring amino acids”. J.Phys.Chem., vol. 79, pp. 2361-2381, 1975.

[3] G. H. Lipkind, S. F. Archipova, E. M. Popov. “Theoretical conformational analisis of metylamides of N-acetyl-L-amides.” J.Struct.Chem., vol. 11, pp. 121-126, 1970.

[4] E. M Popov, V. G. Dashevski, G. M. Lipkind, S. F. Archipova. “The method of calculating the conformation of peptide chains. Potential function.” J.Molek.Biol., vol. 2, pp. 612-620, 1968.

[5] IUPAC-IUB, Quantiby, Units and Sybbols in Physical Chemistry. Blackwell Scientific Publications, Oxford, vol. 39, 1988.

[6] N. A. Akhmedov, Z. H. Tagiyev, E. M .Hasanov, G. A. Akverdieva. “Theoretical conformational analisis of the bovine adrenal medulla 12 residue peptide molecule” J.Molecular Structure, vol. 646, pp. 75-80, 2003.

[7] N. A. Akhmedov, L. I. Ismailova, R. M. Abbasli, N. F. Akhmedov, N. M. Gocayev. “Spatial Structure of Myelopeptides: I.Conformational Analysis of MP-1, MP-2 and MP-3.” Biorg.Chem., vol. 31, pp. 31-38, 2005.

[8] L. I . Ismailova, N. A. Akhmedov, R. M. Abbasli. “Spatial Structure of Glu-Phe-Leu-Arg-Ile-NH_2 and Pro-Phe-Tyr-Arg-Ile-NH_2 izoleysin pentapeptides.”Biofizika, vol. 53, pp. 14-21, 2008.

[9] N. A. Akhmedov, L. I. Ismailova, R. M. Abbasli, L. N. Agayeva, S. R. Akhmedova. “Spatial Structure of Octarphin Moleculeş” İOSR J.of Applied Physics (İOSR-JAP) vol. 8, pp. 66-70, 2016.

[10] S. Celik, A. E. Ozel, S. Akyuz, S. Kecel, G. Agayeva. “Conformational preference, exsperimental and theoretical vibrational spectra of cyclo (Gly-Val) dipeptide.” J.Molecular Structure, vol. 993, pp. 341-348, 2011.

[11] N. A. Akhmedov, Sh. N. Gadjiyeva, R. M. Abbasli. “Structural organization of Asp-Pro-Lys-Gln-Asp-Phe-Met-Arg-Phe-NH_2 molecule.” Peptid and Protein Research, vol. 10, pp. 57-62, 2009.

[12] N. A. Akhmedov, L. I. Ismailova, L. N. Agayeva, N. M. Gocayev. “The spatial structure of the cardio active peptides”. Peptid and Protein Research vol. 11, pp. 87-93, 2010.

Automated Recognition System of Statistic Gestures via Kinect Sensor

Saule Kudubayeva
Department of Informatics
Kostanay State University named after A. Baytursynov
Kostanay, Kazakhstan
saule.kudubayeva@gmail.com

Dmitriy Ryumin
Department of Informatics
Kostanay State University named after A. Baytursynov
Kostanay, Kazakhstan
dl_03.03.1991@mail.ru

Marat Kalghanov
Kostanay State Pedagogical Institute
Kostanay, Kazakhstan
mkkiev@mail.ru

Almatova Assem
Department of Informatics
Arkalyk pedagogical Institute
Arkalyk, Kazakhstan
alasmur01@mail.ru

***Abstract* — The paper presents a system of recognition of elements of a sign language. The description of the elements of Kazakh sign language database is given. The algorithm of allocation of human hands on images obtained via the Kinect sensor and the method of calculation of informative video features of hands gestures are suggested. The implementation of recognition of isolate hand gestures on the basis of the elements of decision theory is given. The results of the automated system operation are illustrated.**

***Index Terms* — Signs, gestures, dactyl, dactylology, vocabulary, semantics.**

I. INTRODUCTION

Currently governments of both developed and developing countries in collaboration with leading research centers and companies start paying due attention to designing intellectual technologies and systems based on speech and multimodal human-machine interfaces.

According to World Health Organization's statistic data of 2015 officially there are 360 million deaf people and people with hearing problems, 328 million of whom are adults and 32 million are children [1].

At present there are not enough applications for reliable use in the sphere of gestures' automated recognition. Besides existing technical and economic obstacles the development of speech technologies (including those of sign language recognition) is first of all influenced by such peculiarities of one or another language and speech as variety of word-formation rules, language variants, etc., which cause difficulties in automated processing.

Availability of information resources (databases) of gestural speech is an obligatory pre–condition for teaching any modern gestural speech recognition system based on probability models as well as on various methods of video stream processing.

Numerous conventional methods and algorithms of digital signal processing are based on the use of signal filtering, correlation and spectral analyses, wavelet transformations. These techniques allow identifying visual images via methods of segmentation, clustering, classification using probabilistic modeling and video signal recognition based on different types of artificial neuronal nets, hidden Markov chains, dynamic programming and pixel video processing. An intellectual system leads to automated recognition and processing of gestural speech.

Developed human-machine interfaces, which incorporate "assistive technologies" [2], provide an opportunity to help people with health problems and individual peculiarities. Consequently, intellectual technologies are directed at social and professional rehabilitation of people with hearing problems, those who need automated systems allowing recognizing gestures.

II. DIGITAL REPRESENTATION OF THE IMAGE

Despite a great number of ways of image recording, their aim is more or less the same: obtaining data using sensitive sensors and building the image based on them. The majority of such sensors transmit an output analog signal in the form of constantly changing voltage. Such a signal should definitely be transformed into the digital form. This procedure includes two processes – sampling and quantization.

The image includes continuous coordinates x, y and amplitude. To transform such an image into a digital form one should present it as references to coordinates as well as to the amplitude. Sampling allows representing coordinates as a known set of points. The amplitudes, represented by a finite set of values, are usually called quantization.

As an example we may take an image and show it as a one-dimensional function, which represents the graph of variable brightness values in the range between 0 and 255. For sampling of this function each line of the image should be divided into equal intervals. As a result there forms a set of discrete values,

obtained at the points of sampling, which describes the one-dimensional function. Then one should resort to quantization, which is carried out via a simple juxtaposition of each discrete value with the closest match from a special grayscale [3].

Thus, there form digital samples on which a two-dimensional image is based. General view of the digital image is to be represented as a numerical matrix, which consists of $M \times N$ size. Such a matrix has a form corresponding to ratio 1

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-0,N-1) \end{bmatrix} \quad (1)$$

A separate element of such a matrix is usually called a pixel or an image element [4].

The origin of coordinates of any digital image is located in the upper left–hand corner. Y–axis goes right and x–axis goes down. This phenomenon can be explained by the fact that the majority of graphic displays scan images from the upper left-hand corner with line–movement to the right.

Normal image processing also presupposes that pixel values should range from 0 to 255 and with greater dynamic range. If we take lowercase p as a pixel designation, then each p has (x, y) coordinates as well as it almost always has four neighbours horizontally and vertically in accordance with Eq. 2.

$$(x+1, y), (x-1, y), (x, y+1), (x, y-1) \quad (2)$$

Such a set is usually called a neighbor quartet of p and is designated as $N_4(p)$. Besides horizontal and vertical neighbours there are diagonal ones with the coordinates represented by Eq. 3.

$$\begin{aligned}(x+1, y+1), (x+1, y-1),\\ (x-1, y+1), (x-1, y-1)\end{aligned} \quad (3)$$

Neighbours of such a type are designated as $N_D(p)$. Totally such points form an eight of neighbours with the $N_8(p)$ designation.

III. Gestures Recognition

While designing automated information systems it is necessary to have access to multimodal databases aimed at information structuring. Structuring presupposes a clear allocation of components (elements), connections between them as well as typing of components and connections at which definite semantics and admissible operations correlate with the type of the component (connection) [5]. Thus, for example, Kogalovsky M. R. In his work "Encyclopedia of Databases Technologies" describes more than 50 varieties of databases [6], however, there are no universal classifications of them, because new varieties of databases are constantly appearing, also many databases are hybrids of already existing ones.

In the course of the research it is ascertained that the optimal model of database design capable of storing three-dimensional signals obtained from Kinect sensor is a hierarchical model. By means of the hierarchical model it is possible to represent all necessary data in total in the form of a file system consisting of root directory and the hierarchy of subfolders and files in various formats on Fig. 1.

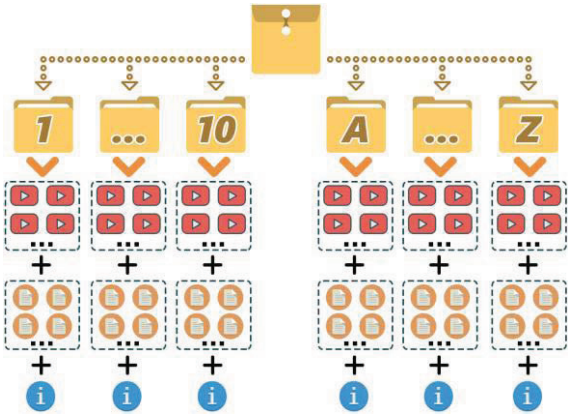


Fig. 1. File System of Hierarchical Database

Recording of the Kazakh Sign language elements was performed via Microsoft Kinect 2.0 sensor at a distance ranging from 1.5 to 2 meters, which is the optimal distance. The wall served as a uniform light background. Capture of video stream was carried out with the resolution of 1920x1080 pixels and at 25 frames per second.

The root directory of the database includes 52 subfolders, 10 of which contain the information about gestures showing numbers from 1 to 10. The remaining subfolders preserve the data about the dactyl of the Kazakh Sign language consisting of 42 letters.

A separate subfolder includes 30 video files with one and the same gesture recorded, as well as the same number of text files with the coordinates of the skeleton of a person found in the stream of frames via Kinect sensor, divided into 25 joints. Each definite point is the intersection of two axes (X, Y) on the coordinate plane and the additional Z designation with double accuracy, denoting the depth of the point which is measured by the distance from the sensor to the point of the object within the range between 0 and 1.

Besides the described files there is a text file bearing service data about the gesture. The average time of one video file is about 4 – 5 seconds.

For teaching the automated system of separate gestures recognition video files in combination with text files with coordinates about necessary skeletal points were used.

The flowchart of hand isolation which is fed with video files recorded via Kinect 2.0 looks like that shown in Fig. 2. Then goes looping of frames, within which there is a checking up for receiving a definite frame on each iteration. In case of an error the algorithm terminates its operation, otherwise threshold brightness clipping is being selected until the optimal value is obtained. Then filling the inner regions of the objects goes. If there are no objects on the image, the iteration is

considered complete and one needs to try to obtain the next frame of the video sequence, otherwise elimination of excess noise and counting of objects based on the calculation of their area are carried out. As the result of the algorithm operation graphic areas of hands on the frames are defined. Output data of the algorithm are the images showing human hands. The average processing speed of a frame is about 0.12 seconds provided that the resolution is FullHD 1920x1080 pixels and calculations are carried out with a single core processor Intel Core i7 3,4 GHz.

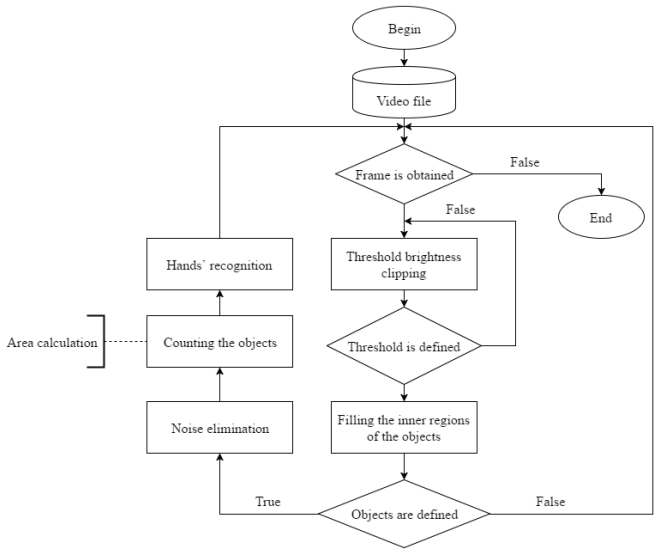


Fig. 2. Flowchart of the Algorithm of Hands Isolation on the Images

Algorithm of Hands Isolation on the Images (Fig. 3) is the first step of processing before the calculation of informative features of a human hand gestures.

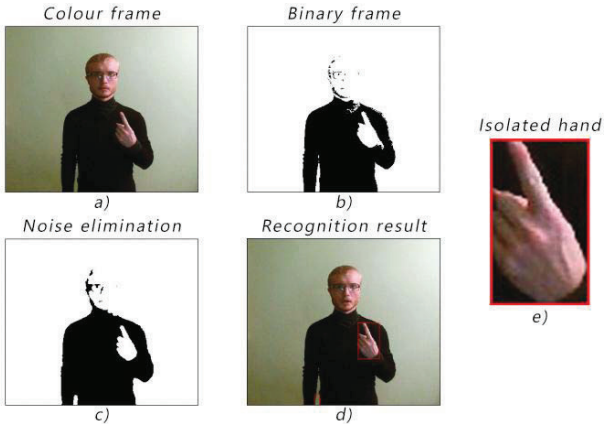


Fig. 3. Presentation of the Algorithm of Hands Isolation in process: a) colour frame, b) binary frame, c) noise elimination, d) recognition result, e) isolated hand

Via skeletal data obtained from Kinect sensor it is possible to show a separate hand as an ellipse, major semi-axis of which passes through points located in the beginning of the hand, the center and tops of middle and ring fingers. Minor semi–axis also passes perpendicularly to the major axis through the point with the coordinates of the hand center.

When the ellipse is defined it is possible to identify the first informative feature as the hand's orientation. The value is the angle between the x abscissa axis of the coordinate plane and the major axis of an object (hand).

Further interconnected features, which are necessary, are the lengths of the ellipse axes. However, Kinect sensor is incapable of defining the vertices coordinates, which will naturally be reflected in erroneous values, therefore it is necessary beforehand via the worked out algorithm of hand isolation to obtain binary image of the hand, which will allow to separate the background from the hand and receive optimal values of the lengths of both major and minor axes of the ellipse. Then one needs to define eccentricity by division of the length of the major axis by the length of the minor axis and to add the obtained value to already identified features.

Other features are topological properties of the binary object, such as the number of holes inside the object and the number of Euler.

Further analyzing of the binary object allows calculating its area and convexity factor, which is defined as the ratio of the object to the area of the quadrangle completely incorporating the object.

At the next step by using the Sobel detector [7] it is possible to identify the border of the hand, to find its length and diameter. Then one should add the obtained values to the array of features.

Collectively obtained informative features will characterize a hand in the general view, without being able to determine with a high probability the gesture shown. That is why it is necessary to carry out breaking [8] by using the structural component [9] in the form of a circle with the diameter of 1/5 of the ellipse minor axis length. This particular diameter will allow cutting the fingers off the center of the hand.

As a result one may obtain the fingers as objects in which it is possible to define the same features as those of the hand. Such a procedure will allow imagining both the hand as a whole and the fingers as its constituent parts. The array of informative video features of hands gestures looks like that shown in Tables 1, 2. The values of the features are stored in identifiers, designations of which do not coincide. This allows using the features in any sequence in the process of recognition.

TABLE I. INFORMATIVE FEATURES OF A HAND

Identifier	Feature
hand_or	Hand orientation
hand_maj_axis_len	Major axis length of the ellipse
hand_small_axis_len	Minor axis length of the ellipse
hand_eccentr	Eccentricity
hand_open	Number of inner holes
hand_eul_num	Number of Euler
hand_area	Area
hand_convex	Convexity factor
hand_bord_len	Border length
hand_bord_diam	Border diameter

TABLE II. INFORMATIVE FEATURES OF A FINGER

Identifier	Feature
fing_or	Finger orientation
fing_maj_axis_len	Major axis length of the ellipse
fing_small_axis_len	Minor axis length of the ellipse
fing_eccentr	Eccentricity
fing_open	Number of inner holes
fing_eul_num	Number of Euler
fing_area	Area
fing_convex	Convexity factor
fing_bord_len	Border length
fing_bord_diam	Border diameter

In the majority of cases the main methods of recognition may be divided into two categories: structural methods and methods based on decision theory.

Structural methods allow identifying the images with qualitative descriptors. These are relational descriptors [10].

If the images are described by quantitative descriptors, such as length, area, texture, etc., one should apply decision theory.

Each gesture is represented by the image. The image in its turn is an ordered set of descriptors forming a feature vector in accordance with Eq. 4.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (4)$$

In which component x_i – i–descriptor, n – the total number of descriptors.

The images with some similar properties form a class. At this stage of the research the system of recognition totally contains 52 classes designated as $w_1, w_2, \dots, w_{52}$. Each class contains 5 prototypes which are the proper standards of showing a certain gesture. The process of comparing the reference vectors with the test ones is based on the calculation of the Euclidean distance between them. Belonging of the object to some class is estimated on the basis of a minimum distance between the standard and the unknown object (Fig. 4).

Fig. 4. Gestures recognition

IV. CONCLUSION

Implementation of the system on the whole was made with the help of tools of mathematical calculations MATLAB [11], which allowed investigating and designing a model with alternative approaches to the solution of the problems mentioned. Auxiliary software and Open Graphics Library libraries [12], Open Source Computer Vision Library [13] were also used. Official software development kit [14] from Microsoft provided an opportunity to obtain the necessary data from the Kinect sensor.

The proposed system of recognition of some static hand gestures was tested on the database including 1560 gestures of some testes showing numbers within the range from 1 to 10 and 42 letters of dactyl of the Kazakh Sign language. Each class includes a gesture shown 30 times. Standards of each gesture are chosen randomly in the amount of 5 samples for each class. The remaining 25 samples are test samples.

The results showed the average recognition accuracy of 87%. These results were obtained on the database which has previously been cut into fragments without additional processing.

This system, with preliminary training, may also be used for other sign languages.

REFERENCES

[1] The World Health Organization. Available at: http://www.who.int/mediacentre/factsheets/fs300/en/ (accessed: 20 may 2016).

[2] A. A. Karpov, “Assistive Information Technologies based on Audio-Visual Speech Interfaces.”, Tr. SPIIRAN, 27 (2013), pp. 114–129.

[3] P. Koenig, “Graphic drawing for professional designers”, SPb.: Peter, 2014, 50 p.

[4] V. Komyagin, S. Lander, “Adobe Photoshop CS5 from scratch!”, M.: The Best Books, 2011, pp. 31–32.

[5] T. Connolly, C. Begg, “Database Systems: A Practical Approach to Design, Implementation, and Management”, M.: Williams, 3th, 2003, 1436 p.

[6] M. Kogalovskii, “Encyclopedia database technologies”, M.: Finance and Statistics, 2002, 800 p.

[7] D. Khryashchev, “A method for edge detection in digital images”, Herald ASTU. Series: Control, computer engineering and computer science, vol. 2, 2010, 182 p.

[8] I. Ognev, N. Sidorova, “Image processing mathematical morphology methods in oscillatory associative environment”, Proceedings of the universities, vol. 4, 2007, pp. 88–89.

[9] Morphological structuring element. Available at: http://www.mathworks.com/help/images/ref/strel-class.html (accessed: 21 may 2016).

[10] H. Verne, V. Vasiliev, K. Safronov and I. Chushenkov, “Hierarchical iterative method of pattern recognition in the analysis of multidimensional data”, Herald USATU, vol. 5, 2007, pp. 3–8.

[11] MathWorks – Makers of MATLAB and Simulink. Available at: http://www.mathworks.com (accessed: 21 may 2016).

[12] OpenGL – The Industry Standard for High Performance Graphics. Available at: https://www.opengl.org (accessed: 21 may 2016).

[13] OpenCV. Available at: http://opencv.org (accessed: 21 may 2016).

[14] Kinect for Windows SDK 2.0. Available at: https://msdn.microsoft.com/en-us/library/dn799271.aspx (accessed: 21 may 2016).

The Technique of Creating Distributed Computing Systems based on Service-Oriented Architecture

Lyashov M.V. Ph.D.
Don State Technical University
Rostov on Don, Russia
e-mail: maxl85@mail.ru

Bereza A.N. Ph.D.
Institute of Service and Business (branch) Don State Technical University
Shakhty, Russian

Babaev A.M., Alekseenko J.V, Nazvantsev D.S., students.
Institute of Service and Business (branch) Don State Technical University
Shakhty, Russian
e-mail: babaev-andrei2011@yandex.ru, ajulyav@mail.ru, nazvantsenv@yandex.ru

Abstract: The paper provides a solution to the problem of uniting heterogeneous computing resources into a unified heterogeneous system using service-oriented architecture (SOA) and the conception of GRID-computing, where computing resources of clients are presented as a service. The use of SOA for creating GRID-systems will allow the clients to implement integration of not only local separated heterogeneous computing resources, but also geographically separated ones. The suggested platform is not intended to replace or displace other GRID-systems. Instead of this, it gives the possibility of using different platforms based on SOA in a unified computing process.

Key words: high performance computing, service-oriented architecture, GRID-computing, heterogeneous system, .NET, Windows Communication Foundation.

I. Introduction

At the present time the need for large-scale calculations has increased not only for scientific purposes, but also in manufacturing, for example in material science, chemistry, development of drugs, engineering of high-tech machines etc. Supercomputer systems can solve a wide range of tasks, but the production and exploitation of these computers require large expenditures, therefore there are only several hundred such computers in the modern world.

GRID-systems have the ability to combine heterogeneous resources. As resources for GRID-systems can be used general-purpose processor (CPU), graphics processor (GPU), special hardware accelerators, based on field programmable gate array (FPGA), memory, disk spaces, networks etc. GRID-computing is widely used for rapid and low-cost processing a large amount of computational tasks. This is usually achieved by dividing complex calculations into a large number of simpler subtasks, which can be processed in parallel using available computing resources [1, 2].

Technologically GRID-systems can be divided into [3]:

- Dedicated, i.e. to fulfill computational tasks it is necessary to dedicate required resources which will solve only one task for a predetermined period of time.
- Volunteer computing: performers donate their available computing resources to process resource-intensive tasks.

In both cases, it is necessary to solve the problem of unifying heterogeneous computing resources into a unified distributed system while organizing the GRID-computing. Now in the IT industry SOA became widespread for creating information systems, because of its main advantage: the possibility of effective integration of information resources.

SOA is a module approach for software development. The approach uses distributed, replaceable and loosely coupled components that interact with each other according to standard protocols (SOAP, etc.). In this case components are implemented as a set of web services [4-6].

In this architecture clients can discover and connect to available services via public interfaces. The positive side of SOA is that the clients do not know about platforms, programming languages and about other technical aspects which are used on the service side. Thus, the client task is just to find and use those services that meet their needs. SOA can be effectively used in GRID-systems, because it will allow clients to get access to GRID-resources easily and flexibly, regardless of what technologies are used to build them.

This paper describes the possibility of the application of SOA concept for distributed computing. The use of SOA for creating GRID-systems will enable to implement the integration of heterogeneous computing resources which are not only local, but also geographically separated.

II. Technologies, Used for Building GRID-Systems on the Basis of SOA

Consider basic technologies that can be used for building GRID-systems based on SOA [4-6].

A web-service is a software system with a standardized interface that is available on a specific web address. Interaction of web services with other services or third-party applications is carried out by SOAP-based protocols, XML-RPC, REST, etc.

To provide the interaction of web-services with other services or third-party applications, the following standards are used:

- XML: an extensible markup language designed for storage and transfer structured data.
- SOAP: a messaging protocol based on XML;
- WSDL: a description language of external interfaces of a web-service based on XML;

A web-service consists of several parts that have their own function and behavior in a system. Fig. 1 shows the structural chart of a web-service.

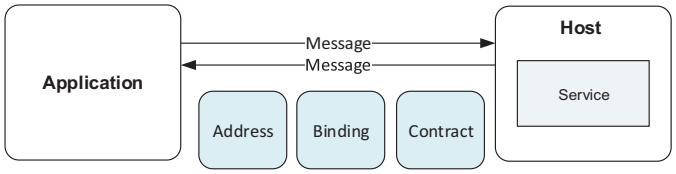


Fig 1. Web-service.

Any software platform, which controls a computing resource, can be as a service for GRID-systems. Clients of this service use a channel, which is compatible with a service channel, to call service methods, transfer necessary data for computing and get back results of these calculations. Messages can be transferred via this channel in any direction.

The main feature of using web-services in SOA is that in order to perform a specific task of a service it is necessary to call a method of this service with a well-defined interface. And web-services do not know anything beforehand about an application, which calls them and the application does not know how web-services perform their task [5].

Consequently, SOA can be considered as a way to build GRID-systems, which enable to unite heterogeneous resources by combining loosely-coupled and interacting resources. These services will interact on the basis of the specific cross platform interface. The interface definition hides the language-dependent implementation of a web-service.

The basis of a service is a set of endpoints, which provide clients with some useful features. An endpoint is a network resource, you can send messages to. In order to use provided possibilities a client sends messages to endpoints in that format which is described by a contract between a client and a service. Services wait for receiving messages to endpoint's address, assuming that the messages will be recorded in a specified format.

Thus, GRID-systems based on SOA will be independent of hardware and development technologies (such as .NET, Java, OpenCL, CUDA, etc.) that facilitates the reuse of components.

XML (eXtensible Markup Language) – a text data format developed for the storage and transmission of information in a structured way, which is both good for reading by a man and suitable for automatic analysis. The main advantage of XML is expansibility by creating dictionaries. The XML Schema specification or XML Schema Definition (XSD) allows describing not only its own dictionary, but any XML-document [3].

WSDL (Web Services Description Language) – a language to describe web-services and access to them, based on XML. For creating GRID-systems it is necessary to have the possibility to describe the communication interactions of its component parts in a certain structured format. So WSDL is used for that [4, 8].

A WSDL-document describes services as a set of endpoints. In order to simplify the reuse, network access points and also messages used by them for exchanging data, are described without binding to concrete network protocols and data formats.

The following set of elements is used to describe network services in a WSDL document:

- Types – a container for describing a data type, which uses some type system.
- Messages – an abstract, typed data description, by which communication is implemented.
- Operation – an abstract description of a specific action supported by the service;
- Port type – an abstract set of operations supported by one or more endpoints;
- Binding – a concrete protocol and data format specification for a particular port type.
- Service – a set of related endpoints.

It should be noted that WSDL is not a new type definition language, but this is a way to describe a format of messages that supports XML Schema as a basic system of types. In practice WSDL, binding and extension mechanisms are implemented as XML dictionaries with using XML Schema.

SOAP (Simple Object Access Protocol) – a simple object access protocol. First of all, the protocol is used to exchange structured messages in a distributed computing system and also to remote calling of procedures or accessing to web-services. SOAP is one of the standards, which technologies of web-services are based on [4, 8].

SOAP gives the possibility to exchange structured messages between applications regardless of a platform, which they work on, and a programming language. More often SOAP is used over HTTP protocol, supported by all internet browsers and servers. SOAP can use other application-layer protocols, such as SMTP, FTP, HTTPS, etc.

Starting from the technologies described above, underlying SOA, it is possible to highlight the following advantages of using SOA for building distributed computing systems:

- Using of SOA allows creating a system in which a service client and the realization of this service have a list of available methods and a specific data structure.
- The client only knows the interfaces of this service, names and types of input and returned values by the function of the service. There is no other dependency.
- Platforms and programming languages can be different at the client and at the service.
- Separation of tasks. Realization of the service and the client can be performed by different developers.

Windows Communication Foundation. Having analyzed all the requirements to the distributed computing system and the most widespread at the present time other analogs, the authors, choosing a software platform, have stopped on .NET and the software framework Windows Communication Foundation (WCF) from Microsoft [8, 9]. The use of the OpenCL standard is expected in realization of the modules of the system. It will ensure to get maximum independence from the architecture of the client PC and at the same time it will allow the transparent use of GPU and FPGA for a significant acceleration of calculations.

WCF is the simplest way to provide and consume services on the Microsoft platform. Using WCF, developers can focus on applications rather than on communication protocols.

It is known that no technology can meet all the needs especially for complex scripts of distributed systems. However, WCF offers a single stack, which includes the following distributed technologies such as: ASP.NET webservices (ASMX); .NET Remoting; EnterpriseServices; WebServicesEnhancements (WSE); COM+; Microsoft MessageQueuing (MSMQ) [8]. This unified stack allows service providers to easily offer that maintenance with different parameters of settings and technical requirements, which are compatible with different service consumers. These parameters can include mechanisms of communication and also of encoding, security, transaction support, etc. The determination of one service with many sets of parameters of settings could be easily done by offering different endpoints to that service. Endpoint relates to a construction which is presented to allow messages to be sent and/or received. Each of these endpoints is available through a unique address (URI) and basic settings that can be set up by service providers or in code or in configuration files. Moreover, WCF provides higher performance in comparison with any other distributed technologies that are offered by .NET Framework. High performance is a key requirement to all computer systems, especially in such as GRID-systems.

III. Proposed Architecture

The system for high-performance calculations, developed according to SOA, was realized as a set of web-services interacting via SOAP protocol. The protocol is used for realization of remote procedure call (RPC) and for exchange of random messages between system nodes in WSDL format [7].

Heterogeneous computing resources (CPU, GPU and FPGA) of the clients are presented as a service and implemented as a set of web-services, i. e. they represent identified by a web-address software system with standardized interfaces. Computing resources of the distributed system, realized as web-services, can interact with each other, with the server and with third-party applications using messages, based on the SOAP protocol.

Our project does not involve the design and implementation of a new framework based on SOA. Instead, it shows a method of providing computing resources with multiple configurable WCF services which can overcome the limitations of traditional XML web-services.

In general, the architecture of the developed hybrid reconfigurable system for high-performance computing is shown in Fig. 2. Any computational problem solving on the presented system consists of the client and server side. The main computing burden is placed on clients who provide their computing resources as a WCF-service. The computing nodes of the clients can be general-purpose processors (further CPU), graphics processors (GPU) and also special hardware accelerators based on field programmable gate array (FPGA). The client side was built on the application deployment technology ClickOnce. Using ClickOnce for deployment of can help to solve other problems of the systems, such as: the complexity in updating the client part of the system, the security of running applications and the need administrator rights to install the client side.

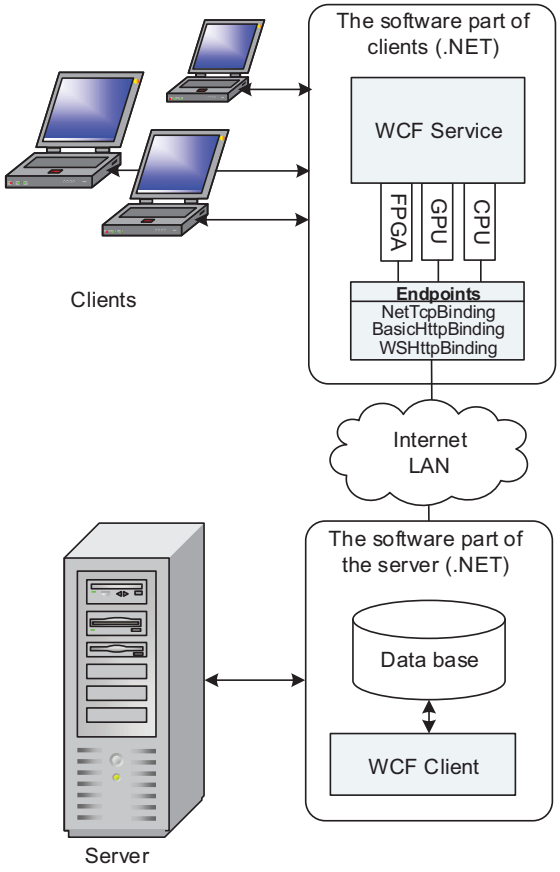


Fig. 2. Proposed Architecture.

The server side is a WCF-client, which refers to the computing resources of the clients over local network or the Internet. The main functions of the server part of the developed hybrid system are: evaluation of client side performance and connection speed, monitoring clients and a computing process, finishing processing of the results of computing received from the clients.

As shown in Fig. 2, the system provides a number of services, which are used as interfaces. Each of the available services has a lot of endpoints to meet different needs. For

example, different parameters of security or transport protocols, methods of serialization and encoding parameters. For basic scripts three main types of WCF binding were used [8]:

- NetTcpBinding: This option is suitable for homogeneous (.NET to .NET) connections over LAN. Although this variant is less compatible with other standards, this is the fastest type of bindings, which is in WCF. NetTcpBinding uses TCP as a communication protocol and binary serialization for data transferred between the client and server.
- BasicHttpBinding: This type of binding provides access to endpoints, as if they were traditional web-services on XML. This binding is slower than NetTcpBinding, because for exchanging simple text messages and UTF-8 encoding are used. Nevertheless, BasicHttpBinding can be set up for using more effective data formats, such as binary, etc. One more advantage of endpoints based on BasicHttpBinding is compatibility with different clients (both homogeneous and heterogeneous).
- WSHttpBinding: This parameter uses HTTP protocol and WS-* specification, so that it is similar to BasicHttpBinding, but with more advanced features. It includes support of secure and reliable sessions. This option is recommended for Internet users (both clients and executors), to guarantee safety and reliability.

Metadata in WCF contains information clearly describing how to address to the service. Having requested metadata from a running service a client can learn about its endpoints and required messages formats. While projecting stage, clients send this request as a message determined in WSMetadataExchange standard and receive a WSDL–document as a response. A client can use the document for proxy-class and configurable file generation, which subsequently will be used to access service while at runtime.

All settings and configurations in WCF can be prescribed either in code or in configuration files. The second method is used for the developed system, because it allows administrators to determine, set up, and switch on/off endpoints easily and flexibly "on-the-fly", without having to modify or recompile the source code.

The interfaces of the components in SOA encapsulate the details of realization (OS, platform, programming language) from other components, thereby providing combining and multiple using components for building complex distributed systems, ensuring the independency from used platforms and developing tools, promoting scalability and controllability of created systems.

IV. Experimental Studies

For experimental studies on the proposed platform Deep Learning algorithms were chosen, because this is an increasingly popular topic in the field of artificial intelligence and one of the most sought fields in computer science [10]. The use of Deep Learning algorithms allowed creating systems, which are able to recognize noisy and distorted objects in photos; to make complex predictions about possible future behavior of a person in a video stream; to analyze text information and etc. At the same time implementation of these algorithms requires high-performance computing.

Currently the TensorFlow library is widely used by developers and researchers in the Deep Learning field [11]. TensorFlow is a new open source library from Google for creating machine learning systems. Google uses this library for: search engines, speech recognition, image search and etc.

For experimental studies the following datasets (table 1) were used, which are most often used when comparing Deep Learning algorithms [10, 11].

MNIST – a set of pictures of hand-written numbers.

CIFAR-10 – a set of colored pictures of 10 different classes.

Molecular – a set of molecular activity of substances for drug development.

In the last three columns of the table the parameters of the implemented deep neural networks are presented.

TABLE I. Datasets

Datasets	Size	Total Parameters	Hidden layers	Hidden Neurons
MNIST	60K x 784	1.8M	2	512
CIFAR-10	60K x 1024	4.1M	2	1024
Molecular	150K x 287	14.2M	3	2048

Figure 3 shows the results of learning these deep neural networks on the developed platform.

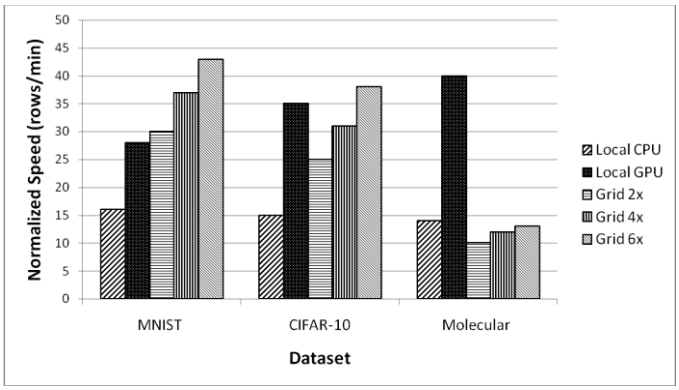


Fig. 3. Results of experimental studies.

For studies networks were built with 2, 4 and 6 clients. From these results it can be seen that the proposed system has the same problem as other similar systems – the greater the volume of data transmitted over a network is, the lower the effect of its use is.

To overcome this problem, it is necessary to consider the size of transmitted data over a network and choose appropriate algorithms.

The stochastic gradient descent (SGD) is a commonly used method of optimization for learning deep neural-networks. The

classic algorithm SGD is generally subsequent, that makes inexpedient to use this for large sets of data, because it takes a lot of time to transmit data in a serial fashion.

The asynchronous stochastic gradient descent was applied for distributed learning of the neural networks. The main feature of the algorithm is that there is its own copy of the data and teaching model on every client. This significantly reduces network traffic. Figure 4 shows the results of the experiments.

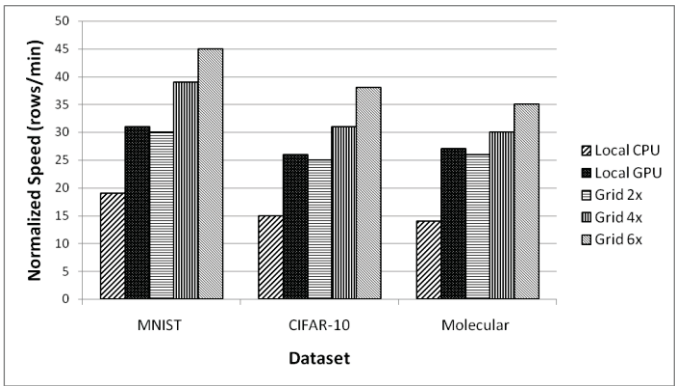


Fig. 4. Results of experimental studies with SGD.

From the diagram it can be seen that taken measures allowed getting the efficiency in computing on the developed platform. I. e. the more clients are in the network, the faster the computing speed is.

The main advantages of uniting distributed computing resources as a set of web-services are:

- Cross-platform. Web-services allow providing interaction of heterogeneous computing resources regardless of a platform. For example, a Windows client on the .NET platform is able to interact with Java-server, working on Linux.
- The use of web-services, which are based on open standards and protocols, can reduce the complexity of the developing and debugging developed systems.
- The use of internet-protocols (also with encoding) allows providing the interaction of heterogeneous computing resources through the firewall. This is a significant advantage compared to such systems as CORBA, DCOM or Java RMI.
- Ease of system deployment.
- The possibility of dynamic adding and cutting off both clients and separate computing units of a client during computing.

V. Conclusion

The proposed GRID-architecture based on SOA can be easily used in all fields requiring high-performance computing. This includes (but not limited by them): biomedical engineering, complex mathematical experiments, finance and economics analysis, numerical modeling, graphics and video rendering and etc [13].

The developed architecture offers big flexibility, ease and compatibility needed to provide the integration between network resources. The platform is not intended to replace other GRID-systems. Instead, the platform gives the possibility to use different platforms (CPU, GPU, FPGA) in unified computing process on the basis of SOA, and the use of OpenCL will allow creating a heterogeneous computing environment within a network. Using the technology of distributed computing and involving unused resources of PC, the system allows decreasing the cost of calculations per unit of performance.

The paper was supported by RFBR projects: 16-37-60016, 14-07-00816.

References

[1] Barbosa, J.G., Dutra, I. Grid computing: Techniques and future prospects (2015) Grid Computing: Techniques and Future Prospects, pp. 1-213.

[2] Shrihari, S.V., Pawan, K., Vishnu, M., Mahajan, S. Technological aspects of grid computing (2015) 2nd International Conference on Electronics and Communication Systems, ICECS 2015, art. no. 7124829, pp. 1463-1468.

[3] Raid A., Hassan A. Design of SOA-based Grid Computing with Enterprise Service Bus. International Journal on Advances in Information Sciences and Service Sciences. – 2010. – Vol. 2. – pp. 71-82.

[4] B.A.D. Gbaffonou, J. Lapalme, R. Champagne, Service-Oriented Architecture: A Mapping Study, 2015 International Conference on Enterprise Systems (ES), Basel, 2015, pp. 33-42.

[5] Amirpour, M., Harounabadi, A., Mirabedini, S.J. Service oriented architecture assessment based on software components (2016) Decision Science Letters, 5 (1), pp. 109-118.

[6] Nurcan, S., Schmidt, R. Service oriented enterprise architecture for enterprise engineering (2015) Proceedings of the 2015 IEEE 19th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations, EDOCW 2015, art. no. 7310675, pp. 88-90.

[7] M.V. Lyashov, A.N. Bereza, J.V. Alekseenko and L.M.L. Blanco, The hybrid reconfigurable system for high-performance computing. Application of Information and Communication Technologies (AICT), 2015 9th International Conference on, Rostov on Don, 2015, pp. 258-262.

[8] Cibraro P. Professional WCF 4: Windows Communication Foundation with .NET 4. – Microsoft Press, 2010.

[9] Stephens R. Fundamentals of Microsoft .NET Programming. – Microsoft Press, 2011.

[10] Li Deng and Dong Yu (2014), "Deep Learning: Methods and Applications", Foundations and Trends in Signal Processing: Vol. 7: No. 3–4, pp 197-387.

[11] Dean J., Rajat M., et al. "TensorFlow: Large-scale machine learning on heterogeneous systems. TensorFlow.org. Google Research. Retrieved 10 November 2015.

[12] Dean J., Corrado G., et al. Large Scale Distributed Deep Networks. NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada, United States, December, 2012, pp 1232--1240.

[13] V. Beglyarov, A. Bereza, L. Blanco Hybrid pareto-evolutionary algorithm for solving mathematical models of high dimensional electronic circuits (HPEA), Design & Test Symposium, 2013 East-West, Rostov-on-Don, 2013, pp. 1-4.

Using Model-Based Methods for Embedded Software Development

S.Jalilian, T.Boroumandnejad
Satellite Research Institute, Iranian Space Research Center
Tehran, Iran
shjalilian@gmail.com

Kazimov Tofig H.
Azerbaijan National Academy of Sciences Institute of Information Technology
Baku, Azerbaijan
tofig@mail.ru

***Abstract*—today, using of embedded systems and complexity of these systems are increased. As a result of this complexity, development of this systems based on model-based process is needed. But the Standard-UML based tools can't be used because there is significant difference between general–purpose software and embedded software. Therefore, the integrated tools must be used for covering non functional, scheduling characteristics of these systems. In this paper we will review three methods that are used for developing embedded software.**

***Index Terms*—Embedded systems, Model-based Development, UML, code generation.**

I. Introduction

Today, Embedded Computer systems are used in a wide range of systems from airbag control systems to automotive and airplanes. The design of these systems is a complex task because these systems include hardware and software components that must be highly optimized for the application needs special in real time embedded systems. Also, the increasing complexity of these systems has impact on its development life cycle. Because of these reasons, system development methodologies must be used in order to manage the team size, the requirement and project's constraints. Furthermore, embedded software is brain and intelligent part of these systems that controls the behavior of embedded systems, so it plays an important role in this type of systems due to its flexibility [4], [5].

Nevertheless to say, Model-based development methods that uses only in classical software engineering such as UML cannot be applied for creating embedded software and must be integrated with another tools. Because the development process of general-purpose software and embedded software is differ significantly. The embedded software is deal with non-functional issues such as performance, reliability, safety, timeline and concurrency in a physical context. Therefore the integrated methodology must be used for covering these issues [1], [7].

Standard UML-based modeling has several shortcomings. For example while deployment diagram attempt to provide a modeling capability for supporting component and system configuration but it is weak because it can't map component to processors, component to threads and so on. Also, UML cannot express system behavior under fault conditions or can't define concurrent behavior of component that is important in the embedded systems [2],[6].

Because of reasons that mentioned above, hybrid methodology or tools must be used for making embedded software. In this paper these kinds of methods will be reviewed.

II. Embedded Software Design Methodology

Some useful methods and techniques for embedded software development is explained in this section. These methods are as follow:

- *Embedded System Modeling Language(ESML)*
- *An Integrative approach with UML and Simulink*
- *Template-based code generation*

A. Embedded System Modeling Language(ESML)

ESML is a system modeling language for embedded systems that is used for integrated modeling of software component, component interaction, hardware configurations, scheduling policies and other aspects relevant for system developer. Procedure invocation via component and event propagation through public/subscribe are two mechanisms that is used in this language. These two mechanisms which is used in ESML are combined in a "push-directed-pull" interaction pattern. As illustrated in fig.1, in this mode of operation, subscribers are notified by a publisher that data is available for retrieving. After that, the subscribers will be triggered and a call "back" will be send to their facets by their receptacles. This technique provides more adaptive behavior to support non functional characteristic of embedded systems [2].

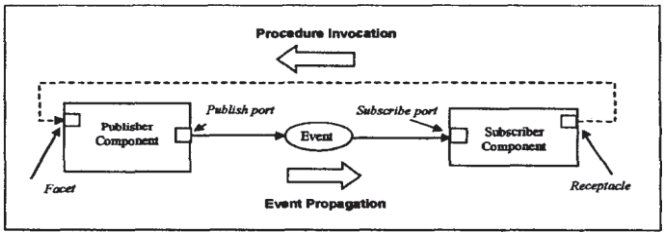


Fig. 1. Model of computation in ESML

Nevertheless, ESML isn't a replacement for UML or implementation language. It is in conjunction with these tools for designing embedded software. Figure 2 illustrates how this is done. In development process of this kind of software, components are designed by Rational Rose and header files are generated through its code generator. UML model must be annotated with "tagged values". Next, annotated model will be import in ESML modeling environment and ESML component will be construct by interprets the annotations. Now, the modeler can construct other models for example interaction diagrams and scheduler diagram [2].

B. An Integrative approach with UML and Simulink

This approach tries to integration of exiting artifact like a C legacy code and the migration to a model-based development. Specially, it shows when modeled and non-modeled artifacts exist how UML & Simulink is used for functional specification and automatic code generation for control device. Because of using UML for modeling the architecture, simulation and integration aspects of embedded software artifacts, this method defines an extensible UML profile. This profile has different stereotypes such as Matlab, LegacyClass to reference several artifacts such as Matlab function, legacy C code. The process of development steps is given in Fig. 3. A system developer designs new capability based on exiting artifacts and models. Next, this artifact must be integrated into the architecture view of UML model. The legacy code, network models and real time operating system (RTOS) models are directly included in the UML via references. Finally, code will be generated automatically. This approach uses a stepwise migration model-based development due to the fact that industries can't switch-over immediately from C-code to model-based development.

This method is useful for existing embedded systems that has huge amount of legacy C-code and want switch to model-based development as soon as possible [8].

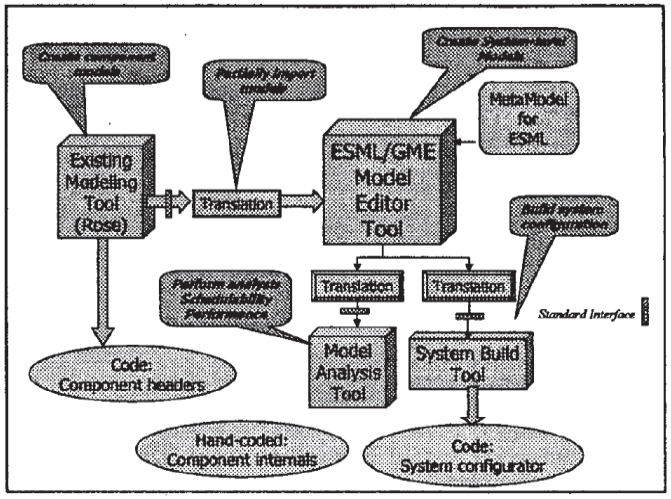


Fig. 2. The process of using ESML

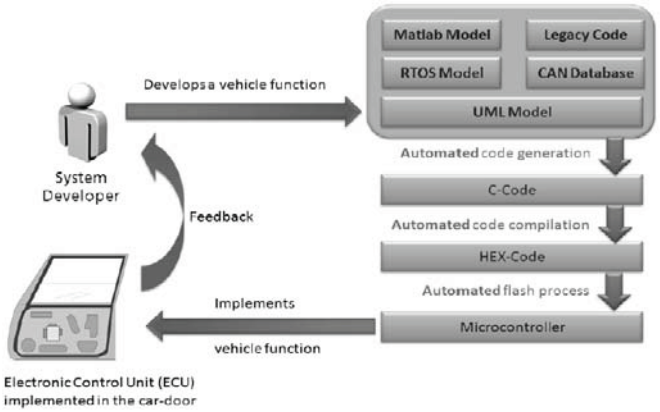


Fig. 3. Process of automated development step

C. Template-based code generation

There are many repetitive problems in the context of embedded systems like process management, scheduling and so on. Solutions for these problems exist but the heterogeneity of embedded software doesn't allow reusing components. This approach, as shown in Fig. 4, uses a technique like preprocessor macros for solving this problem which is called template-based development. This method use application-independent templates instead of generating machine code from an application model or attempt to reuse precompiled libraries. One of the advantages of this approach is that it can be adapted automatically to the application requirements on the base of the model. Also flexibility is big advantage of this method [3].

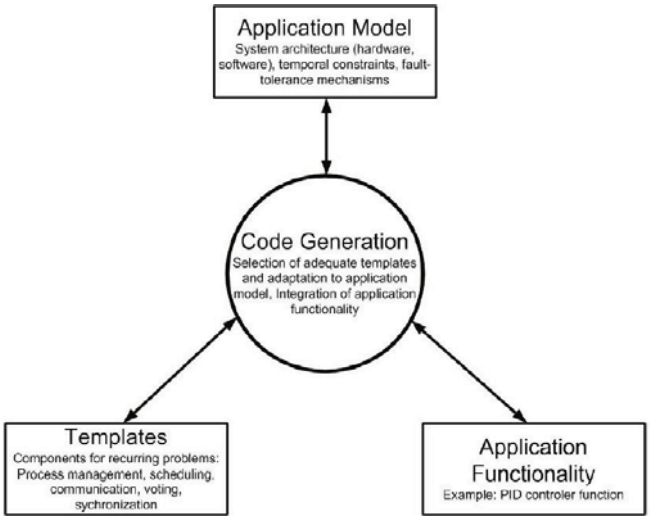


Fig. 4. Code Generation Architecture

The code generation process steps for developing embedded software as shown in Fig. 5 is as bellow:

- First, developer specifies the model of system. The model must be included the information about the software architecture, the tasks of application and timing constraints.
- The information about non-functional aspects like OS, programming language, performance is also needed for selecting the right template.
- In the next step, the correctness of model will be check by model parser and the information will be store into database.
- The second step for developer is to implement the application code like control function. This task can be performed by other model-based tools such as Matlab/Simulink.
- At the last step, the parser combines the generated code with implemented code by the developer. The result is source code that can be compiled and executed on any platform [3].

III. Conclusion

As we mentioned in this paper, to cope with complexity, uncertainly and encompassing heterogeneity of embedded systems, model-based approaches are more convenient than traditional approaches. General-purpose model-based development is an ongoing approach for developing embedded software. It can't cover all aspects of embedded software such as non-functional characteristics; therefore new approaches such as integrative or template-based approach have been provided to cope with UML weaknesses. In this paper 3 approaches were discussed. The advantages of these approaches are high-level modeling of the software, possibility facilitating to achieve early simulation and formal verification based on model and flexibility in source code generating. Main problem in this area is tools support. Although UML is widely supported by software development tools, tool facilities for developing embedded software are not as rich as UML. There are some commercial tools such as IBM Rhapsody or Eclipse model-based development tools which support UML for embedded software development. They help developer to cope with software complexity but to cover non-functional aspects of embedded software development, advanced model-based methods must be included.

In the future, we are going to inspect aspect-oriented approaches for developing embedded software, its advantages and disadvantages.

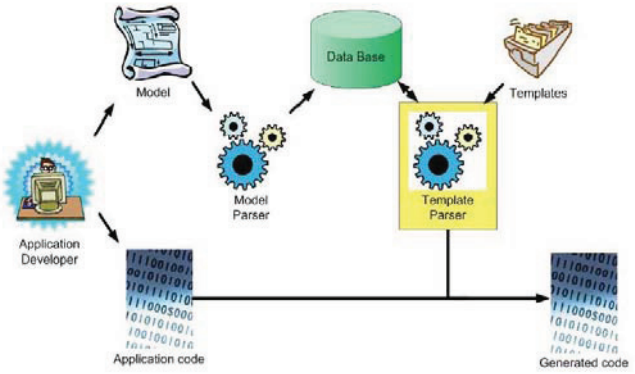


Fig. 5. Code Generation Step

References

[1] B. Dion, Th. Le Sergent, B. Martin, "Model-based development for Time-Triggered Architectures," the 23rd digital Avionics Conference, vol. 2, pp.6. D.3-6.1.7, IEEE, Oct 2004.

[2] G. Karsai, S. Neema, B. Abbott, D.Sharp, "A Modeling Language and its supporting Tools for Avionics Systems", the 21rd digital Avionics Conference, vol.1, pp.6A3.1-6A3.13, IEEE , 2002.

[3] Ch. Buckl, A. Knoll, and G. Schrott, "Model-Based development of Fault-tolerant Embedded Software", second international symposium on leveraging applications of formal methods, IEEE, pp. 103–110, Nov 2006.

[4] L. Cordeiro, C. Mar, E. Valentin, F.Cruz,"A Platform-Based Software Design Methodology for Embedded Control System: An Agile Toolkit,"15th annual IEEE International Conference and workshop on the Engineering of Computer Based Systems, pp. 408–417, Apr 2008.

[5] M. A. Wehrmeister, C. E. Pereira, L. B. Becker, "Object-Oriented to the Development of Embedded Real-Time Systems," 8th annual IEEE International symposium on object-oriented Real-Time Distributed Computing, pp.125-128, May 2005.

[6] M. A. Wehrmeister, C. E. Pereira, F. Rammig, "Aspect-Oriented Model-Driven Engineering for Embedded Systems applied to Automation Systems," IEEE Transactions on Industrial Informatics, vol.pp, jan 2013 .

[7] Ph. Conmy, R. F. Paige, "Challenges when using Model Driven Architecture in the development of Safety Critical Software,"4th international workshop on Model-Based Methodologies for pervasive and embedded software, IEEE, vol. 2, pp. 127–136, March 2007.

[8] T. Farkas, C. Neumann, A. Hinnerichs, "An Integrative Approach for Embedded Software Design with UML and Simulink,"33rd Annual IEEE International Conference on Computer and Software, vol. 2, pp.516-521, July 2009.

Computer Modeling of Spatial Structure of Dermorphin

G. A. Akverdieva
Institute for Physical Problems, Baku State University,
Baku, Azerbaijan,
Hagverdigulnara@gmail.com

***Abstract*—The possibilities of the theoretical study of the spatial structure of biomolecules using a computer technology advances are demonstrated on the example of the opioid peptide dermorphin. The comparative conformational analysis of dermorphin and its active analogs have been investigated by molecular mechanics, molecular dynamics and semiempirical quantum-chemical methods. The conducted research has allowed to assess the bioactive conformation and the relative positions of pharmacophore groups of this peptide molecule.**

Keywords—dermorphin; dermorphin analogs; conformational analysis; spatial structure; bioactive conformation; pharmacophore

I. INTRODUCTION

The activity of biomolecule is directly related to their particular spatial structure with its characteristic electronic parameters and dynamical conformational properties that play an important role in the enzyme-substrate interactions. Currently, using the different theoretical calculation methods, the recent advances in computer technology, including programms with a graphical representation of the spatial structures allow researchers to construct the various models of biomolecules. In this work the possibilities of a theoretical study of the spatial molecular structure with the using of the computer programms are demonstrated on the example of the bioactive molecule—dermorphin. Dermorphin is linear peptide, isolated from a skin of the Southern-American frogs belonging to family Phyllomedusa, has high relationship and selectivity to μ-opioid binding places, but possess certain affinity to receptors δ- type [1-5]. Dermorphin shows strong analgetic influence as on central and peripheral nervous systems that is reflected in various physiological functions of an organism: cardiovascular, immune, anti-inflammatory, termoregulatory and etc. This peptide is about 30-40 times more potent than morphine [6]. In order to put a step forward in the desigh of generation peptidomimetics it is necessary to assess the bioactive conformation of dermorphin and propose a pharmacophore of its interaction to the opioid receptor. In order to investigate the geometry characteristics of the bioactive conformation of dermorphin the conformational profiles of dermorphin and its active tetrapeptide analogs (1-4) dermorphin, [D-Arg2] (1-4) dermorphin and [D-Arg2, Beta-Ala] (1-4) dermorphin [7] were studied and compared pairwisely. The received results were detailed by molecular dynamics simulations and quantum-chemical calculations. The conformational-electronic properties common to these molecules were found. On the basis received results and data of SAR studies the bioactive conformation of dermorphin was assessed and the model of pharmacophore for its interaction with the opioid receptors was proposed.

II. METHODS

Conformational energy calculations were made with an IBM computer using version of ECEPP (Empirical Conformational Energy Program for Peptides) [8,9]. The proqram was developed from the matrix method principle of Hermans and Ferro [10]. The investigations were carried out within molecular mechanics framework as described in [11,12]. The backbone was described by the "shape" symbols *e* and *f* corresponing to extended and folded configuration of virtual bonds C^{α}_{i}- C^{α}_{i+1}- C^{α}_{i+2}- C^{α}_{i+3}, respectively. The nomenclature and conventions adopted are those recommended by IUPAC-IUB [13]. The molecular dynamics of dermorphin was spent with use of force field AMBER in the temperature interval 293-313K with step 5K during 10 nanoseconds by means [14]. Procedure solvation with application of model of water in the set spherical volume TIP4P [15] has been used. The quantum- chemical calculations of these molecule was conducted by method CNDO [16], used the demonstration version of software package HyperChem (http://www.hyper.com).

III. CALCULATIONS AND RESULTS

A. Conformational profiles of dermorphin and its analogs

In first step the conformational analysis of dermorphin was carried out through a fragmental calculation on the basis of stable states of monopeptides. The starting structural variants for dermorphin yielded more than 400 conformations belonging to 32 shapes of peptide skeleton, 31 dihedral angles were exposed to rotation. The presence of proline residue in peptide sequence causes of the bend of peptide chain on the C-terminal tripeptide fragment and reduces the number of sterically allowed structural types for this part of molecule: the conformations only *ee* and *ef* shapes are possible for it. Despite rather limited conformational possibilities of the N-terminal tetrapeptide fragment of this molecule, thanks to presence residue Gly in its sequence the effective interactions are realised between the terminal segments of molecule. The low-energy conformations were clustered from the values of the rmsd between each pair of structures into 12 shapes (Table I). It was found the optimal structures of dermorphin ($E_{rel.}$= 0.0 kcal/mol, 0.4 kcal/mol, 0.5 kcal/mol, 0.7 kcal/mol and 0.9 kcal/mol) are the folded or semifolded conformations stabilized by hydrogen bonds between the atoms of peptide skeleton. These results well coordinate to the data of spectroscopic works and theoretical researches [17-24], that offer as the preferable the semifolded or the folded forms of peptide skeleton of dermorphin molecule.

TABLE I. LOW-ENEGY CONFORMATIONS OF DERMORPHIN

No	Shape	Backbone form	$E_{nb.}$	$E_{el.}$	$E_{tors.}$	$E_{rel.}$
1	ffffef	LPBLLRR	-24.6	1.3	4.2	0
2	feefee	LLBRBBR	-24.5	3.1	2.7	0.4
3	feefef	LLBRBRR	-24.2	3.2	2.4	0.5
4	ffffee	LPRRBBR	-23.8	2.4	3.0	0.7
5	feeeef	BLBRBRR	-23.7	3.2	2.3	0.9
6	efffef	RPRRBRR	-24.0	3.8	2.2	1.0
7	fefeee	LLBLBBR	-24.1	3.0	3.1	1.1
8	efffee	RPRRBBR	-24.3	3.9	2.6	1.2
9	ffefef	LPBRBRR	-23.1	1.8	3.4	1.3
10	fffeee	BPBLBBR	-23.7	1.5	4.5	1.4
11	ffefee	LPBRBBR	-22.8	1.5	4.0	1.7
12	feffef	LLRRBRR	-23.8	3.6	2.8	1.7

The physiologically active N-terminal tetrapeptid of dermorphin which defines the specificity of its interactions with opiate receptor has regular structure - α-spiral in two optimal conformations and semifolded structure in others optimal structures of dermorphin. As in this segment of the molecule the majority of the residues have folded backbone structure, the interactions of Tyr1 residue with the subsequent residues Ala2 and Phe3 by the contributions from -1.9 to -2.9 kcal/mol and from -2.2 to -5.2 kcal/mol, respectively and also the interactions of Tyr5 residue with the subsequent residues Pro6 and Ser7 by the contributions from -3.7 to-4.7 kcal/mol and from -2.9 to -3.8 kcal/mol, respectively, are effective. In the global conformation, belonging to *ffffef* shape of peptide skeleton, the residues Tyr1 and Phe3 with the aromatic side chains approach in space and form quasicyclic structure with formation of two hydrogen bonds between amino and carboxyl groups of these residues. Therefore, in this conformation the interactions of mentioned residues are stronger, than in other structures of molecule. This structure is characterized also by more effective contribution of interactions of residues D-Ala2 and Pro6 (-2.1 kcal/mol). In the conformations 2, 3 and 5 the dispersion contacts of residue Phe3 both with Tyr5 and with Pro6 are effective owing to their spatial approach and the interactions between atoms of aromatic side chains. These structures are characterized by existence of hydrogen bond (Phe3) CO…HN(Tyr5). For conformation 2 the hydrogen bond between the atoms of the main chain of the residues in fifth and seventh positions, namely (Tyr5) CO... HN(Ser7) is characteristic too, but thus HN group of Ser7 participates also in formation of hydrogen bond with O atom of own side chain. Conformation 4 (*ffffee*-shape) represents the particular interest. In this conformation the form of peptide chain of the N-terminal tetrapeptide part is similar to global conformations and the form of peptide chain of C -terminal tripeptide part is similar to conformation 2. Therefore, the hydrogen bonds, characteristic for both specified conformations are realized in this structure. But distinctive features of conformation 4 are effective tetrapeptide interaction and formation of hydrogen bond (Tyr1) NH...OC(Gly4). As seen from the results, this conformation is distinguished from others by saturation of hydrogen bonds. In this structure the phenolic ring of the residue Tyr1 hangs over a conformationally rigid ring of the proline residue in the parallel position to it, and as result the interaction between these residues is found to be -4.1 kcal/mol. Here, α-amino group and the carbonyl residue Tyr5 are spatially close together, forming the salt bridge. The interactions of residue Tyr1 with Gly4 (-2.3 kcal/mol) and also with Tyr5 (-2.8 kcal/mol) bring the appreciable contributions to stabilization of this structure. Conformation 5 (feeeef-shape) is characterized by folded form of the N- and C-terminal dipeptide parts and by completely extended form of central part of the molecule. For this structure the intraresidual hydrogen bonds for residues Tyr1, Phe3, Tyr5 and also hydrogen bond between carboxyl group of main chain of Ser7 and H-atoms of the terminal amid group NH2 are characteristic. All optimal structures of dermorphin are characterised by folded of conformation of the dipeptide segment Tyr1-DAla2 that is reason of strong dipeptide interaction. Apparently, the specified minimum structural requirement is important for physiological activity of dermorphin molecule, as it is necessary for protection of this peptide bond from splitting action by enzymes in the process of metabolism of peptide. Apparently, from the Table 4, thanks to presence DAla and Pro in peptide chain a number of preferable structures have turns on the terminal parts of molecule. The particular interest is represented by the structures belonging *feefee* and *feefef* shapes of peptide skeleton with the relative energy, equal 0.4 and 0.5 kcal/mol. In these conformations β - turn is formed on the Gly4-Pro5 segment, therefore in them the aromatic side chain of residue Phe effectively interacts with atoms of the subsequent residues and the residue is shaden from solvent amid proton of Gly4.

In a further step within the framework of mechanical model, a conformational dynamics of the side chains of the amino acid residues of dermorphin peptide was investigated. For this purpose, a series of conformational maps, or sections of the potential surfaces were constructed over the dihedral angles for amino acid residues of the molecule [25]. Thus, the dihedral angles around the N-C^{α}, C^{α}-C^{β}, C^{β}-C^{γ}, C^{γ}-C^{δ} bonds of the side chains were varied, and the backbone of the peptide residues was fixed in accordance with the coordinates of the atoms of optimal structures of the peptide molecule. Thus, the analysis of the conformational maps has revealed the degree of mobility of the functional residues. It is established that the dynamics of Phe3 and Tyr5 is limited owing to realization of effective stabilizing interactions of its aromatic side chains atoms with atoms of other residues of the molecule. So, the conformational freedom of Phe3 residue is limited by effective interactions with Tyr1, and conformational freedom of Tyr5 residue is limited by effective interaction of its aromatic ring with the atoms of the rigidly fixed prolin residue in sixth position. For this reason, the conformational energy is rather sensitive to orientation of the side chains of specified aminoacid residues. The side chains of Tyr1 and Ser7 are characterized by the large conformational dynamics. Evidently, such mobility of the terminal residues is necessary for realization of interactions with atoms of environment or receptor. As seen from the presented data, the conformational dynamics of any segment of a molecule is defined by efficiency of its interaction with other segments of molecule. The rotary possibilities of side chains of the residues are various among the considered optimal structures dermorphin,

but because of presence of the hydrogen bonds the conformational balance is between them.

In the subsequent investigation step molecular dynamics of dermorphin has been studied and the most probable conformational states of this molecule are defined in vicinities of local minima at physiological temperature. As a result of simulation process the ranges of change of dihedral angles and distances between atoms of aminoacid residues are defined. The quantitative estimation of distances between atoms shows that at modelling the dispersion contacts between the aminoacid residues in the first and fourth positions of peptide chain are invariable. The N-terminal tetrapeptide fragment of the molecule preserves the folded structure throughout molecular dynamics simulation, the distance between C^{α}-atoms of the specified residues does not exceed 7Å during simulation. The structure of this part of molecule upon termination of modelling process practically does not differ from the initial stage. It is possible to assume the conformational stability of specified fragment plays important role in functional activity of dermorphin molecule and defines the specificity of its interaction with the receptor. Because of mobility of the C-terminal tripeptide fraqment of investigated molecule the distance between the atoms of the side chains of the residues Phe3 and Tyr5 is approximated to 2.8 Å during simulation. As result, the hydrogen bond between atoms of hydroxyl group of residue Tyr5 and the main chain of residue Phe3 is formed. Such conformational behavior may be due presence glycine in fourth position of peptide chain that is characterized by conformational freedom of rotation around C^{α}-$C^{/}$ and $C^{/}$-N bonds, giving certain flexibility to peptide chain. One can conclude, as the system is in an equilibrium state, the possible conformational changes are reversible. The cited data though testifies about stability of optimal conformations of dermorphin, but assumes also the probability of realizations of each of them depending on polarity of environment.

It was established, that the optimal conformations of dermorphin has low values of dipole moment and of energy $E_{HOMO,}$ characterizing its electro-donor properties and the energy of activation. It is possible to draw a conclusion that the low-energy structures of dermorphin have the insignificant chemical reactionary ability and the weak elektrodonor properties. It was found the conformational transitions between optimal structures of dermorphin are resulted by fluctuations of density of a charge on certain groups of the atoms, not exceeding 18%. The distinctions of charges on the atoms C-terminal dipeptide residues and amid group are revealed. The distribution of charges on the atoms of the α - amino group and atoms of the side chains of tyrosine and phenylalanine reidues are similar for optimal structures, that may be correlate with their biological activity - ability to participate in ligand-receptor interactions. Two-dimensional contour of electrostatic potential for this molecule is shown in Fig.1. Green and violet colours represent, accordingly, positive and negative charged segments of the molecule. Apparently from the presented figure, there is the zone of high electronic density subjected to attack of the electrophil reagent.

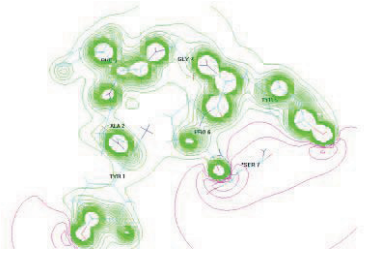


Fig. 1. Two-dimensional contours of electrostatic potential in dermorphin molecule

In a further step the conformational possibilities of the tetrapeptide analogs (1-4) dermorphin, [D-Arg2] (1-4) dermorphin and [D-Arg2, Beta-Ala] (1-4) dermorphin have been investigated on the basis of low-energy states of monopeptides [26, 27]. As may be inferred from these data, the structures of tetrapeptide with the folded and semifolded backbone shapes are the most preferred ones and correspond to stabil states of this sequences. They are representatives of *fee, ffe, fff, fef, eff*--shapes with relative energies in within 0-2 kcal/mol.

B. Bioactive Conformation of demorphin and its analogs

Since dermorphin and its active analogs have similar functions, it is plausible to assume these peptides perform their activity by the same structural domains to the receptor – pharmacophore groups. In order to characterize the bioactive conformation of dermorphin, the subset of optimal conformations within 2 kcal/mol above global conformation found for dermorpin and its analogs were paiwisely compared. Rmsds were calculated for all backbone atoms of the four N-terminal residues of the peptide. Comparison of of the optimal structures of the dermorphin with the set of optimal structures common to its active analogs revealed the conformations with rmsd ranging from 0.09 Å to 1.35 Å. Visual verification of all pairs of conformations superimposed indicated that pair conformation with rmsd values lower 0.8 A may be considered as similar. It was found that an α - amino group and the side chains of the tyrosine and phenylalanine residues are in the identical positions in space from each other in similar structures of dermorphin and its active analogs. Dermorphin molecule forms compact structure, in which the residues Tyr1 and Ser7 point away from the core surface of molecule, that explains abilities of its OH-groups to interact with the environment. As these residues are capable to participate also in effective interresidual interactions, it is possible to assume that they should possess defined conformational dynamics. This conformation was considered the bioactive conformation.

The dihedral angles of the putative bioactive conformation of the demorphin and its active analogs are listed in Table II.

C. Model of pharmacophore for the interaction of dermorphin with the opioid receptors

On the basis of the received results and data of SAR studies the model of opioid pharmacophore for its interaction with opioid receptors was proposed (Fig.2). The proposed model contains five areas which are occupied by pharmacophore elements marked by numbers 1-5. The pharmacophore element 1 represents the α - an amino group, namely, the protonated nitrogen atom participating in electrostatic interaction with negatively charged residue of opioid receptor.

TABLE II. BACKBONE DIHEDRAL ANGLES (IN DEGREES) OF THE PUTATIVE BIOACTIVE CONFORMATION OF DERMORPHIN AND ITS ACTIVE ANALOGS

Dihedral angles	Dermorph	(1-4) dermorphin	[D-Arg2] (1-4) dermorp	[D-Arg2, Beta-Ala] (1-4) dermorphin
φ_1	66	63	78	-110
ψ_1	128	124	163	-66
ω_1	-173	-172	-178	174
φ_2	76	78	75	78
ψ_2	-64	-60	-74	-77
ω_2	167	170	180	179
φ_3	-114	-107	-133	-139
ψ_3	162	170	-57	-61
ω_3	175	180	169	180
φ_4	-73	-77	-93	-91
ψ_4	76	-79	-90	90
ω_4	178			
φ_5	51			
ψ_5	71			
ω_5	-178			
ψ_6	-62			
ω_6	-178			
φ_7	-107			
ψ_7	-61			
ω_7	-179			

Comment: Residues are numbered with respect to the dermorphin sequence

Thus there is also the formation of hydrogen bond with participation of protonated nitrogen atom. Such interaction in turn can lead to rupture of hydrogen bond between the transmembran spirals of opioid receptor AspIII:7... TyrVII:11 that leads to its activation [28]. The pharmacophore element 2 is presented by hydroxyl group of the side chain of residue Tyr1, it participates in the transit of a charge as the donor of electronic density and in the formation of the hydrogen bond. The pharmacophore elements 3 and 4 represent, accordingly, a phenolic ring of residue Tyr1 and an aromatic ring of residue Phe3, they participate in hydrophob interactions. The pharmacophore element 5 - a phenolic ring of the residue Tyr5 can participate in hydrophob interactions or in the formation of hydrogen bond, that defines the selectivity of this ligand. The relative positions of pharmacophore areas of dermorphin molecule are characterized by a set of distances, angles and dihedral angles [29]. The proposed model defines the presence of the similar structural elements of the bioactive molecules participating in the interaction with opioid receptors. It is expected that received results can be used for search of the ligands of opioid receptors as pharmacological preparations with effective action.

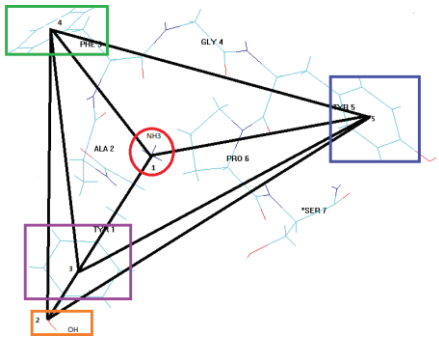


Fig. 2. The model of proposed pharmacophore for the recoqnition of dermorphin to the opioid receptors

IV. CONCLUSIONS

- Optimal structures of a molecule dermorphin are characterised by folded conformation of the dipeptide segment Tyr1-DAla2. Apparently, the specified minimum structural requirement is necessary for protection of this peptide bond from splitting action by enzymes in the process of metabolism of peptide.
- The physiologically active N-terminal tetrapeptide fragment of dermorphin has regular structure - α-spiral in two optimal conformations and semifolded structure in others. Conformational stability of this fragment, apparently, plays important role in functional activity of dermorphin molecule and defines the specificity of its interaction with the opiate receptors.
- The conformational dynamics of Phe3 and Tyr5 is limited owing to realization of effective stabilizing interactions of its aromatic side chains atoms with the atoms of other residues of the molecule.
- The residues Tyr1 and Ser7 are localized on a surface of molecule and characterized by large conformational dynamics.
- The distribution of charges on the atoms of pharmacophore elements in the optimal structures are similar that confirms their identical arrangement in space from each other in all of them and may be correlated with their ability to participate in ligand-receptor interactions.
- On the basis of the received results and SAR studies the bioactive conformation of dermorphin was assessed and the model of pharmacophore for interaction of this molecule with the opioid receptors was proposed.

The reperesented results are helpful for the study of the biology active forms of opioid peptides and for the design of opioid peptiomimetics.

REFERENCES

[1] Montecucchi P.C., De Castiglione R., Piani S., Gozzini L., Erspamer V. "Amino acid composition and sequence of dermorphin, a novel opiate-like peptide from the skin of Phyllomedusa Sauvagei", Int. J. Peptide Protein Res. 17 (1981) 275–283.

[2] Amiche M., Delfour A., Nicolas P. "Opioid peptides from frog ski", EXS. 85(1998) 57–71.

[3] Mignogna G., Severinib C., Simmacoa M., Negric L. et al."Identification and characterization of two dermorphins from skin extracts of the Amazonian frogvPhyllomedusa bicolor", Febs Letters, 302 (1992) 151–154
[4] Choi H., Murray T.F, Aldrich J.V., "Dermorphin-based potential affinity labels for mu-opioid receptors", J Pept Res. 61 (2003) 40-45
[5] Melchiorri P, Negri L.," The dermorphin peptide family", Gen. Pharmacol. 27 (1996) 1099–1107
[6] Broccardo M., Erspamer V., Falconieri Erspamer G. et al., "Pharmacological data on dermorphins, a new class of potent opioid peptides from amphibian skin", Br. J. Pharmacol. 93, (1981) 625–631
[7] Salvadori S., Sarto G., Tomatis R. "Syntesis and pharmacological activity of dermorphin and its N-terminal sequences", Int J Pept Protein Res. 19 (1982) 536-542
[8] Godjayev N.M., Maksumov I.S., Ismailova L.I., "Program of semiempirical calculations of conformations of molecular complexes", J.Struct.Chem., 4(1983) 147-148 (in Russian)
[9] Akverdieva G.A."Improvement of program of calculation of molecular conformation". Theoretical and applied aspects of modern science. Collection of scientific papers of VI nternational Scientific-Practical conference, Belgorod (2015) 11-14(in Russian)
[10] Hermans J. and Ferro D. Representation of a protein molecule as a tree and application to modular computer programs which calculate and modify atomic coordinates, Biopolymers, 10(1971)1121-1129
[11] Godjayev N.M., Akyuz S., Akverdieva G.A.,"A molecular mechanics conformational study of peptide T", J. Mol. Structure 403 (1997) 95-110
[12] Akverdieva G.A., Godjayev N.M.,Akyuz S.," Conformational Dynamics of peptide T molecule". Journal of Molecular Structure, 609 (2002) 115-128
[13] IUPAC-IUB. "Quantities, Units and Symbols in Physical Chemistry", Blackwell Scientific, Oxford (1993), 168
[14] Shaitan K.V., Saraykin S.S., "Molecular dynamics method", 1999, http://www.moldyn.ru (in Russian)
[15] Horn H. W., Swope W. C., Pitera J. W. et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. 120 (2004) 9665-9678
[16] Allinger N.L.,Yuh Y., QCPE 395, "Quantum chemistry program exchange", Indiana Univ., Indiana, 1982
[17] Pattabiramann N., Sorensen K., Langridge R. et al. "Molecular mechanics studies of dermorphin", Biochem. Biophys. Res. Commun. 140 (1986) 342-349
[18] Pastore A., Temussi P.A., Salvadori S., "A conformational study of the opioid peptide dermorphin by one- dimensional and two-dimensional nuclear magnetic resonance spectroscopy", Biophysical Journal 48 (1985) 195–200
[19] Arlandini E., Ballabio M., De Castiglione R. et al. "Spectroscopic investigations on dermorphin and its [L-Ala2] analog", Int.J.Pept. and Protein Res. 25 (1985) 33-46
[20] Toma F., Dive V., Fermandjian S., Darlak K. et al., "Preferred solution and calculated conformations of dermorphin and analysis of structure-conformation-activity relationships in the series [Ala]-dermorphin analogs", Biopolymers 24(1985) 2417-2430.
[21] Ranghino G., Tosi C., Barino L. et al. "Conformational properties of the analgesic peptide dermorphin and their relationships to biological activity", J. Mol. Structure (THEOCHEM), 164 (1988) 153-164
[22] Renugopalakrishnan V., Rapaka R.S., "Conformational studies of dermorphin", NIDA Res. Monogr. 87(1988) 74-82.
[23] Segawa M., Ohno Y., Doi M. "Comparative conformational analyses of mu-selective dermorphin and delta-selective deltorphin-II in aqueous solution by 1H-NMR spectroscopy". Int J. Pept. Protein Res. 44(1994),295-304.
[24] Shenderovich M.D., Liyepinya I.T., Nikiforovich G.V. "Conformation-function relationships for dermorphin and its pentapeptide analogues", J.Bioorganic Chemistry, 15 (9) (1989), 1161-1172 (in Russian)
[25] Akverdieva G.A., "Structure-dynamical properties of dermorphin", Journal of Qafqaz University, 32(2011)7-12(in Azerbaijan)
[26] Akverdieva G.A.," Molecular mechanics study of [D-Arg2] – dermorphin tetrapeptide analogues". 5-th International Conference on Biological Physics (ICBP) 2004, August 23-27, Gothenburg, Sweden, p.89, B05-158.
[27] Akverdieva G.A., Nabiyev N.S., Godjayev N.M., Eminzade U.T. "Structure-functional relationship of the dermorphin tetrapeptide analogs", Baku University News, Ser. of Physical and Mathematical Sciences, 3(2005), 197-203(in Azerbaijan)
[28] Pogozheva I.D., Lomize A.L., Mosberg H.I."Opioid receptor three-dimensional structures from distance geometry calculations with hydrogen bonding constraints", Biophys. J. 75 (1998) 612-634.
[29] Akverdieva G.A. "Insights into bioactive conformation of dermorphin", Azerbaijan Journal of Physics, vol.XXII, 2(2016)12-20.

Interface for Interaction of GIS Technology for Electronic Master Plan Management and Third-Party Software Systems

Pavel V. Senchenko[1], Yury B. Gritsenko[1], Oleg I. Zhukovsky[1] , Yury P. Ekhlakov[1]
[1]Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russian Federation
pvs@tusur.ru

***Abstract* — The paper studies the possibilities for integration of software for electronic infrastructure master plan management with related information systems. The paper offers description of methods of cross domain interaction and discusses use of application programming interface for application integration.**

***Index Terms* — geographic information systems, electronic master plan, information system integration, application programming interface, document management.**

I. Introduction

A geographic information system for electronic master plan management (EMP GIS) is a unified corporate information system for processing of geographic data from spatial and attributive descriptions of an enterprise's production facilities. As distributed web-oriented software, the system requires an application programming interface (API) in order to be integrated with other software systems employed by the enterprise. One example of an application that requires such interface is an electronic document management system (EDMS) that, after integration with the GIS, can be used to generate and organize specific documents required for master plan management.

Another form of system interaction is data exchange by means of specific-format files. This form of interaction becomes necessary when third-party software only requires the data used by the GIS (without the API integration), for example, attributive descriptions of master plan objects that are necessary for data analysis and recovery. This paper analyses the key stages of implementation of EMP GIS to third-party software interface.

II. Statement of the Problem

One of the requirements for EMP GIS is system scalability, which means that the GIS should consist of independent software units responsible for specific functions of the system. The primary goal in development of an external system API is to develop such interaction model that would make it possible to start a light version of the EMP GIS as a widget (GUI element) of the third-party software.

Integration of web-oriented systems (where the client end is the web browser) is made difficult by the need for cross-domain system interaction between such systems.

The new EMP GIS can be installed and used in networks managed through the Microsoft Active Directory service. Active Directory has a hierarchical structure that is made up of objects (resources, services and user and group accounts) [1, 2]. A set of such objects is known as a forest – the top level of an enterprise's network infrastructure. A forest is a collection of one or more trees that share a trust relationship. The trees, in turn, include one or more domains that are identified by their DNS name structure (Domain Name Service).

As dictated by this enterprise network structure, various web-applications can run in various domains. Interoperability of such systems will be determined by domain constraint rules (Same Origin Policy). Such rules are used in all today's web browsers and do not have a single standard [3]. Their primary objective is to restrict access to administration of the content of web pages and web application within the domain from untrusted sites.

Thus, the choice of interaction model is determined by support of domain constraint rules in all modern web browsers and by sufficient security of such interaction.

III. Methods of Cross Domain Interaction

There are several basic methods of cross domain interaction:

1. JSONP technology (Java Script Object Notation with Padding). As opposed to JSON, which is a format for data transmission in the form of key:value pairs (objects appear as {key: value, key2: value2, …}), this technology make it possible to execute GET requests by using the HTML (HyperText Markup Language) tag <script>, which loads the JavaScript application and defines the request parameter as the name of the function that needs to be performed to process data [4]. One of the key drawbacks of this technology is vulnerability of sites using JSONP to cross site request forgery by malicious web pages, which often makes use of JSONP an unsafe method of cross domain interaction [5]. Moreover, JSONP does not allow data transfer with POST requests.

2. CORS (Cross-origin resource sharing) technology [6]. This technology makes it possible to define a list of trusted

client applications for the web application server using the Access-Control-Allow-Origin parameter (for example, Access-Control-Allow-Origin: http://example.com http:// another. example.com:8080). This method is safer to use than JSONP and it makes it possible to use the usual XMLHttpRequest. Its main disadvantage is that it is not supported by Microsoft Internet Explorer 6 and 7 and only partially supported by Internet Explorer 8.

3. Own cross domain request management mechanisms of individual web browsers. The key drawbacks of this method is that it is not supported by Microsoft Internet Explorer, and it requires and individual API for each browser, which significantly hinders software development and support.

4. Frames (HTML tag <iframe>). Using this approach, an iframe window is created on the page of the parent application, which can service GET requests by changing the URL, and POST requests by setting property as form.target with data referenced in the iframe window identifier.

5. PostMessage function. This mechanism is available as part of the HTML5 standard [5] and is supported by all browsers that support the standard. The key drawback of this method is that it is not supported by Microsoft Internet Explorer.

IV. API Development

Based on the above review of potential solutions, the authors have chosen to use the HTML5 PostMessage function. This approach does not require individual APIs to be developed for specific browsers, and is a standard recommended by W3C (World Wide Web Consortium) [5]. As an alternative for client applications that do not support HTML5, one of the following three methods can be used:

- Adobe Flash-assisted data exchange;
- data exchange through special HTML files on the client and the server sides;
- data exchange using the hash part of the URL bar in the iframe element.

The approach makes it possible to ensure that the API is supported by all contemporary web browsers by selecting the optimum interaction option for each individual situation.

Interaction is based on socket-class objects (special objects initiated on the client and server application side) provided by the JavaScript library of EasyXDM tools [7]. Let us discuss the interaction mechanism in more detail (Fig. 1).

In accordance with the above algorithm, connection is initiated by an external software system (client application), which creates a socket for connection with the EMP GIS. The main parameter of the socket is the URL of the web EMP GIS software that will be used as a widget. Parameters specified for the GET request are the name and the hash function of the user account password required for authorization in the system, for example:

http://gis/mapwidget?user={USER_NAME}&auth={SECRET_KEY}, where SECRET_KEY = md5({IP-address}.{USER_PASSWORD}).

Thus, user account data is transferred through an untrusted network securely.

After socket initialization, a light version of the web EMP GIS software is created in the external system that processes requests through the socket. After connection is established and confirmed, systems can interact by transmitting messages.

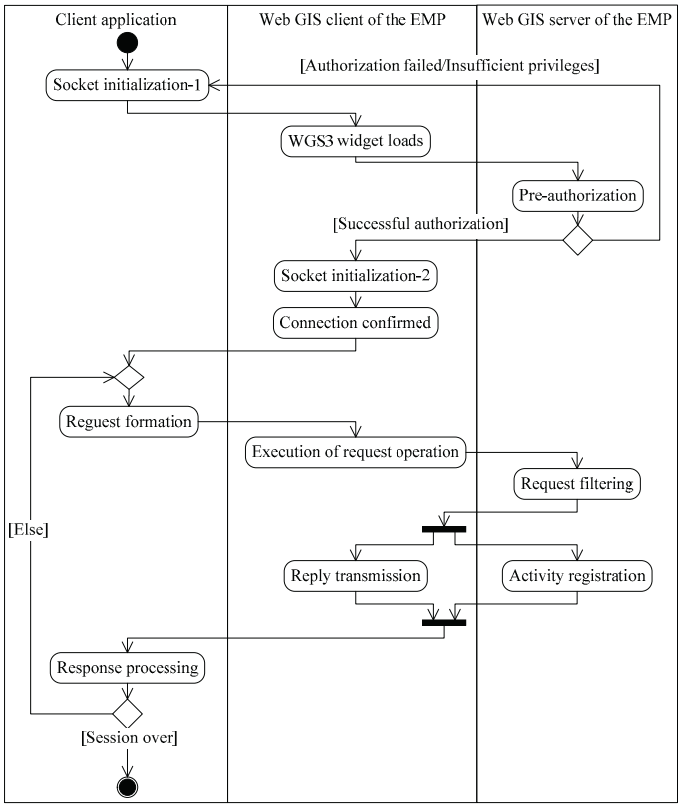


Fig. 1. Access control subsystem data model fragment

V. Message Exchange

The new API offers processing of two core types of requests:

- requests to select specific areas in the master plan;
- requests to select individual objects in the master plan.

Each of these types, in its turn, includes requests for current display parameters for cartographic description of the master plan.

The request is transmitted in the form of a fixed-structure message. Such message is presented as a JSON notation line (string representation of the JavaScript object).

The message is composed as follows: {type: message_type, content: message_content}, where message_type is the number literal determining the type of request, message_content is the request parameters. Request parameters include the map context (list of active layers, scale and the center point), and data for identification of the selected area or object in the master plan. The system supports four values of message_type:

1) get the selected area of the master plan and its current context;

2) set the selected area of the master plan and its current context;

3) get the ID of the selected object and the current map context;

4) set the ID of the selected object and the current map context.

VI. Map Context Transmission

Map context is used to unambiguously define the visible area in the master plan and list the active layers of the map. The area is unambiguously defined by coordinates of the central point in the map and its scale. The layers are listed using a 32-bit cyclic redundancy code (crc32). Each layer is identified by a crc32 checksum of its core properties (name, layer group, location in geodata repository). The approach makes it possible to avoid dependence of API operation on current properties of the spatial description data base and values of spatial layer IDs that can change with time.

Thus, the context of the cartographic object is represented by the following line: <coordinate X of the map center>,<coordinate Y of the map center> | <map scale> | < checksum of active layer 1>,< checksum of active layer 2>,…,< checksum of active layer n>.

In addition to these properties, the context can include a description of the selection: "|<selection type>| <selection property 1>, < selection property 2>, …, < selection property т>", where <selection type> can have two values: radius – circle selection, and polygon – polygonal selection. The selection is described through its properties which are the x and y axis coordinates of the circle center and the circle selection radius and the vertex coordinates of the polygonal selection.

VII. Use of the EMP GIS API for Integration with Third-Party Applications

Let us analyze the problem of software development for electronic master plan document management as an example of GIS API application.

In general, document flow management includes processing of incoming, outgoing and internal correspondence: receipt and preprocessing of incoming documents, coordination and approval of draft documents filed within the enterprise, formation and control of assignments, dispatch of outgoing documents. Specifics of document flow processes are determined by the properties of the business processes being automated. For example, if infrastructure management processes are being automated, the enterprise EDMS must offer the following solutions specific to the area [8]:

- ensure registration of all document types in the overall document flow of the EMP;
- maintain the electronic archive of EMP documents;
- ensure interaction with other specialized subsystems of the EMP software complex;
- ensure automated control of document execution;
- maintain the technology of electronic interaction between enterprise divisions (EMP users);
- monitor documents – identify the stage of document review;
- identify connection between documents executed and various levels;
- store all document information and ensure access to any stage of document processing;
- keep records of document execution and performance discipline of employers.

Delivering the above solutions requires the establishment of data connection between infrastructure facilities and documents circulating in the general document flow. There can be two types of such connection: 1) document link to a specific master plan object, which is described by its unique ID in the EMP system; 2) document link to a spatial area in the master plan that is characterized by a set of points (polygon). These connections in the EDMS can be reflected in an additional database table with the following structure (Table I)

TABLE I. Structure of table detailing links between documents in the EDMS and EMP spatial data

PK*	Field name	Data type	Comment
+	ID	NUMBER(10)	Link ID
	LinkType	NUMBER(2)	Link type code
	DocumentID	NUMBER(10)	Document ID
	GIS_ObjectID	NUMBER(10)	EMP object ID
	GIS_Area	BLOB	Master plan area description
	GIS_Context	BLOB	EMP context
	AddedAt	DATE	Time of connection

Note. PK* – Primary Key.

Please note the inclusion of information about the electronic map context in the above structure. Context information does not characterize the link between a document and an EMP object, but if included, it ensures unambiguous definition of the visible master plan area when the electronic map is displayed.

Link between a document and an EMP object is reflected in the user interface of the electronic document flow system with the help of the EMP GIS widget, as shown in Fig. 2.

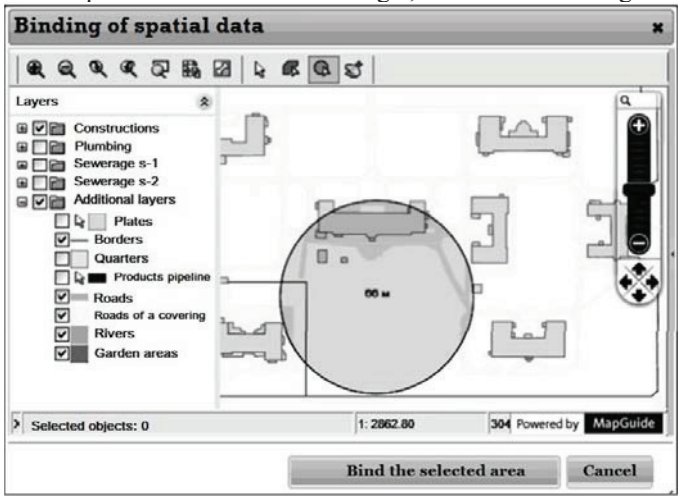


Fig. 2. Association of spatial data with a document in the EDMS

This user element is integrated in the EDMS interface with the help of a special class GEOSelectorControl, which encapsulates the EMP GIS API interaction logic and electronic map display in the EDMS.

The process of master plan fragmenting can be used to demonstrate the link between spatial data and document data. When a fragment request is registered in the EDMS, the user has to specify the territory for fragment formation. Another

example would be the process of adding a new spatial object to the master plan and registering the supporting document in the EDMS.

VIII. Conclusion

As a result of the work described and within the W4GIS project, the authors have developed an interaction model and a programming module for integration of the electronic master plan management GIS in user interface of third-party web applications. The interaction technology used in the project is based on W3C standards and supported by all contemporary web browsers.

This article is designed as part of the national project of the Ministry of Education and Science of the Russian Federation № 3653 "Models, algorithms and software to support decision-making in risk management in the socio-economic and industrial-technological system".

References

[1] Gritsenko, Y. B., Milikhin, M. M., Senchenko, P. V., & Zhukovsky, O. I. (2015). Model of role-based access to spatial data of electronic master plan. Paper presented at the 9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings, 196-200. doi:10.1109/ICAICT.2015.7338545

[2] Li, H. -., Li, J. -., & Zhu, H. -. (2005). Access control of secure operating system for electronic government application. Shanghai Jiaotong Daxue Xuebao/Journal of Shanghai Jiaotong University, 39(8), 1292-1295+1299.

[3] Same Origin Policy. Accessed 21 April 2016. Available: http://www.w3.org/Security/wiki/Same_Origin_Policy.

[4] RIAspot JSON P for Cross Site XHR. Accessed 21 April 2016. Available: http://www.ria-spot.com/blogs/entry/JSONP-for-Cross-Site-XHR.

[5] Json-p.org. The problem. Accessed 19 April 2016. Available: http://json-p.org.

[6] Cross-Origin Resource Sharing. Accessed 18 April 2016. Available: http://www.w3.org/TR/cors.

[7] EasyXDM. Home. Accessed 19 April 2016. Available: http://easyxdm.net.

[8] Gritsenko, Y. B., Zhukovsky, O. I., Lazarev I.V., Milikhin, M. M. & Senchenko, P. V. "Interface between GIS technology of electronic master plan with other software systems". Reports TUSUR. #1, vol. 27, 2013. pp. 129-134. (In Russian).

Existence of repetitions of digits in the construction of periodic sequences related to 3x+1 problem

Proof of the theorem on the periods in the 3-adic representation of terms of periodic sequences that appear in Collatz problem

Yagub N. Aliyev
Dept. of Mathematics and Informatics Teaching
Qafqaz University
Khyrdalan AZ0101, Azerbaijan
yaliyev@qu.edu.az

***Abstract*—The paper is the continuation of the previous paper on this topic. We first review some of the recent developments on 3x+1 problem. In the remaining part of the paper we prove the result announced at the end of the first paper. The proof uses the finite version of the method of mathematical induction.**

***Index Terms*—3x+1 problem, Collatz problem, p-adic numbers, 3-adic numbers, periods.**

I. Introduction

In one of the Informatics Olympiads in USSR the following problem was proposed to the contestants:

Let x_0 be a positive integer and for n>0 define recursively

$x_n=x_{n-1}/2$ if x_{n-1} is an even integer and
$x_n=(3x_{n-1}+1)/2$ if x_{n-1} is an odd integer.

For example if one starts with $x_0=7$ the successive numbers will be $x_1=11$, $x_2=17$, $x_3=26$, $x_4=13$, $x_5=20$, $x_6=10$, $x_7=5$, $x_8=8$, $x_9=4$, $x_{10}=2$, $x_{11}=1$, $x_{12}=2$, $x_{13}=1$, …. Write a computer program for the calculation of the terms of the sequence. Try the numbers $x_0=27$ and $x_0=2000007$.

We do not know the author and what was the official or students' solutions of this problem but we have found an algorithm which is sufficiently fast and will not have difficulties with overflow of digits. We have explained our algorithm in [1] but for completeness we will give it here, too. We must first note that this is not the first appearance of this recursive sequence. This recursive sequence has a long history. It is conjectured that for arbitrary positive integer x_0 there is a number n such that $x_n=1$. This problem, named as "3x+1 problem", is known for more than 60 years and it is still unsolved. For more information see [1-3]. We are more interested with a different problem that states: Prove that if the sequence x_n is periodic then $x_0=1$ or $x_0=2$ that is the only periodic sequences satisfying the above conditions are 1,2,1,2,…. and 2,1,2,1,…. This problem remains unsolved, too.

II. Division by 2 algorithm in 3-base number system

We have already described this algorithm in [1] (with typos unfortunately). We will just state its main steps here. First write the number in 3-base number system. Only the digits 0,1,2 can appear in this representation. We use the following substitutions to find the digits of $x/2$ from the digits of x:

$$A=\begin{pmatrix}0 & 1 & 2\\ 0 & 2 & 1\end{pmatrix},\ B=\begin{pmatrix}0 & 1 & 2\\ 1 & 0 & 2\end{pmatrix}.$$

Unfortunately the substitution B appeared in [1] as identical to the substitution A. We hope that this has not caused a confusion to the readers of [1]. We will explain the algorithm using the following even number written in 3-base number system: x=1022120121200102012. We start from right with 1st digit and apply the substitution A. We write the digits obtained in this and all the next steps under the corresponding digits of the given number. If n^{th} digit (counted from right) is even then we apply to the next $(n+1)^{th}$ digit the same substitution which was applied to n^{th} digit.) But if n^{th} digit (counted from right) is odd then we apply to the next $(n+1)^{th}$ digit the substitution which is different from what was applied to n^{th} digit i.e. if A is applied to n^{th} digit then we apply B to $(n+1)^{th}$ digit and vice versa. So for the given number x the result will be:

x=1022120121200102012
x/2=0122210022100012121

It is possible to generalize this algorithm to the other bases with odd order $2k+1$. We have explained the algorithm in [4]. The substitutions A and B in this case will be

$$A=\begin{pmatrix}0 & 2 & 4 & \dots & 2k & 1 & 3 & 5 & \dots & 2k-1\\ 0 & 1 & 2 & \dots & k & k+1 & k+2 & k+3 & \dots & 2k\end{pmatrix},$$

$$B=\begin{pmatrix}0 & 2 & 4 & \dots & 2k & 1 & 3 & 5 & \dots & 2k-1\\ k & k+1 & k+2 & \dots & 2k & 0 & 1 & 2 & \dots & k-1\end{pmatrix}.$$

This substitutions were also discussed in [5] in a different context.

III. Theorem and its proof

This part of the paper will be heavily relying on the notations of our previous paper [1]. Therefore we recommend the readers to read the paper [1] first and then return to the current paper.

At the the end of [1] we have stated the following theorem without proof. We have discussed the idea of the following proof in various conferences and seminars [6, 7].

Theorem. *Suppose that a sequence of operations with m operations S and n operations T in some order is given. Then the corresponding* 3*x*+1 *construction contains the main period of the construction of the special numbers* $U_i = \frac{2^i}{2^{n+m}-3^m}$ ($i = 0,1,2,\ldots,n+m$).

Proof. Let us first prove the claim for the special sequence $P_0 = \underbrace{SSS\ldots S}_{m}\underbrace{TTT\ldots T}_{n}$ which is the sequence of *m* operations *S* followed by *n* operations *T*. This sequence is called in [7, Hints and Solutions to Puzzle 7] as (*m*,*n*) passage. The proof of impossibility of (*m*,*n*) passages in [7] is unfortunately very difficult to understand and it is probably erroneous or at least not complete. We have proved the impossibility of (*m*,*n*) passages for sufficiently large *m* and *n* in [9].

Note that the first and last rows of the construction will be 3-adic represantations of the number $\frac{3^m-2^m}{2^{n+m}-3^m}$. Indeed, one can try to solve the following linear equation:

$$\frac{\left(3\left(\ldots\left(3\left(\frac{3\left(\frac{3x+1}{2}\right)+1}{2}\right)+1\right)\ldots\right)+1\right)}{2}}{2^n} = x.$$

The number of 3s in the equation is *m*. Let us denote this and the following numbers in the construction by $A_0, A_1, A_2, \ldots, A_{n+m} = A_0$. Then

$$A_0 = \frac{3^m-2^m}{2^{n+m}-3^m} = U_{n+m} - U_m - 1$$

$$A_1 = \frac{3A_0+1}{2} = 3\left(U_{n+m-1} - U_{m-1}\right) - 1$$

…

$$A_m = \frac{3A_{m-1}+1}{2} = 3^m\left(U_n - U_0\right) - 1$$

$$A_{m+1} = \frac{A_m}{2} = 3^m U_{n-1} - U_{n+m-1}$$

$$A_{m+2} = \frac{A_{m+1}}{2} = 3^m U_{n-2} - U_{n+m-2}$$

…

$$A_{m+n} = \frac{A_{m+n-1}}{2} = 3^m U_0 - U_m = A_0$$

.	.	.	1	2	2	1	1	1	2	1	2	1	0	1	1				S
.	.	.	0	2	2	2	0	2	1	0	2	2	0	0	2	1			S
.	.	.	0	1	1	1	0	1	0	1	2	2	1	1	2	2	1		S
.	.	.	0	0	2	0	1	2	0	0	2	2	2	0	2	2	2		T
.	.	.	1	1	2	1	2	1	0	0	1	1	1	0	1	1	1		T
.	.	.	0	2	1	0	2	2	0	0	0	2	0	1	2	0	2		T
.	.	.	0	1	0	1	2	2	1	1	1	2	1	2	1	0	1	1	

(1)

In this and the following constructions the vertical lines separate the group of digits that appear first from the groups that appear later (see [1] for more details). The pair of horizontal short lines indicate the main periods and they can be drawn at any place of the construction with arbitrary length. They indicate the repeating digits of the construction that are separated by the minimum number of columns (in this example it is 4).

By the definition of the main period, the construction of $U_0, U_1, U_2, \ldots, U_{n+m}$ and its cyclic permutation

$$U_m, U_{m-1}, U_{m-2}, \ldots, U_0, U_{n+m-1}, U_{n+m-2}, \ldots, U_m$$

contain this main period. The example of the construction corresponding to the cyclic permutation with m=n=3 is the following:

.	.	.	1	0	1	0	1	2	2	1	1	1	2	2					$=U_3$
.	.	.	0	1	2	0	0	2	2	2	0	2	1	1					$=U_2$
.	.	.	1	2	1	0	0	1	1	1	0	1	0	2					$=U_1$
.	.	.	0	2	2	0	0	0	2	0	1	2	0	1	0	0	1		$=U_6$
.	.	.	1	2	2	1	1	1	2	1	2	1	0	0	1	1	2		$=U_5$
.	.	.	0	2	2	2	0	2	1	0	2	2	0	0	0	2	1		$=U_4$
.	.	.	0	1	1	1	0	1	0	1	2	2	1	1	1	2	2		$=U_3$

(2)

We have underlined the main period in the last example, too.

From the expression of A_k in terms of U_j and from the last observation about the presence of the main period in the

construction of $U_0, U_1, U_2, ..., U_{n+m}$ and its cyclic permutation

$$U_m, U_{m-1}, U_{m-2}, ..., U_0, U_{n+m-1}, U_{n+m-2}, ..., U_m$$

It follows that (*m*,*n*) passage contains the main period. The change caused by the subtraction of 1 in the expression for $A_0, A_1, A_2, ..., A_m$ will not affect the main period. When the numbers $U_0, U_1, U_2, ..., U_{n+m}$ are not integers, this subtraction of 1 can change only finitely many digits at the end of these numbers. Note that the cases when the numbers $U_0, U_1, U_2, ..., U_{n+m}$ are integers, occur only for some special known choices of m and n $\left(2^{n+m} - 3^m = \pm 1\right)$, and these cases can be studied individually.

It remains only to prove that when one changes any pair of operations ST into TS in a sequence of operations P_1, whose construction enjoys the main period, one obtains a sequence of operations P_2, whose construction also enjoys the same period. We can suppose that the pair ST, that will be later transposed, is at the end of P_1. Indeed, we can always achieve this by doing cyclic permutation of the sequence of operations P_1. This cyclic permutation of P_1 will do cyclic permutation in the rows of corresponding construction and this does not affect the presence of the main period in the construction. For example we can replace the construction (1), corresponding to the sequence of operations SSSTTT with the following construction corresponding to the sequence TTSSST:

. . .	1	2	1	0	0	1	1	1	0	1	1	1			*T*
. . .	0	2	2	0	0	0	2	0	1	2	0	2			*T*
. . .	1	2	2	1	1	1	2	1	2	1	0	1	1		*S*
. . .	0	2	2	2	0	2	1	0	2	2	0	0	2	1	*S*
. . .	0	1	1	1	0	1	0	1	2	2	1	1	2	2	1	*S*
. . .	0	0	2	0	1	2	0	0	2	2	2	0	2	2	2	*T*
. . .	1	1	2	1	2	1	0	0	1	1	1	0	1	1	1	

(3)

Let us denote by R_i the sequence formed by the first *i* operations in P_1. Similarly, denote by σ(i) the number of S operations in R_i. In particular, σ(0)=0. Finally, denote by R_i[b] the number obtained from application of the operations of the sequence R_i to the number b the in the given order (from left to write). In particular, R_0[b]=b. As we said the sequence P_1 ends with ST that is $P_1=R_{n+m-2}ST$. The sequence P_2 is obtained from P_1 from transposition of these S and T at the end of P_1. That is $P_2= R_{n+m-2}TS$. Let us denote by b_0 and d_0 the numbers obtained in the first (or the last) row (except the last digit 1 appearing if the first operation in the sequences P_1 and P_2 is S) of the constructons corresponding to the sequence of operations P_1 and P_2. In other words $P_1[b_0]=b_0$ and $P_2[d_0]=d_0$. Let us find difference of these numbers and simplify:

$$b_0 = P_1[b_0] = \frac{\frac{3R_{n+m-2}[b_0]+1}{2}}{2},$$

$$d_0 = P_2[d_0] = \frac{3\frac{R_{n+m-2}[d_0]}{2}+1}{2},$$

$$d_0 - b_0 = \frac{3}{2^2}\left(R_{n+m-2}[d_0] - R_{n+m-2}[b_0]\right) + \frac{1}{4} =$$

$$= \frac{3^m}{2^{n+m}}\left(d_0 - b_0\right) + \frac{1}{4}.$$

Now we can express one of the numbers in terms of the other number:

$$d_0 = b_0 + \frac{1}{4}\frac{2^{n+m}}{2^{n+m} - 3^m} = b_0 + U_{n+m-2}.$$

It is now easy to check that if

$$b_i = R_i[b_0] \ \left(i = 0,1,2,..., n+m-2\right)$$

Then

$$d_i = R_i[d_0] = R_i[b_0] + 3^{\sigma(i)}U_{n+m-2-i}$$
$$\left(i = 0,1,2,..., n+m-2\right).$$

So the numbers in the first n+m-2 rows of the construction corresponding to the sequence P_2 are the sums of the numbers in the first n+m-2 rows of the construction corresponding to the sequence P_1 and the corresponding numbers of the construction obtained from cyclic permutation of the construction of the numbers

$$U_i = \frac{2^i}{2^{n+m} - 3^m} \ (i = 0,1,2,...,n+m).$$

In our example m=n=3 the last construction will look like this

. . .	2	0	2	1	0	2	2	0	0	0	2	1			$=U_4$
. . .	1	0	1	0	1	2	2	1	1	1	2	2			$=U_3$
. . .	0	1	2	0	0	2	2	2	0	2	1	1			$=U_2$
. . .	1	2	1	0	0	1	1	1	0	1	0	2			$=U_1$
. . .	0	2	2	0	0	0	2	0	1	2	0	1	0	0	1	$=U_6$
. . .	1	2	2	1	1	1	2	1	2	1	0	0	1	1	2	$=U_5$
. . .	0	2	2	2	0	2	1	0	2	2	0	0	0	2	1	$=U_4$

(4)

Since the first n+m-2 rows of the construction of P_1 and the first n+m-2 rows of the cyclic permutation (4) of the

construction of numbers U_i have the main period, their sum that is the first n+m-2 rows of the construction of P_2 will also have the main period. To complete the proof it remains only to check the claim for the last two rows namely for the rows n+m and n+m+1 which is identical to the first.

$$d_{n+m-1} = T[d_{n+m-2}] = T[b_{n+m-2} + 3^{m-1}U_0] =$$
$$= T[b_{n+m-2} - 1 + 3^{m-1}U_0 + 1] =$$
$$= T[b_{n+m-2} - 1] + T[3^{m-1}U_0 + 1].$$

The number $T[b_{n+m-2} - 1]$ is identical to $b_{n+m-1} = S[b_{n+m-2}]$ if we ignore the last digit in the last number. Similarly, the number $T[3^{m-1}U_0 + 1]$ (ignore the last digit) is identical to the number $T[3^m U_0 + 1] = T[U_{n+m}] = U_{n+m-1}$ if we again ignore the last two digits in the last number. So we see that the row n+m-1 is the sum of two rows of the constructions having the same main period. For the last row we can write

$$d_{n+m} = d_0 = b_0 + U_{n+m-2} = b_{n+m} + U_{n+m-2}.$$

The proof is complete.

IV. Conclusion

We will summarize the results of this and previous paper [1] here. The proved theorem shows that in a certain sense the structure of the construction of the numbers

$$U_i = \frac{2^i}{2^{n+m} - 3^m} \ (i = 0,1,2,\ldots,n+m).$$

repeats itself in all constructions of the sequences with m operations S and n operations T in any order. Below we will show two constructions. The first (SSTTST) is one of the conctructions corresponding to the case m=n=3. The second is the construction of the numbers

$$U_i = \frac{2^i}{2^6 - 3^3} = \frac{2^i}{37} \ (i = 0,1,2,3,4,5,6).$$

. . .	0	1	2	2	1	1	1	2	1	2	1	1	1		S
. . .	0	0	2	2	2	0	2	1	0	2	2	0	2	1	S
. . .	0	0	1	1	1	0	1	0	1	2	2	1	2	2	T `
. . .	0	0	0	2	0	1	2	0	0	2	2	2	1	1	T
. . .	1	1	1	2	1	2	1	0	0	1	1	1	0	2	1	S
. . .	2	0	2	1	0	2	2	0	0	0	2	0	1	2	2	T
. . .	1	0	1	0	1	2	2	1	1	1	2	1	2	1	1	

. . .	0	2	2	0	0	0	2	0	1	2	0	1	0	0	1	$=U_6$
. . .	1	2	2	1	1	1	2	1	2	1	0	0	1	1	2	$=U_5$
. . .	0	2	2	2	0	2	1	0	2	2	0	0	0	2	1	$=U_4$
. . .	0	1	1	1	0	1	0	1	2	2	1	1	1	2	2	$=U_3$
. . .	0	0	2	0	1	2	0	0	2	2	2	0	2	1	1	$=U_2$
. . .	1	1	2	1	2	1	0	0	1	1	1	0	1	0	2	$=U_1$
. . .	0	2	1	0	2	2	0	0	0	2	0	1	2	0	1	$=U_0$

One can easily see the similarity in the structure (the triangular blocks of zeros and twos) of digits in these constructions.

References

[1] Aliyev Y.N., Suleymanov V.A., Construction of periods for 3x+1 problem: Use of division algorithm by 2 in 3-base number system for construction of 3-adic numbers as periods of Collatz sequence, 7th International Conference on Application of Information and Communication Technologies (AICT), Conference Proceedings, 2013, 413-415.

[2] Conway J.H., On Unsettleable Arithmetical Problems, The American Mathematical Monthly, Vol. 120, No. 3 (March 2013), pp. 192-198.

[3] Tezer C., Collatz Problem, Matematik Dünyası, 1999-I, 11-18. (in Turkish)

[4] Aliev Ya.N., Suleymanov V.A., New algorithm for division by two in base three number system (in Russian), Actual problems of Mathematics and Mechanics, BSU, 2012, Baku, 23-26.

[5] Briskin Ya., Vasiliev N., Problem M1342, Kvant, 11 (1992) 20-21. (in Russian)

[6] Aliyev Y.N., 3x+1 problem and construction of periods of 3x+1 sequence, Proceedings of 9th Ankara Matematics Days, Atilim University, Matematics Department, 12-13 June 2014, 16.

[7] Yagub Aliyev, Construction of 3X+1 sequence periods and patterns in repetitions of digits in these constructions, Proceedings of 2nd International Conference of Young Researches, 18-19 April 2014, Baku Azerbaijan, 546-547. (in Azerbaijani)

[8] Arsac J., "Jeux et Casse-tête a Programmer", Bordas, Paris, 1985. (in French)

[9] Aliyev Y.N., Aliyev N.Y., "On some applications of Gelfond's theorems, J. Qafqaz Univ., Number 17, Volume 1 (2006), 11-14.

Selection of time series forecasting model, using a combination of linguistic and numerical criteria

T. Afanasieva
Information Systems Department
Ulyanovsk State Technical University
Severny Venec, 32, 432027 Ulyanovsk, Russia
tv.afanasjeva@gmail.com

A. Sapunkov
Information Systems Department
Ulyanovsk State Technical University
Severny Venec, 32, 432027 Ulyanovsk, Russia
sapalks@gmail.com

Abstract – **One of the problems in forecasting is selection a subset of suitable models, which perform accurate results not only in tested part of time series (TS), but in real forecast. In previous our paper a TS forecasting technique was proposed, where a framework of TS model selection using linguistic and numerical criteria was proposed. To choose a subset of suitable models from a given set of TS models we developed a linguistic description of TS behavior based on identification of TS general tendency. In this paper we focus on extension of a linguistic description of TS behavior and study it's efficiency in out of sample TS part. The study of a proposed framework in forecasting of 91 TS showed the improvement of accuracy in comparison with TS model selection, which used a numerical criterion only.**

Keywords— **Time series, model selection, forecasting, linguistic description, general fuzzy tendency.**

I. Introduction

One of the traditional practical tasks in different domains is the analysis of time series (TS), stored in data sets. These TS could have different lengths, behaviors, frequencies, so, some set of TS models is needed to obtain the appropriate forecasts. To our days there are various forecasting techniques, models and their combinations for TS with stationary and non-stationary properties. Many of them can produce forecasts of given TS with almost the same good accuracy being tested. But sometimes these TS forecasting models are not good, when they are applied to real practical task. In this case TS model overfitting or incorrect type of TS model could be a cause. The problem is how to choose from the set of given forecasting techniques the appropriate one (or their combination) to obtain adequate and accuracy TS predictions in real practical task.

Well-known statistical models [1], [2], [3], [4] are optimum for a wide class of TS. In statistical techniques regressions and (or) auto regressions are used to predict TS points and global trends. However, they aren't rather effective and sometimes inadequate for short-term time series (typically, 7- 40 time points). To solve the problem of short-term TS forecasting fuzzy models were presented in [5], [6], [7], [8] and a combination of statistical models and fuzzy sets are discussed in [9], [10], [11], [12]. In [12] three groups of fuzzy time series models are considered: (1) a regression model with fuzzy regression coefficients [13], [14], (2) a Box-Jenkins model, based on a fuzzy autocorrelation coefficient [15], [16], (3) a fuzzy reasoning in fuzzy TS model [5], [7]. The notion and application of TS fuzzy tendencies in performing short-term forecasts were considered in [17], [18].

In the paper [19] a Hybrid Forecast Model Combining Fuzzy Time Series and a New Smoothing Technique is proposed. Authors described a transformation of given numerical TS into new TS, where each value is a label of some cluster. Then new TS of labels is forecasted by fuzzy model. The forecast method is based on last two known values that belong to clusters *Li* and *Lj*, respectively.

The Adaptive Fuzzy C-Regression Modeling technique, presented in [20] is based on the first order Takagi-Sugeno fuzzy model. Authors construct fuzzy model by learning the rule antecedent using a fuzzy clustering.

There is the fact, that now many different techniques and models can be used to TS forecasting, but the problem is how to choose the most adequate and accurate ones.

To choose the adequate and accurate model from the set of given TS forecasting models the various numerical criteria are used usually. Sometimes they are contradictory: the best TS model, estimated by one of criteria, is not the best by another one.

Another problem is overfitting, when a model outputs out of sample represent the TS, not similar to given TS behavior. So, that model has poor predictive performance and not good in tendency generalization. This problem occurs also when the TS length is small and there are not enough data points to train or to test a forecasting model.

On the other hand, one of the developing trends in TS analysis is aimed not only on forecasting, but also on their linguistic description [23], [24]. These linguistic descriptions are based on the notions of fuzzy sets and linguistic variables and perform information about TS behavior. Moreover, the linguistic description of a TS behavior can be an additional criterion in TS model estimation. In [18] we proposed and applied an algorithm of TS linguistic description, based on fuzzy tendencies to select the best TS model. We take in account five types of TS tendencies: "increase", "decrease", "stability", "fluctuation" and "chaos". These linguistic descriptions characterize different types of TS behavior. However, the study of efficiency of proposed TS linguistic

description [18] applied in selection of TS model has received far less attention and is still an open problem. In practice most difficult problem is forecasting TS, which is characterized by linguistic term "chaos", so any new information about behavior of such TS will be useful.

The aim of this contribution is to develop the extended algorithm of TS linguistic description, in particular with details about behavior "chaos", and to study accuracy of it's application in selection TS forecasting model.

II. Framework of choosing TS forecasting model

Like in the classical TS analysis, the main task is to choose the adequate and accurate model from the finite set of TS models. This means the process, where the type of the best model from the set of optimal models for the given TS is selected. In this paper we will distinguish three types of TS forecasts: (1) in sample, (2) out of sample and (3) real forecasts. The first one is determined on the part of TS, which used as training part to fit a model (in sample forecasts). The second type of forecasts performed on some last part of given TS (out of sample forecasts) as usually, which can be used to test a model. For the first and second types of TS forecasts the accuracy by some numerical criteria can be calculated. The third type of TS model forecasts is the real forecasts of given TS, determined by some horizon. The real values for given TS are unknown in that TS part and must be predicted in practice. Therefore, the TS model accuracy of the real forecasts can't be calculated by any numerical criteria.

Usually the forecasting quality of the optimal TS model is estimated by the minimum value of some numerical criterion. In practice sometimes the real forecasts of TS best model represent a behavior, not similar to given TS behavior. This problem arises in case if not sufficiently effective TS model was chosen or in case of model overfitting. To deal with this problems, a framework, named two step TS model selection was proposed in [18].

In that framework a combination of linguistic and numerical criteria was used to select the best model from the given set of optimal in their class TS models. The main idea in our contribution [18] is imitation a TS model evaluation as it done by analysts. Traditionally, at first, analyst visually estimates TS behavior in terms "increase", "decrease", "stability", "fluctuation", "chaos", and then chooses one or several types of models. The latter are estimated by criteria of two types. The first type of criteria, which takes in account by the analyst, is a visual criterion of tendencies (behaviors) similarity of given and predicted TS, involving in sample, out of sample and real forecasts. The second type of criteria is the numerical criterion of TS model accuracy (in sample and out of sample forecasts), such as SMAPE, RMSE and MAPE. To imitate this process two step TS model selection was proposed [18], using algorithm of identification of TS behavior. The algorithm of identification of TS behavior, described in [18] performs only five types of TS behavior: "increase", "decrease", "stability", "fluctuation" and "chaos". It was assumed that "stability" and "fluctuation" described a behavior of stationary TS, "increase", "decrease" are corresponded to non-stationary TS with trend. Linguistic term "chaos" determines non-stationary behavior of TS. There is a question how to derive more detailed description about non-stationary TS behavior "chaos". To perform more information about TS behavior "chaos" extended algorithm of identification of TS behavior, based on the notion of fuzzy tendency [12], will be proposed in next Section. This algorithm extends the proposed algorithm given in [18] by performing more detailed linguistic description of TS behavior "chaos" and by using TS transformation into TS of elementary fuzzy tendencies. Then the extended algorithm will be applied in TS model selection as was presented in [18] to improve accuracy of real forecasts, performed by fuzzy TS models. The results of forecasts, advantages and lacks will be discussed at the last Section of the paper.

III. The algorithm of identification of TS general tendency

The proposed extended algorithm of identification of TS behavior is based on a notion of fuzzy tendency [12] and TS of fuzzy elementary fuzzy tendencies [17]. This algorithm outputs the linguistic description of TS general tendency, so it was named as an algorithm of identification of TS general tendency.

In this Section the notion of fuzzy tendency and TS of fuzzy elementary fuzzy tendencies will be recalled. Then a transformation of given TS into TS of fuzzy elementary fuzzy tendencies will be presented firstly. Then the latter will be used in extended algorithm of TS behavior identification to derive the linguistic description of TS general tendency. The result of this algorithm is represented by six types of linguistic description of TS general tendency "decrease", "stability", "increase", "fluctuation", "fluctuation with growth", "fluctuation with fall", which will be implemented in two step TS forecasting model selection, represented in Section 4.

A. The notion of fuzzy tendency and TS of elementary fuzzy tendencies

The notion of fuzzy tendency of a fuzzy time series was described in [12] as a quality characteristic of a TS behavior in linguistic terms. It was developed and successfully used in forecasting technique of short TS in combination with the F-transform [17], [18].

Let us remind a description of fuzzy tendency, given in [18].

Let $Y = x_t$, $x_t \in X$, $X \subset R^1$, $t = 1, 2, \ldots, n$ be a given numerical TS, on which fuzzy TS $\tilde{x}_t, \tilde{x}_t \in \tilde{X}, t = 1, \ldots, n$, is defined. In accordance with the basic provisions of the fuzzy TS theory, proposed by Q. Song and B.Chissom [5], any finite discrete TS can be transformed into FTS $\tilde{x}_t, \tilde{x}_t \in \tilde{X}, t = 1, \ldots, n$, provided that its set of values X can be covered by special functions (membership function of fuzzy sets) of some linguistic variable $\tilde{X} = (\tilde{x}_i | i = 1,2, \ldots, r)$. We assume each $\tilde{x}_i$ satisfy the properties of normality and convexity. Let us define for a given numerical TS $x_t \in X, X \subset R^1, t = 1, 2, \ldots, n$, $\Delta x = \max|x_i - x_j|, x_i, x_j \in X, i, j = 1,2, \ldots, n$.

Definition 1 [18]. A fuzzy tendency (FT) for a fuzzy TS $\tilde{x}_t, \tilde{x}_t \in \tilde{X}, t = 1, \dots, n$, derived by some linguistic variable $\tilde{X} = (\tilde{x}_i | i = 1,2, \dots, r)$ is a set of the components:

$$< \tilde{V}, \tilde{A}, \Delta t > \quad (1)$$

where $\tilde{V}$ and $\tilde{A}$ are the linguistic variables, which characterized a type and an intensity of FT respectively. Here $\tilde{V}$ is a linguistic variable with the possible linguistic values "decrease", "stability", "increase", "fluctuation", "chaos", defined on the set $X_V \in R^1, X_V = [-\Delta x, \Delta x]$. The component $\tilde{A}$ is a linguistic variable, which is defined on the set $X_A \in R^1, X_A = [0, \Delta x]$ with a possible linguistic values "zero", "very small", "small", "medium", etc.

The variable Δt in formula (1) determines duration of FT and kinds of FT:

In case if $\Delta t = n - 1$, then FT is a "general fuzzy tendency". If $\Delta t = 1$, then FT is a an "elementary fuzzy tendency".

A general FT characterizes the behavior of TS in whole. An elementary FT characterizes the elementary difference between two neighboring of fuzzy TS values.

It is easy to see that any general FT of a TS can be considered as a sequence of an elementary FT. Moreover, the sequence of elementary FT of fuzzy TS in respect to (1) represents TS of an elementary FT by means of two fuzzy TS [18]:

$$(\tilde{v}_t, \tilde{a}_t), \tilde{v}_t \in \tilde{V}, \tilde{a}_t \in \tilde{A}, t = 2,3, \dots, n. \quad (2)$$

This TS of elementary FT will be used to output the linguistic description of given TS behavior.

Notice, that for elementary FT the following statements are fulfilled. The set of types of elementary FT involved only three linguistic values $\tilde{V} = \{\tilde{v}1, \tilde{v}2, \tilde{v}3\}$, $\tilde{v}1$="decrease", $\tilde{v}2$="stability", $\tilde{v}3$="increase". The set of intensity of elementary FT involved $(r-1)$ linguistic values $\tilde{A} = \{\tilde{a}_j | j = 1,2, \dots, (r-1)\}$.

B. The transformation of numerical TS into TS of elementary fuzzy tendencies

We consider the task of TS transformation into TS of an elementary FT as a TS pre-processing. This transformation is needed for a linguistic description of a general FT of a given TS.

Let us use the following statement of the task. For a given numerical TS $x_t \in X, X \subset R^1, t = 1, 2, \dots, n,$ where n is a quantity of TS points, TS of an elementary FT must be derived in the form (2).

The algorithm of TS transformation into TS of an elementary FT includes following steps.

Step 1. The first step is smoothing of given TS by some technique to remove the noise with small intensities.

Step 2. Then a linguistic variable $\tilde{X} = \{\tilde{x}_i | \ i = 1,2, \dots, r\}$ must be determined on the set X and used to construct a fuzzy TS. We propose the following formula to construct each fuzzy TS value (r is a quantity of fuzzy values):

$$\tilde{x}_t = \tilde{x}_s, s = argmax\big(\tilde{x}_i(x_t)\big), i = 1,2, \dots, r. \quad (3)$$

Step 3. On the third step, using a fuzzy TS $\tilde{x}_t, t = 1,2, \dots, n$, TS of an elementary FT is derived. Let s1 be a number of a linguistic value for $\tilde{x}_{t-1}$ and s2 be a number of a linguistic value for $\tilde{x}_t$, where s1, s2 belong to the ordered set {1,2,…,r}. Below we recall [18] the following expressions to output the each value of TS of elementary FT for any time points t=2, 3, … ,n:

$$\tilde{v}_t = \begin{cases} \tilde{v}_1, s_1 > s_2 \\ \tilde{v}_2, s_1 = s_2 \\ \tilde{v}_3, s_1 < s_2 \end{cases} \quad (4)$$

$$\tilde{a}_t = \tilde{a}_g, \tilde{a}_g(|x_t - x_{t-1}|) \geq \tilde{a}_j(|x_t - x_{t-1}|) \quad (5)$$
$$\forall j, g = 1,2, \dots, (r-1),$$

where $\tilde{v}_1$ = "decrease", $\tilde{v}_2$ = "stability", $\tilde{v}_3$ = "increase".

C. The technique for linguistic description of a TS general fuzzy tendency

To identify a TS behavior in linguistic terms we propose extended heuristic algorithm of TS general tendency identification, described previously in [18]. The proposed algorithm extended algorithm given in [18] by performing more detailed linguistic description of TS behavior and by using TS transformation shown in Section B. We split the linguistic description of TS general tendency “chaos” given in [18] into linguistic descriptions "fluctuation with growth", "fluctuation with fall" to derive more information about TS behavior.

A given numerical TS Z is the input and the linguistic value B(Z)∈{"decrease", "stability", "increase", "fluctuation", "fluctuation with growth", "fluctuation with fall"} is the output of extended algorithm of a general tendency identification. The linguistic values "stability" and "fluctuation" are corresponded to behavior of stationary TS, while other ones give the linguistic description of non-stationary TS. Below extended algorithm of identification of TS general tendency, based on the notion of elementary FT is described.

Step 1. Transform a given TS Z into TS of elementary FT, using the pre-processing algorithm (see Section B).

Step 2. Defuzzify the intensities of elementary FT for each time point $t = 2,3, \dots, n$: $a_t = deFuzzy(\tilde{a}_t)$.

Step 3. Calculate the cumulative intensity for each type of elementary FT for $t = 2,3, \dots, n$:

if ($\tilde{v}_t$= ”Increase”), then $S_{inc} = S_{inc} + a_t$,

if ($\tilde{v}_t$= ” Decrease”), then $S_{dec} = S_{dec} + a_t$,

Step 4. Define the type of the general FT B (Z) = "Stability", if ($S_{dec} = 0$ and $S_{inc} = 0$), otherwise Step 5.

Step 5. Define the type of the general FT B (Z) = "Increase", if $S_{inc} > (2 * S_{dec})$, otherwise Step 6.

Step 6. Define the type of the general FT B (Z)="Decrease", if $S_{dec} > (2 * S_{inc})$, otherwise Step 7.

Step 7. Define the type of the general FT B (Z) = "Fluctuation", if $((0.85 * S_{dec}) < S_{inc} < (1.15 * S_{dec}))$ or $((0.85 * S_{inc}) < S_{dec} < (1.15 * S_{inc}))$, otherwise Step 8.

Step 8. Define the type of the general FT B (Z) = "Fluctuation with growth", if ($\tilde{v}_t$= ”Increase”) , otherwise Step 9.

Step 9. Define the type of the general FT B (Z) = "Fluctuation with fall", if ($\tilde{v}_t$= ” Decrease”).

In the Section 4 it will be shown that a linguistic description of TS general fuzzy tendency is useful to choose the best TS forecasting model.

IV. Two step TS model selection technique

In that Section a combination of linguistic and numerical criteria is used in two step selection of the best model from the set of optimal ones in their class. As was mention in Section 2 we will take in account three parts for given TS: (1) in sample, (2) out of sample and (3) real forecasts.

Let L be a length of given numerical TS Y and $L = L_1 + L_2$. For given TS Y we will distinguish following parts:

a) Y_1 is a part of TS with length L_1, known for a model and used to train on the in sample values of Y.

b) Y_2 is a part of TS with length L_2, known for a model and used to test on the out of sample values of Y.

c) Y_3 is a part of TS with length L_3, unknown for a model and is real forecasts of given Y with the horizon equal L_3.

d) Y_4 is a part of TS with length $L_4 = L_2+L_3$.

Let B(Y) be a linguistic description of behavior of a given TS Y with the length L and $B(Y_4)$ be a linguistic description of behavior of predicted parts of TS Y, which are performed by the algorithm of TS general tendency identification, described in Section 3.

At the first step to select the adequate TS models from the given set of TS models with different types and orders the linguistic criterion of general FT similarity is used [18]:

$$K_{ft} = \begin{cases} 1, if\ B(Y) = B(Y_4) \\ 0,\ otherwise \end{cases} \quad (6)$$

where $B(Y)$, $B(Y_4)$ are the outputs of the algorithm, presented in Section 3 for time series Y and Y_4 respectively.

The model for which $K_{ft} = 1$ is considered as an adequate one and is adding to subset of adequate models.

At the second step of TS model selection we use the numerical criterion of similarity of TS model accuracy in sample and out of sample, based on criterion SMAPE [18] to choose the best model from the subset of adequate models:

$$K_{SMAPE} = \frac{abs(SMAPE_1 - SMAPE_2)}{\max(SMAPE_1, SMAPE_2)},$$

where $SMAPE_1$ is in sample average error, $SMAPE_2$ is out of sample average error, computed as follows:

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2},$$

where F_t is t-th forecasted value, A_t is t-th actual value, and n is the number of TS points. We assumed that the value of K_{SMAPE} has to be less than 0.3. In results only one fuzzy time series model with minimum K_{SMAPE} is considered as the best forecasting model.

V. Results and conclusions

In this study the linguistic description of TS behavior was used and applied to select the best fuzzy TS model in real forecast, when the TS points were unknown for TS models. This linguistic description of TS behavior, derived by extended algorithm of identification of TS general tendency, was used to calculate linguistic criterion.

The study of efficiency of that extended algorithm is focused on three basic fuzzy TS models: model with fuzzified TS values [5], model with fuzzified first differences of TS values [7] and model, based on the elementary fuzzy tendency (FT) [18]. These models were chosen, as they are not good in real forecasts often.

The testing dataset consisted of 91 time series from banking, web hosting, and health domains [http://irafm.osu.cz/cif/main.php]. The frequencies of the time series (i.e. the frequencies at which the measures were reported) were daily, monthly, quarterly, and yearly. The forecasting horizons (i.e. the number of future values to be forecasted) was set by taking into account the length of the time series and its frequency: horizon differs from 4 for short yearly time series to 180 for quite long daily time series. These TS available at CIF-2015 competition [http://irafm.osu.cz/cif/main.php].

In comparison with the traditional fuzzy TS models evaluation, when a numerical criterion was used only, the applying of two step model selection with extended algorithm of TS behavior identification, showed that prediction error in real forecasts was decrease from 0.3421 to 0.1662 (SMAPE (Y_3) was calculated, when real values of TS of CIF–2015 were open).

As a benchmark, well-known forecasting models (ARIMA, ETS, the Theta method, and Random Walk) were used on the competition data set CIF–2015. Table 1 shows the results of real forecasts sorted by mean and standard deviation of SMAPE(Y_3) in ascending order (lower values are better). The result of applying of combination of fuzzy and numerical criteria to choose fuzzy forecasting model (named Proposed method) demonstrate not bad accuracy in comparison with well-known forecasting methods. Moreover, fuzzy TS models with two step selection algorithm produced the good accuracy in average (SMAPE(Y_3)=0.126), when TS behavior was described as “increase”.

TABLE I. The results of the CIF-2015 competition sorted by Mean and Standard Deviation of SMAPE

Method	*Mean*	*Std. Dev.*
ETS	0.12760	0.11883
ARIMA	0.12906	0.12323
Theta	0.13043	0.13221
RW	0.14216	0.13537
Proposed method	0.16621	0.15251

However, the linguistic criterion is not very good when the TS frequency is very low. In this case it is difficult to define the similarity of linguistic descriptions using three parts of TS, described in the paper. The future study will be aimed at development and study of TS model selection from the given set to improve efficiency and accuracy of forecasting results for TS with very low frequency. Another aim of future work is

to apply algorithm of identification of TS behavior to select appropriate models with known properties.

ACKNOWLEDGEMENTS

The authors acknowledge that this paper was supported by by the projects no. 16-07-00535 of the Russian Foundation for Basic Research.

REFERENCES

[1] J. D. Hamilton, Time series analysis. New Jersey: Princeton University Press, 1994.May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro–Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg

[2] Box, G. and Jenkins, G., Time Series Analysis: Forecasting and Control, Holden- Day, San Francisco, 1970.

[3] S. Makridakis, S. Wheelwright, and R. Hyndman, Forecasting: methods and applications. USA: John Wiley & Sons, 2008.

[4] Anderson T. W. Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc., 1971.

[5] Song, Q. Fuzzy time series and its models / Q. Song, B. Chissom // Fuzzy Sets and Systems. 1993, vol. 54, pp. 269–277.

[6] Chen, S. M. Forecasting enrollments based on high-order fuzzy time series / S.M. Chen // Cybernetics and Systems: An International Journal. 2002. – vol. 33, pp. 1–16.

[7] Hwang, J.R., Chen S.M. and Lee, C.H. Handling forecasting problem using fuzzy time series, Fuzzy Sets and Systems, 100, 217-228, 1998.

[8] Novak, V., Stepnicka, M. Dvorak, A., Perfilieva, I. and Pavliska, V. Analysis of seasonal time series using fuzzy approach, International Journal of General Systems, 39, pp.305–328, 2010.

[9] Pedrycz W., Chen S.M. (Eds). Time Series Analysis, Modeling Applications, A Computational Intelligence Perspective (e-book Google). – Springer–Verlag, Berlin Heidelberg, 2013. – (Intelligent Systems Reference Library, Vol. 47). – 354–412 pp. 2013.

[10] Bardossy, A. Note on fuzzy regression / A Bardossy // Fuzzy Sets and Systems. – 1990. – Vol. 37 – pp. 65–75.

[11] Wagner N., Michalewicz Z., Schellenberg S., Chiriac C., Mohais A. Intelligent techniques for forecasting multiple time series in real-world systems, International Journal of Intelligent Computing and Cybernetics, Vol. 4 Iss:3, pp.284–310, 2011.

[12] Yarushkina, N., Afanasieva, T., Perfilieva, I. Computational Intelligence in the analysis of time series. – ID "FORUM": INFRA–M, Moscow, 2012, 160 pp.

[13] Tanaka, H. Linear regression analysis with fuzzy model / H. Tanaka, S. Uejima, K. Asai // IEEE Transactions on Systems, Man and Cybernetics. – 1982. – Vol. 12(6). – pp. 903-907.

[14] Diamond, P. Fuzzy least squares / P. Diamond // Information Sciences. – 1988. – Vol. 46(3). – pp. 141–157.

[15] Tsenga, F. M. Fuzzy ARIMA model for forecasting the foreign exchange market / F. M. Tsenga, G. H. Tzengb, H. C. Yu // Fuzzy Sets and Systems. – 118 (2000) 9–19.

[16] Khashei, M. Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks / M. Khashei, M. Bijari, G. Rassi Ardali //Neurocomputing, 72(4), 2009, 956-967.

[17] Perfilieva, I., Yarushkina, N., Afanasieva, T., Romanov A. Time series analysis using soft computing methods, International Journal of General Systems, 42: 6, 687-705, 2013.

[18] T. Afanasieva, N. Yarushkina, M. Toneryan, D. Zavarzin, A. Sapunkov and I. Sibirev Time series forecasting using fuzzy techniques / // International Joint Conferece IFSA-EUSFLAT (16th World Congress of the International Fuzzy Systems Association (IFSA), 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)), June 30th-July 3rd, Gijon (Asturias) Spain), 2015. – P. 1068–1075.

[19] F. J. J. Santos and H. A. Camargo, "A hybrid forecast model combining fuzzy time series, linear regression and a new smoothing technique," in Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015), ser. Advances in Intelligent Systems Research. Gijo´n: Atlantic Press, 2015.

[20] L. Maciel, A. Lemos, R. Ballini, and F. Gomide, "Adaptive fuzzy c- regressiion modeling for time series forecasting," in Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT 2015), ser. Advances in Intelligent Systems Research. Gijo´n: Atlantic Press, 2015.

[21] Perfilieva, I. Fuzzy transforms: Theory and applications / I. Perfilieva // Fuzzy Sets and Systems – 2006. – Vol. 157.

[22] Mamdani, E.H., Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies, 7(1):1–13, 1975.

[23] J. Kacprzyk, A. Wilbik, and S. Zadroz´ny, "Linguistic summarization of time series using a fuzzy quantifier driven aggregation," Fuzzy Sets and Systems, vol. 159, no. 12, pp. 1485 – 1499, 2008.

[24] V. Nova´k, V. Pavliska, M. Stepnicka, and L. Stepnickova´, "Time series trend extraction and its linguistic evaluation using f-transform and fuzzy natural logic," in Recent Developments and New Directions in Soft Computing (Proc. of 2nd World Congress on Soft Computing), ser. Studies in Fuzziness and Soft Computing. Springer, 2014, vol. 317, pp. 330–339.

A FC^{2}TA Model for Congestion Management in Confidence AODV Protocol

Rahim Rashidi, Farugh Katani
Department of Computer, Boukan Branch, **Islamic Azad University**, Boukan, Iran
Rahim.rashidy@gmail.com, f_katani828@yahoo.com

Saeid Usefpour, Faroogh Ashkooti
Department of Computer, Mahabad Branch, **Islamic Azad University**, Mahabad, Iran
Saeid.usefpour@gmail.com, ashkooti_f@yahoo.com

Abstract— **Ad-hoc networks, as well as many wireless networks and wired networks applied here to do and the proper function of routing data packets forwarding, maintenance and update routing information, the security need. Mobile ad hoc networks due to the special properties of operations and placement of often unstructured meshes, or self-organizing ad-hoc networks are known. In these networks in order to extend the communication range of nodes, more than a step, configure routing algorithms are used. In fact, security was a prerequisite for the proper functioning of network operations and without it there is no guarantee for the proper conduct of the exercise and Attackers can easily disrupt the integrity of the network. The trust model proposed by Pirzada and his colleagues route based on at least step is not created. It will be nodes in the path which other nodes are more confident, the old Confidence models can be use that have three part of the Confidence section, Share section and Merge section. So there are high chances of congestion in nodes with confidence. In AODV routing protocol congestion management is vital, so in this paper, a FC^{2}TA model is presented. The FC^{2}TA model increase nodes energy, network stability and balancing on the nodes that have high confidence. Here we apply the FC^{2}TA model to the latest version of AODV. Because of that these protocols constantly improving and now in the development phase. The simulation model is done in Fuzzy Logic Tools in MATLAB.**

MANTE, fuzzy system, congestion control, Confidence

I. INTRODUCTION

Ad-hoc networks are said to be temporary networks that are created for a specific purpose. In fact, the wireless network that the nodes are moving. The major difference Ad-hoc networks with common 802.11 wireless networks in Ad-hoc network is a collection of wireless mobile nodes without any central infrastructure access point or base station for wireless data transmission in the interval specified are connected.

Ad-hoc networks, as well as many wireless networks and wired networks applied here to do and the proper function of routing data packets forwarding, maintenance and update routing information, the security need. Ad-hoc network nodes to perform key management in a managed environment need. In fact, there should be one or several Confidenceed center to new nodes in the network and register their malicious nodes from the network line. Thus, based on healthy nodes existing network security to provide, because malicious nodes are located in the revocation list. A number of routing protocols use of cryptographic and Confidence to reliability the routes [1]. The presented model by Pirzadeh can be functionally divided into the three components: Confidence section, Share section and Merge section [11]. This protocol enables nodes to forward packets onto Confidence routes, rather than the default MIN hop routes. Therefore, there is the probable of congestion on the nodes which have higher Confidence. A developed Confidence protocol namely FC^{2}TA model has been presented in this article. the suggested model considers energy, to ensure stability and balancing on the network with nodes that have MAX confidence. Now to show the performance of the FC^{2}TA model, it is done on the Confidence AODV protocol. The simulation model in MATLAB shows the improving performance of AODV [11].

II. THE PROPOSED MODEL FC^{2}TA

The four components of the proposed system are as follows: Confidence section, Share section, merge section and congestion management section. Operating confidence, the confidence levels of the events that have been experienced by a node directly obtains. [8]. Share section shares Confidence data with other nodes in network. The composition of cumulative Confidence in a node based on information received from the Confidence and Share section separately [7, 10]. Based on the Fig 1 three section are associated with Fuzzy Block. The Fuzzy section (FC^{2}TA) based on the receive data from the Merge section dose balancing on the nodes with MAX confidence. The final decision is made by Fuzzy Blocks. The proposed model design is shown in Fig 1.

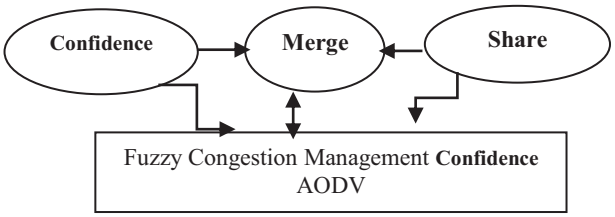


Fig. 1. Schematic of the FC^{2}TA Model

A. confidence section

The Confidence section consists of Row Extraction, Row Description and Row Calculation. Extraction of confidence is done in Passive mode, i.e. do not need to search for packages. Essential information about other nodes can be collected by

checking local traffic network. Events that can be measured and properly recorded in Passive mode is the following.

TABLE I. TYPE OF EVENTS FOR RECORD

Data Frames sent	Frames Expect
Data Frames expect	Control Frames Sent
Streams recognize	Control Frames Expect

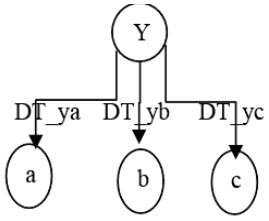
In the above event information acquired split into one or more Confidence sections [7, 11].

Row Description: Never Confidence networks are not static, because Confidence relationships constantly changing due to the dynamic topology. Time interaction with other nodes is important. It is essential that Confidence be displayed as a continuous range to distinguish between nodes by comparing the levels of Confidence. Confidence generally describes a continuous process, so the display is discrete enough to display confidence. As was explained in the first distributed Confidence model, Confidence using a six-valued range of no confidence to Confidence the symbolism [9,11].

Row Calculation: In the calculation of confidence, and a little weight to monitor events previously allocated. The weight of all nodes dynamically, depending on the situation and allocate themselves. This weight in a continuous range from -1 to +1 shown that 1 + 1 non-serious and important measures. Then to calculate a position of trust classes together in a special node is dumped directly confidence. the direct Confidence in node y, by node x, as Txy defined and as follows:

$$\frac{\mu^{1}_{T_{xy}}}{i} = \sum \min\left(\frac{\mu^{1}_{w_{xy}}}{i}, \frac{\mu^{1}_{T_{xy}}}{i}\right)$$

Where Wxy (i) is the weight of the ith Confidence section of node y to node x and Txy (i) is the situational Confidence of node x in the ith Confidence category of node y [28]. The total number of classes n protocol confidence in the scenario that is applied depends on the model. Each node in the network itself to place confidence rate in relation to other nodes in the network maintenance. Direct trust and confidence in the current situation to display the status of nodes and dynamically updated maintenance is dumped. from the view of node y, the situational and direct Confidence table is shown in table 2.



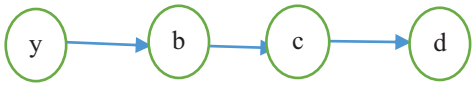
Node	Straight Confidence
a	DT_ya
b	DT_yb
c	DT_yc
…	…

Fig. 2. Straight Confidence Links.

The derived and cumulative Confidence values are preserved by the Merge and dynamically updated whenever a modification occurs either in the situational and Straight Confidence table or the Share table. The direct, recommendation and cumulative Confidence table maintained by node y is exposed in bellow. [12]

$$\frac{\mu^{1}_{T_{yb}}}{i} = 1 - \left(1 - \mu^{i}_{T_{ya}}\right)$$

$$\frac{\mu^{1}_{T_{yb}}}{i} = 1 - \left(1 - \mu^{i}_{T_{yb}}\right) \qquad \frac{\mu^{1}_{T_{yb}}}{i} = 1 - \left(1 - \mu^{i}_{T_{yc}}\right)$$



$$\frac{\mu^{1}_{T_{yd}}}{i} = 1 - \left(1 - \mu^{i}_{T_{yb}}\right)\left(1 - \mu^{i}_{T_{bc}}\right)\left(1 - \mu^{i}_{T_{cd}}\right)$$

B. Share Section

Most events experienced by a node occurs through the interaction between neighboring nodes. These events help to build trusting relationships directly next door. Although a very small number of events directly between nodes that are experiencing is more than a step away. It is imperative that when calculating confidence in the other nodes, the nodes share must be considered. Offers from reputable agent node to other nodes request is based on trust. Two likely mechanisms that could be used to play a direct confidence values the share of integrated trade and exchange are independent reputation

C. Merger Section

The merge of the values of self-confidence and share agent receives. To calculate the amount of cumulative confidence a target node, we may computational methods, and accepted by Beth will change. Although reliable computational method may be used to combine the values of trust [4].

III. CONGESTION MANAGEMENT USING FC²TA MODEL

A. Fuzzy system

The theory of fuzzy sets and fuzzy logic Prof. Lotfizadeh first in a thesis called “fuzzy - Information and Control introduced in 1965”. His primary goal at the time, developed a model to describe the process more efficient natural language processing. He and reform concepts such as fuzzy, fuzzy events and fuzzy numbers enter the mathematical sciences and engineering. I ask <is cloudy or sunny?> First, partly cloudy. I ask "Is everything you said to me yesterday, right?> First, much of it was true. We often fuzzy logic we use in everyday life. The fact is that the world of zero and one, abstract and imaginary world. Absolutely right or absolutely wrong subject is rarely, because in the real world most of the time, everything is neatly in place. Fuzzy logic (fuzzy logic) formulated a new solution to provide qualitative parameters. In fact, with this new logic, the possibility of progress in robotics and computer sciences has been provided.

Fuzzy logic is a specific place among other sciences, even the social and economic sciences; find and perhaps many social and economic parameters with this logic is formulated. In fact, a new method for formulating concepts and sensory quantity and quality of offers. What man has his theories are based, was that only the "quantity" to be formulated and qualitative concepts and sensory and imprecise and even vaguely like a nice, long, hot, cold, old, young, and so forth Can be formulated! If the human brain by taking into account various factors and based on deductive thinking, especially for the reasonable quantities say. Today, dishwashers and many other

household appliances also use this technique. Fuzzy Logic frequency applications are in automotive industry. For example, ABS brakes and uses fuzzy logic in some cars. One of the most famous examples of the use of fuzzy logic in global logistics systems, network monorail (single-rail train) Tokyo in Japan. Load displacement motor and other systems, like elevators also use fuzzy logic. Fuzzy inference systems may be used as a feedback control design approach in places where no strict control method cannot be used is observed. Control algorithm as a set of IF-THEN fuzzy rules have been arrested. Fuzzy expert system for non-linear control systems may be too complex is popular [2,6].

B. Reasons for using Fuzzy inference systems

For what we use fuzzy logic? Below is a list of general observations about fuzzy logic is:

Fuzzy logic can easily be understood conceptually. Mathematical concepts are fuzzy reasoning is very simple. What made that nice fuzzy "naturalness" approach rather elusive complexity of his hand. Fuzzy logic is flexible. For any given system, adding more functionality without restarting from scratch is not easy. Fuzzy logic is tolerant of imprecise data. If you look around carefully to see that everything around us imprecise. Moreover, the more things even in the imprecise carefully inspected. Fuzzy reasoning builds this understanding into the process rather than remove it from the process. Fuzzy logic can model the non-linear functions with any degree of complexity. You can get a fuzzy system to match any set of input data exit. This process, especially through techniques such as adaptive neural fuzzy inference systems to easily implement adaptive. Fuzzy logic can be built on the experience of experts. Unlike neural network training data to produce models that are ambiguous impenetrable, Fuzzy logic gives you permitted to rely on the experience of people who already understand your system, trust. Fuzzy logic can be controlled with conventional techniques of combination. Fuzzy systems are not necessarily replacing conventional control measures. In many cases fuzzy systems complement and strengthen them and make them easier to implement.

C. Where and how fuzzy systems are used?

Knowledge-based systems are fuzzy systems or rules. The heart of a fuzzy system is a knowledge base that is composed of fuzzy IF-THEN rules. A rule fuzzy if-then statement if - then that some words are marked by continuous membership functions. For example, consider the following phase: if the car speed is high then less force to enter the gas pedal. Fuzzy logic provides a method for processing non-deterministic events, exactly what the nature and everyday life with which I am associated. In fuzzy logic and approximate working with the uncertain value; a range of possibilities that may happen. Fuzzy logic Boolean logic is the logic of binary or binary [12].

D. Fuzzy inference systems implementation

Fuzzy Inference System, systematic process to become a knowledge base to provide a non-linear mapping. For this reason, knowledge-based systems (fuzzy system) is used in engineering applications and decisions. A fuzzy system has the following components:

Fuzzfier a numeric value in the input variables to make a fuzzy set. Fuzzy rule base is a collection of if-then rules. Fuzzy inference engine inputs and outputs with a series of actions to make. Defuzzification fuzzy output to a certain number of turns.

An ad_hoc network is a huge spreading complex system with so much difficulty nonlinear, time-varying and wrong behavior. Dynamic or static modeling of such a system is really complicated to control. Measuring the partial network mode, and latency is often weak. In order to enable wireless image and multimedia communications, to communicate information over wireless image must be addressed. Network congestion control is a matter of high priority, especially in increasing size, low speed (broadband) network gradually integrated. Recent solutions in existing networks are becoming useless and simply cannot scale up. Traditional approaches to congestion management for TCP / IP and ATM separately action [1].

Two units together under the rule base of database and knowledge base (Knowledge base) are known. Fuzzy inference systems can be divided into three classes, "Mamdani, Sugeno and Takagi divided. Many FIS of Mamdani type involved in this type of fuzzy set out to predict. Fuzzy Inference System to successfully operate the system that can be easily achievable analytical models or model is very complex and non-linear control is used. In this paper, a fuzzy model is confidence, which is applied in hop-by-hop Ad Hoc networks [2].

E. Fuzzy inference system

In this paper FC^{2}TA model is based on a queue management scheme to provide transmission speed control in nodes with high confidence. The FC^{2}TA model is shown in Fig 6. Speed node is confident that with the information returned by the source node is adjustable. Confidence source node by node, end node shift to the node system is produced. Units are confident this information to the next node current confidence back. Feedback will be given of a buffer state node confident in the next scan, and the previous node how to change the speed of transmission. Units of confidence, after receiving feedback information to adjust your speed to MAX and MIN. The proposed fuzzy model, looks at the current state of the buffer queue length and speed ahead.

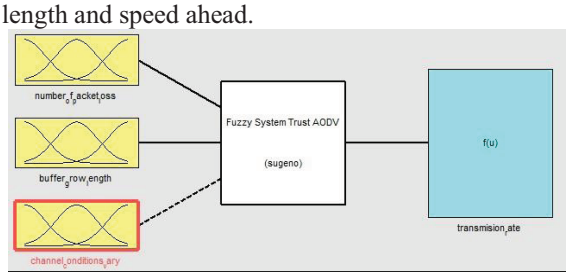


Fig. 3. Input variable for FC^{2}TA in MATLAB

Accordingly, the FC^{2}TA model suggested will be more effective than models they only have threshold. In the view of the membership function. Membership functions can be changed by using existing elements. In books and articles fuzzy talk usually comes from three systems:pure fuzzy systems,

Takagi-Sugeno fuzzy systems and Kang (TSK) and systems with fuzzy defuzzification mechanism building, summarize these three systems we describe. The main problem is that in terms of pure fuzzy system inputs and outputs are fuzzy sets (words in a natural language). While the engineering systems, inputs and outputs are placeholders with real values. To resolve this problem, Takagi-Sugeno and Kang fuzzy systems have introduced another type of inputs and outputs are placeholders with actual values.

F. Rule descriptions

After determining the membership function must define Rule to do this, enter the edit menu, and begin to define the roles. In this page through the Edit menu to enter the number of inputs and outputs are importing inputs and outputs. With it comes in the form below. The Make of FC^{2}TA model is perform into three steps: Create the Input variables, Create the membership functions, Create the rules for FC^{2}TA

IN #1: Packets_Loss	Fix (Fx).
MIN_S - Bellow 0.3%	Growth slow (-GS).
MID_A - Range 0.3-0.6%	Growth Fast (+GF).
MAX_H - Over 0.6%	IN #3: Channel_Conditions
IN #2: Buffer_Length	MIN - Under 0.3% (L)
Reduction Fast (-RF).	Mid - Range 0.3-0.6% (M)
Reduction Slow (-RS).	MAX - Over 0.6% (H).
OUT2: Sliding Win (L=0 ... H=10).	OUT1: Broadcast speed (0 - 400 Mbps)

TABLE II. LINGUISTIC RULES MATRIX

		Queue_Length or **Channel_Status**				
		-RF	-RS	Fx	+GS	+GF
		MIN		Mid	MAX	
Packet_loss	MAX_H	395 Win=9	355 Win=7	215 Win=4	95 Win=3	62 Win=2
	MID_A	395 Win=9	310 Win=7	315 Win=6	205 Win=4	62 Win=2
	MIN_S	395 Win=9	395 Win=9	355 Win=8	225 Win=5	110 Win=3

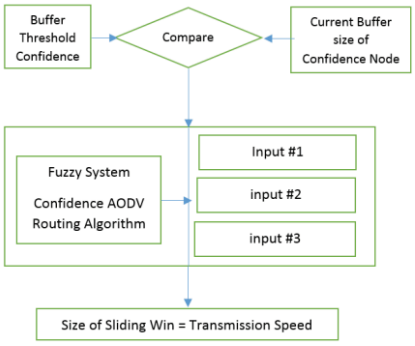


Fig. 4. Block diagram of FC^{2}TA

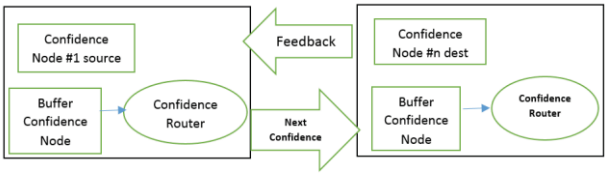


Fig. 5. The channel speed control using FC^{2}TA

The syntax of a rule base (IF–THEN) consists of two parts: First, the linguistic rules are set (table 3) and afterwards, membership functions of the linguistic values are determined.

IV. APPLY FC^{2}TA MODEL TO AODV PROTOCOL

AODV routing protocol is an On-Demand routing protocol where all paths are discovered only when needed and only during the time that used to be kept. Flooding paths are discovered during a search process during which a route to the destination network nodes are questioned. When a node has discovered a way to track it back to ninety departures that the question had to be my report. AODV is designed to fulfill the objectives as follows:

Minimal overhead control, Minimal processing overhead, Several-routing capabilities, Dynamic topology maintenance Freedom from the ring.

Because resources are scarce Ad hoc mobile networks AODV attempts to control overhead by limiting Updates alternate route as well as the use of the On-Demand messages to a minimum. To minimize processing overhead, messages, and AODV simple structure requiring little computing. The route discovery process when a source node needs a route to a destination node and route in the routing table is not valid, a source node route request packet (RREQ) to the destination node is broadcast. When each node receives RREQ, an entry in a reverse direction to the source node to create or update the routing table and if a valid path to the destination node in the routing table, not RREQ is broadcast again. When Flooding RREQ packets from the source node to the destination node is reached, the input destination node to create or update the reverse route and a route reply packet (RREP) is an enhanced Shmarhtrtyb in the reverse direction unicast.

When RREP to the source node and turns along the way, on the contrary, a forward path to the destination or refresh and communication starts. In the process of keeping track of each node periodically broadcast a packet Hello for local connections and RREP with TTL = 1, such as Hello packets are broadcast. When ninety packet from a neighbor and does not get in a few seconds, assumes that neighbor connection is broken. In addition, when a broken connection node to a neighbor notification based on its MAC layer, the path to the destination node detects a failure The next step route, the node's neighbors. When the node detects a connection failure, is close to the destination node needs a new direction to the process as local repair (Local Repair) is called. We modify this rule on based FC2TA model and have assigned the direct fuzzy Confidence value as cost of each HOP. FC2TA Model would guarantee that the load is balanced in the Ad Hoc network. [1,10]. Here we apply the FC2TA model to the AODV routing protocols. For the reason that these protocols are now in the development phase.

V. FUZZY LOGIC TOOLBOX

In making using fuzzy logic toolbox in MATLAB's graphical user interface is only available in the toolbox we use fuzzy logic. Keep in mind that you can use to build command line system, but making the systems easier and more

understandable graphical form. Five major graphic tools to create, view and edit the system is provided in the toolkit: Fuzzy Inference System FIS Editor Editor or Editor, Membership function editor (Membership, Function Editor), Editor Rules (Rule Editor), Display Rules (Rule Viewer), Shop surface (Surface Viewer).

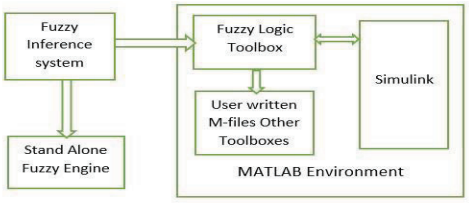


Fig. 6. Fuzzy logic in MATLAB

VI. Simulation and Results

After defining the rule of time to reach conclusions. To evaluate the results of the view menu options we chose rules and opened them on the output variables and results can be viewed. The results are illustrated in figure 7.

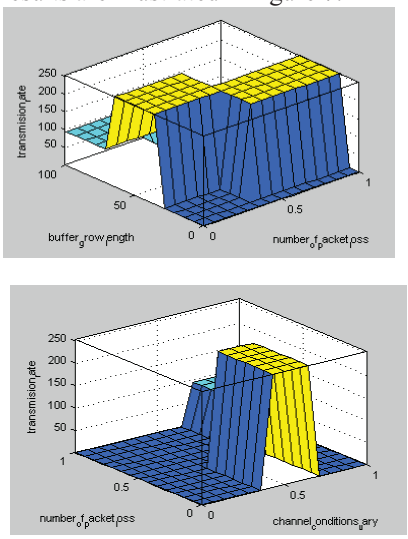


Fig. 7. Adjusting the transmission speed using FC^{2}TA

VII. Rules of the Fuzzy System

Once the input and output fuzzy sets and membership functions are assembled, the rules are then formulated. The rules are formulated based on the input parameters (Number of packets loss, Buffer develop length, channel conditions vary) and the output i.e. level of sliding window and transmission speed. IF a is A THEN b is B, where a and b are the variables while A and B are the values determine by fuzzy system.

VIII. Conclusion

Due to the inherent nature Ad-hoc networks can be easily inferred that network performance is highly dependent on the behavior of network nodes. This means that if a node does not perform its duties properly, greatly compromising network performance efficiency and share vital information may be at risk. On the basis of some of the models proposed to secure the majority use logic and incorrect behavior of nodes on the background of their actions are examined. If this record is worse than a threshold applied to the average behavior of malicious node is detected. Of course, these decisions are somewhat relative and never absolute cannot be determined that any behavior that is correct or incorrect node his head.

AODV algorithm is based on DSDV routing protocol with the exception that the route only when necessary to reduce the Broad Casting. The only route discovery algorithm since the inception of the path between two nodes does not exist. So to say that reaction. In the original AODV, route hop is done according to MIN. In this case, the nodes in the path of the money to create more confidence. Therefore, there is a possibility of condensation on the nodes confidence. Accordingly, the temporary network congestion and packet loss have occurred that caused a long delay. A FC^{2}TA model has been suggested in this article. In this paper, in order to solving this problem, a technique of regulating the transmission speed to avoid buffer overflow of Confidence nodes using fuzzy model was presented.

Currently, congestion management, a way to prevent buffer overflow nodes confidence in using the system FC^{2}TA the paper suggested. The results show that confidence can be flexible node transfer speed to adjust the size of bandwidth.

References

[1] Rashidi, Rahim, et al. "Confidence routing protocol based on congestion control in MANET." Application of Information and Communication Technologies, 2009. AICT 2009. International Conference on. IEEE, 2009.

[2] Frankel, David S., Model Driven Architecture: Applying MDA to Enterprise Computing, OMG Press, Wiley Publishing, 2003.

[3] Clark N. Taylor, Sujit Dey, “Adative Image Compression for Wireless Multimedia Communication”, Electrical and Computer Engineering, University of California, San Diego, La Jolla, California, USA.

[4] A Pitsillides, Y.A, Sekercioglu, G. Ramamurthy, “Effective control of traffic flow in ATM networks using fuzzy explicit rate marking (FERM)”, IEEE Journal 15 (1997) 209.

[5] O. Rioul, M. Vetterli, “Wavelets and signal processing”, IEEE Signal Processing Magazine 1991.

[6] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, “Image coding using wavelet transform,” IEEE Trans. Image Process 1 (1992) 205.

[7] C. Lee, “Fuzzy logic in control systems: Fuzzy logic controller”, IEEE Trans. Systems 20 (1990) 404.

[8] Y.-C. Hu, A. Perrig, and D. B. Johnson. Ariadne : A secure on-demand routing protocol for ad hoc networks. Proceedings of MOBICOM 2002., 2002. 5.1

[9] M. Burmester and A. Yasinsac. Confidence infrastructure for wireless mobile networks.WSAES Transactions on Telecommunications, pages 377{381, 2004. 1.5.2, 1.5.2.1,1.5.4

[10] P. Ning and K. Sun. How to misuse AODV: A case study of insider attacks against mobile ad-hoc routing protocols. Proceedings of the 4th Annual IEEE Information Assurance Workshop, pages 60{67, 2003. 1.3.1

[11] T.Beth,M.Borcherding , and B.Klein. Valuation of Confidence in Open Networks. In Proceeding of the Tired European Symposium on Research in computer Security pages 3-18, springer 1994.

[12] Asad Amir Pirzade ,Confidence based Routing in Wireless ad-hoc network's . university of western Australia, 2007

[13] C. Lee, “Fuzzy logic in control systems: Fuzzy logic controller”, IEEE Trans. Systems 20 (1990) 404

Simulation of Identification Process of Parameters of Movement of the Wagon on Monorail Transport

Moroz N., Starostenko A.
Control systems laboratory
Moscow Institute of Physics and Technologies
Moscow, the Russian Federation

Gorbachev R.A., Kuznetsov N.A.
"GMK Innovations" LLC
Moscow, the Russian Federation

***Abstract* — the work is devoted to construction of mathematical model of movement of the passenger wagon and algorithm of identification of parameters of mathematical model, as well as demonstration of efficiency of this algorithm through simulation of the process of identification for selected mode of acceleration.**

***Index Terms*—mathematical model, monorail transport, law of motion control, identification of parameters of model, the method of least squares, a sequence of independent Gaussian variables.**

I. Introduction

Currently the richest and the most technologically advanced countries, such as the USA, Japan, Germany and China, use monorail transport systems to solve their traffic problems. This type of transport is actively developing in Malaysia, Singapore, the United Arab Emirates and the others [1-3].

The primary advantage of monorail transport systems is that this type of transport does not require any allocation of "scarce" areas and fits perfectly into the existing infrastructure, has a high level of comfort, is more economical in terms of cost of construction and operation in contrast to the metro. In addition, monorails have a high degree of curvature of the track compared to the underground transport systems and railway, and are well suited for rough terrain.

The speed of the monorail in theory can exceed the speed of conventional rail trains significantly. However, monorail transport often moves at a low speed and cannot cope with the required passenger traffic, which indicates the need to find more efficient and reliable engineering solutions, taking into account all local conditions as well as the automation of trains at a higher speed with high level of safety.

II. Problem statement

Let us assume that the aim of control of wagon is to reach identity or minimum deviation between the current value of speed of the wagon and the set one. The set speed is a function of the distance traveled by the wagon from the previous station and is determined by conditions of safety and efficiency, taking into account the profile of the road cloth (descents, climbs, turns etc.). To determine the current speed level, sensors are installed in the wagon. The set speed is estimated by measurement of the wagon's location.

Let us assume that the actual information about the process of wagon movement comprises measured values of traction and braking forces. These forces constitute the control action on the wagon.

III. Mathematical model of wagon movement

A random process that meets the system of difference equations can describe the movement of the wagon [4]:

$$V_{n+1} = V_n + \frac{g}{1000}\left(\frac{F_n - B_n}{Pg} - A_0 - A_1 V_n - A_2 V_n^2 - i_n\right)\Delta t + \varepsilon_{n+1},$$
$$X_{n+1} = X_n + V_n \Delta t,\ n = 0, 1, 2, ..., V_0 = 0,\ X_0 = 0. \quad (1)$$

F_n – traction force, and B_n – braking force assumed to be known quantities, V_n – wagon speed, X_n – distance from the wagon to the previous station, Δt – time sampling step. i_n – slope of the track, assumed to be known. ε_n, n = 1, 2 – a sequence of independent Gaussian random variables with zero mean $M\varepsilon_{1,n} = 0$ and equal variances $M\varepsilon_{1,n}^2 = \sigma^2$.

The force of aerodynamic resistance of the wagon (when there is no wind) is equal:

$$A_2 V_n^2 \cdot gP = c\frac{\rho}{2} V_n^2 s, \quad (2)$$

c – wagon resistance factor, s – cross- sectional area of the wagon, ρ – air density. All these quantities are assumed to be constant. Hence $A_2 gP = c\frac{\rho}{2}s$ is fixed and known value for the wagon.

Next we are considering the effect of wind on aerodynamic resistance to movement of the wagon. The speed of its motion relative to the air mass is equal to the algebraic sum of the speed of the wagon relative to the rail and the projection of the wind speed in the direction of movement of the wagon. Therefore, the mathematical model of the aerodynamic force will be:

$$F_{аэр} = c\frac{\rho}{2}s\left(V_n + V_{ветра}\right)^2 = c\frac{\rho}{2}s\left(V_m^2 + 2V_n V_{ветра} + V_{ветра}^2\right). \quad (3)$$

Taking into account Eq.3 the mathematical model of the motion of the wagon in the presence of wind will have the form:

$$V_{n+1} = V_n + \frac{1}{1000P}(F_n - B_n - c\frac{\rho}{2}sV_n^2 - (A_0 Pg + c\frac{\rho}{2}sV_{wind}^2) - (A_1 Pg + c\rho sV_{wind})V_n - i_n Pg)\Delta t + \varepsilon_{n+1},$$
$$X_{n+1} = X_n + V_n \cdot \Delta t,\quad n = 0,1,2,...,\ V_0 = 0,\ X_0 = 0. \quad (4)$$

The measured values are V_n and X_n, known quantities are i_n – slope of the track, and g – the acceleration of gravity. The unknown quantities required evaluation are P, A_0, A_1,. A_0 and A_1 may vary during the passage of the wagon between the stations as their values depend on the geometry and conditions of the track, wind speed and direction etc. The rate of change of these parameters is assumed to be slow. The weight of the wagon with passengers P is constant between adjacent stations and is measured in tones.

IV. The Identification of Parameters of Mathematical Model of the Wagon Movement

For the estimation of the unknown parameters of the model P, A_0, A_1 we are applying the results of the method of least squares (OLS) [5,6], the implementation of which in relation to this problem is as follows. Known forces (F_n - B_n), $n = 0, 1, \ldots, k-1$ are applied to the wagon, located on the station ($V_0 = 0$, $X_0 = 0$) at regular intervals Δt to reach the set value of speed. At the same intervals the speed of the wagon is measured V_n^{msr}, $n = 1, 2, \ldots, k$.

The Eq. 4 will have the form:

$$F_n - B_n - c\frac{\rho}{2} sV_n^2 - \left(A_0 Pg + c\frac{\rho}{2} sV_{wind}^2\right) - \\ - (A_1 Pg + c\rho sV_{wind})V_n - i_n Pg = \\ = \frac{1000P}{\Delta t}(V_{n+1} - V_n - \varepsilon_{n+1}),\ n = 0, 1, \ldots, k-1. \quad (5)$$

For convenience, we introduce vectors:

$$\Theta = (\frac{1}{P}, \widehat{A_0}, \widehat{A_1})^T, \text{ где}$$

$$\widehat{A_0} = A_0 + c\frac{\rho}{2Pg} sV_{wind}^2,\ \widehat{A_1} = A_1 + c\frac{\rho}{Pg} sV_{wind},$$

$$\varphi(n) = (F_n - B_n, -c\frac{\rho}{2} sV_n^2,\ -g, -gV_n)^T,$$

$$Y(n) = \left(\frac{1000}{\Delta t}(V_2 - V_1) + i_1 g, \frac{1000}{\Delta t}(V_3 - V_2) + \right. \\ \left. i_2 g, \ldots, \frac{1000}{\Delta t}(V_{N+1} - V_N) + i_N g\right)^T = \\ (y(1),\ y(2),\ \ldots y(N))^T \quad (6)$$

In the new variables the Eq. 1 will have the form:

$$y(n) = \Theta^T \times \varphi(n) + e(n), n = 0, 1, 2, \ldots,$$

$$e(n) = \frac{1000}{g \times \Delta t}\varepsilon_{n+1}.$$

Predictive model of the method of least squares can be written as:

$$\hat{y}(n) = \Theta^T \times \varphi(n), n = 0, 1, 2, \ldots$$

The criterion to be minimized is:

$$\sum_{j=1}^{n}(y(j) - \hat{y}(j))^2 = \sum_{j=1}^{n}(y(j) - \Theta^T\varphi(n))^2.$$

by Θ to get an estimate $\widehat{\Theta(n)}$ of the vector Θ.

The algorithm for obtaining estimates can be represented by the following sequence of calculations. Let us write the first series of N measurements $Y(n) = (y(1),\ y(2),\ \ldots y(N))^T$ and the regression matrix corresponding to these measurements:

$$\Phi(n) = \begin{pmatrix} F_1 - B_1 & -\frac{\rho}{2} sV_1^2 & -g & -gV_1 \\ F_2 - B_2 & -\frac{\rho}{2} sV_2^2 & -g & -gV_2 \\ \cdots & \cdots & \cdots & \cdots \\ F_N - B_N & -\frac{\rho}{2} sV_N^2 & -g & -gV_N \end{pmatrix}. \quad (7)$$

Estimates $\widehat{\Theta(n)}$ of the vector Θ на based on the first set of N measurements are defined by the equation:

$$\widehat{\Theta}(N) = \left(\Phi^T(N)\Phi(N)\right)^{-1}\Phi^T(N)Y(N). \quad (8)$$

The sequence of estimates $\widehat{\Theta(n)}$ of the vector Θ can be rewritten recursively [7]:

$$\widehat{\Theta}(n) = \widehat{\Theta}(n-1) + \frac{1}{n}K(n) \times \left(y(n) - \widehat{\Theta}(n-1)^T\varphi(n)\right),$$

$$K(n) = \frac{R(n-1)^{-1}\varphi(n)}{1+\frac{1}{n}(\varphi(n)^T R(n-1)^{-1}\varphi(n)-1)},$$

$$R(n) = R(n-1) + \frac{1}{n}\left(\varphi(n)\varphi(n)^T - R(n-1)\right),$$

$$R(N) = \frac{1}{N}\Phi^T(N)\Phi(N). \quad (9)$$

When replacing 1/n in the Eq. 8 by a sequence of positive numbers μ (n) input sequence μ changes contribution of old estimates in OLS criterion in comparison with 1/n sequence.

V. The Simulation of the Process of Identifying the Parameters of the Mathematical Model of the Wagon Movement. The Solution of the Direct Problem.

Let us suppose that at the initial section of the wagon movement ($0 \le t \le 30$) the change in acceleration is subject to the following law:

$$a(t) = \begin{cases} 7\ м/с^2, 0 \le t \le 10 \\ 0\ м/с^2, 10 < t \le 20 \\ 7\ м/с^2, 20 < t \le 30. \end{cases}$$

The value of traction force F, according to (4), will be:

$$F_n = 1000 \cdot Pa + \left(A_0 Pg + c\frac{\rho}{2} sV_{ветра}^2\right) + \\ + \left(A_1 Pg + c\rho sV_{ветра}\right)V_n + c\frac{\rho}{2} sV_n^2.$$

Furthermore, we assume the following values of the coefficients to be equal $\widehat{A_0 g}$ = 21,07, $\widehat{A_1 g}$ = 0,0833, which corresponds to the experimental data for rail transport, $P = 9$, g = 9,8. A table of values of $v(t)$, $a(t)$, $F(t)$ is built with a sampling step Δt = 0.05 sec.

VI. The Simulation of the Process of Identifying the Parameters of the Mathematical Model of the Wagon Movement. The Solution of the Inverse Problem.

Let us conduct the simulation of the process of identifying the parameters of the mathematical model of the wagon movement for two situations. The first situation corresponds to the so-called "run-in" the wagon when an empty new wagon is released on the line, and the motion parameters are measured using high-precision measuring technology. For this case, we can consider the weight of the car to be constant and

equal to the weight of the empty wagon (9 tons), and the error of measurement of speed and distance to be zero.

Let us estimate the value of the parameters $\widehat{A_0}$ and $\widehat{A_1}$ (considering the weight of the wagon to be known quantity) using Eq. 8. The simulation results are shown in Figure 1. Both graphs show the dependence of the identifiable parameters on the wind speed. The dots indicate the values of the parameters obtained from the calculations, whereas the line provides the true values of the parameters. As can be seen from Figure 1, model results converge to the true values of motion parameters in the absence of speed measurement errors.

Let us conduct simulation of the process of identifying the parameters of the mathematical model of the wagon movement for normal operation of the wagon. Define vector $\Theta = (\frac{1}{P}, \widehat{A_0}, \widehat{A_1})^T$ to be the required vector (the speed of the wagon is unknown to us in this situation), suppose the value of the measurement error to be equal to 1% of the measured speed. Figure 2, 3 show the simulation results for P = 11, 13. For each parameter results of 4 experiments are presented.

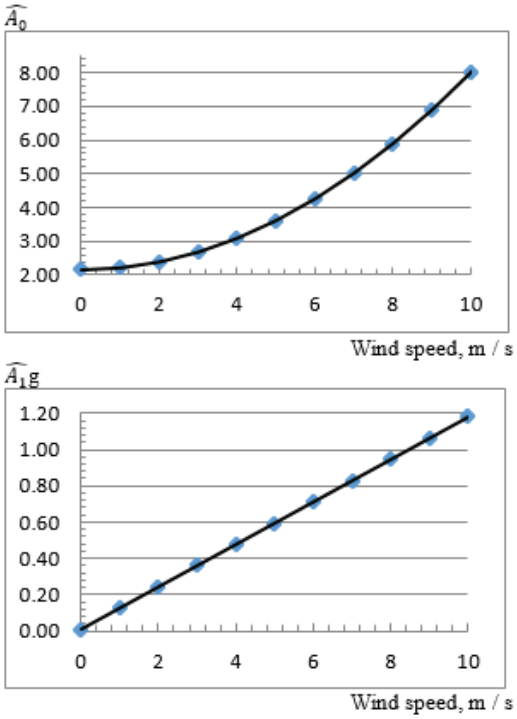


Fig. 1. *The results of identification of the parameters* $\widehat{A_0}$ and $\widehat{A_1}$ *depending on the wind speed.*

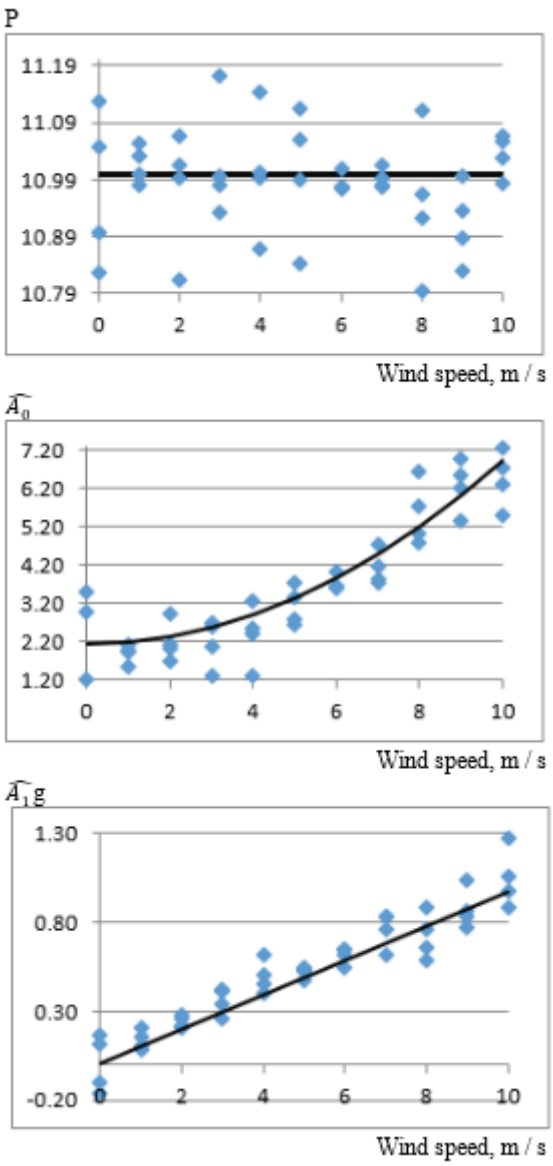


Fig. 2. *The results of identification of the parameters P,* $\widehat{A_0}$ and $\widehat{A_1}$ *depending on the wind speed when the measurement error is 1% of the measured speed (P =11).*

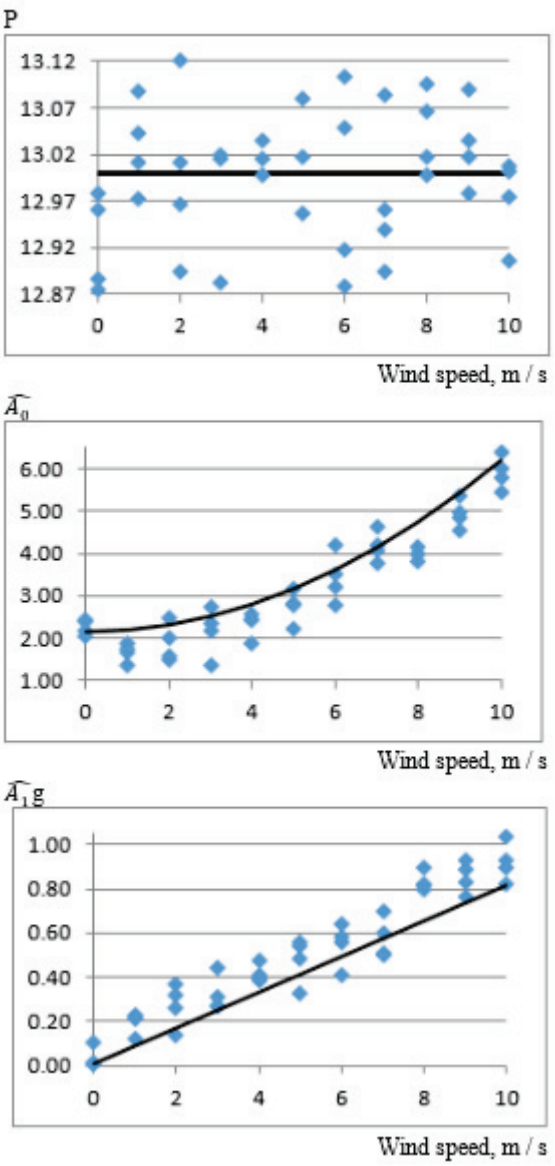


Fig. 3. *The results of identification of the parameters* P, $\widehat{A_0}$ и $\widehat{A_1}$ *depending on the wind speed when the measurement error is 1% of the measured speed (P=13).*

VII. Conclusion

To begin with the mathematical model of the wagon movement and the algorithm of identification of the model parameters are developed.

Furthermore, performance of the identification algorithm is confirmed for the selected schedule of set speed.

In addition, linearized model of the wagon movement is obtained.

Finally, the simulation confirming the efficiency of the algorithms is conducted.

Acknowledgment

Research was supported by The Ministry of Education and Science of Russia, unique identifier - RFMEFI58214X0003.

References

[1] D.O. Volkov, S.N. Garichev, R.A. Gorbachev, N.N. Moroz, “Mathematical Simulation of Transport Network Load with a View to Assessing Viability of Building New Types of Network Systems”, IEEE Proc. Of International Conference on Engineering and Telecommunication (EnT) 2015, pp. 20-23.

[2] Peter E. Timan, “Why monorail systems provide a great solution for metropolitan areas”, Urban rail transit, March 2015, vol. 1, pp. 13-25.

[3] Z. Yi-ying, C. Jian, T. Wei, “Study of feasibility of suspended monorail transit system in mountain cities”, Railway transport and Economy, July 2015.

[4] N.A. Grechiskkina, F.N. Grigoriev, N.A. Kuznetsov, “Adaptive traffic control of the car of the monorail road”, Information Processes, vol. 15, pp. 440-446, 2015.

[5] U.V. Linnik, “The least squares method and the basics of mathematics and statistical processing of observations theory”, Moscow, Fizmatgiz, 1962, p. 349.

[6] S.R. Rao. “Linear statistical methods and their applications”, Moscow, Nayka, 1968, p. 547.

[7] A. Albert, “Regression, pseudoinverse and recurrence assessment”, Moscow, Nayka, 1977, p. 226.

Using of Multi-Agent System to Model a Process of Maintenance Service and Repair of Equipment of a Service Company

Vladislav V. Panteleev, Valery A. Kamaev, Alexey V. Kizim, Anna V. Matokhina
Department of Computer-Aided Design Systems
Volgograd State Technical University
Volgograd, Russia
panteleev.vlad@gmail.com; vkamaev40@mail.ru; kizim@mail.ru; matokhina.a.v@gmail.com

***Abstract*—This report examines the issues of improving the efficiency of use of available funds for the maintenance and repair of equipment, reducing the burden on the staff. We have investigated and analyzed the process of repair and maintenance of a service company. We have defined basic models and its set of functions. This work is based on agent-based approach to simulate MRO process. Logical and conceptual models of multi-agent systems of MRO process have been developed.**

***Index Terms*— training, testing, maintenance and repair, maintenance and repair, the organization of maintenance and repair, general industrial equipment, software and information support, multi-agent system (MAS).**

I. Introduction

In our country, the necessity of improving the quality of maintenance and repair (MRO) is very high, because of depreciation of equipment. In this regard, MRO staff services of large and medium-sized industrial companies have to analyze a huge amount of data every day and it makes it harder to plan, register and prepare the repair work, especially in case if units are separated in the area. Proper management can improve the quality of MRO and reduce time and financial losses [1].

MRO Management covers a wide range of problems, such as structural organization, methods of research, control and improvement of the reliability of the equipment, scheduling.

One of the most pressing problems of industrial enterprises at present is to ensure the continuous operation of the equipment during the production cycle. This is due to high levels of wear and tear, which reaches under 70%. Employees repair services large and medium industrial enterprises on a daily basis has to handle huge amount of data, which greatly complicates the planning, preparation and registration of maintenance activities, especially in the territorial distribution units. Proper management of maintenance and repair (MRO) MRO can simultaneously improve the quality and reduce the resulting time and financial costs. In recent years, since the reorganization of the USSR, maintenance and repair of the equipment has changed quite significantly, and the old methodological developments is not always possible to apply, there is a conflict of goals and subjects of interest [2].

Without use of modern methods and ways of automation of implementation of the it's Impossible to reduce operating costs and accidents at the same time [3]. For a complex solution of these problems a flexible information system is required which can adapt and which considers the specifics of MRO management at a large industrial company [4].

Application of modern technologies of simulation is one of the effective ways of solving the problem of forecasting and the study of technical condition of rotating equipment. Since the dynamic equipment - an object with a complex structure and its status is made up of the states of its elements, for modeling the dynamics of the technical condition appropriate to use "Agent based modeling" approach.

This approach allows you to simulate the behavior of complex-structured objects and for modeling systems containing a large number of sub-systems that have individual behavior.

It is reasonable to use technology of multi-agent systems to model and informational support of distribution process.

At the same time in order to ease the information burden of the MRO staff such system needs a support of decision-making processes, it means that it must have some features of an expert system.

Relevance of this problem makes it necessary to increase the efficiency and quality of information processing, which provides intellectual support for MRO processes of equipment of industrial facilities [5].

The agent-based modeling, the system is modeled as a set of autonomous units capable of autonomous decision-making, called agents. Each agent evaluates the individual's own situation and the situation on the basis of the assessment shall take a decision, which is based on a set of defined rules. Agents may behave differently depending on what kind of system they are, for instance, they can sell, consume or produce. At the heart of the main ideas of agent-based simulations are repeated interaction between agents, Competitive and nature. The implementation of this idea lies in the ability of modern

computers that allows us to study the dynamics of the system out of reach of mainstream purely mathematical methods.

At the simplest level, agent model consists of agents of the system and the relationship / interaction between them. Even a simple agent model can detect the complex behaviors and provide important information about the dynamics of real-world systems that it simulates. In addition, agents are able to develop, showing unexpected behavior. The complex agent-based models sometimes include neural networks, evolutionary algorithms or other learning technology to simulate real processes of learning and adaptation.

We can say that agent-based modeling is a way of thinking rather than technology. The mentality of agent-based modeling is the description of the common system in terms of its component parts. Synonymous with agent-based simulations could be called the "microscopic modeling".

Modern trends in considered area have been reviewed such as:

- The development of a simulation model ability to predict changes in the condition of equipment for more efficient use of money for repairs and maintenance.
- Development simulation of the MRO process models using agents based on intelligence.
- Create of systems to support MRO process decision-making.
- Using this approach to study the process of inspection and repair of various enterprises.
- Simulation of MRO process in different areas such as metal-working and machine building, oil refining, astronautics, aviation, and etc.

II. Methodology

A process of maintenance service and repair of equipment is the main object of ongoing research.

To carry out of our research, we chose a service company. The chosen service company carries out maintenance and repair of various equipment for different customers.

The process of repair and maintenance and repair service of the company is the subject of our research. The process in question MRO repair and maintenance service of the enterprise used a team, which is engaged in the repair of a particular piece of equipment. For example, the dynamic-mechanical equipment: pumps. In their work they use the methodology of repair "on the actual state of" equipment.

Technical staff of the enterprise service performs tasks necessary for the maintenance and repair of equipment, such as:

- timely inspection equipment;
- ordering spare parts in stock;
- execution of works required to repair;
- equipment stops for repair \ start after repair;

Process MRO repair team can be divided into the following subtasks:

- timely inspection equipment;
- deciding on necessity of repair;
- equipment stops for repair;
- order for the repair parts in stock;
- transmission of applications for the purchase of the missing parts in a warehouse to a supplier;
- ordered the transfer of parts supplier;
- acquisition of parts in stock and transfer master;
- obtaining the necessary parts and repair equipment;
- start equipment after repair.

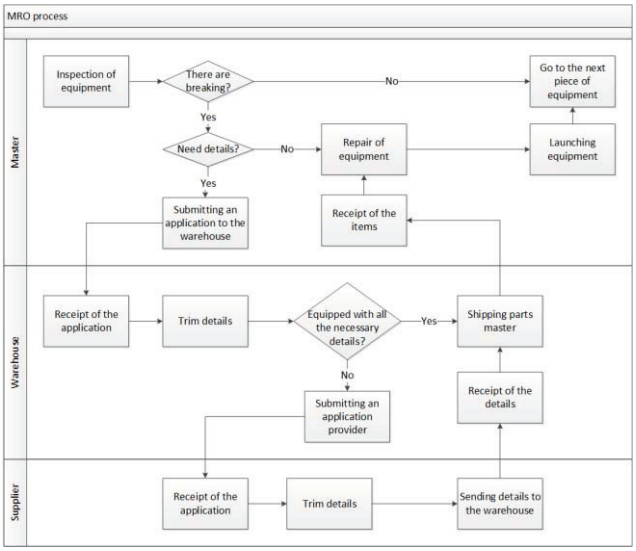


Fig. 1. MRO process of repair and service facility

The main essence of the process MRO was highlighted:

- Equipment is a device that is designed to perform the various process steps (manufacturing, procurement, finishing, thermal, and paints al.), which are necessary to obtain the required accuracy and product quality.
- Master is a specialist carrying out engineering activity. Engineers involved, as a rule, in all life cycle processes of technical devices, which are the subject of engineering, including applied research, planning, design, construction, development of production technology (equipment), preparation of technical documentation, manufacturing, commissioning, testing, operation, maintenance , repair and disposal of devices and quality management;
- Warehouse is a building, construction and a variety of devices designed for the reception, distribution and storage of goods received for them, preparing them for consumption and supply to consumers.Warehouses are one of the most important elements of logistics systems. The objective necessity in specially equipped places for keeping inventory exists at all stages of the flow of materials, ranging from the primary source of raw materials to the end consumer. This explains the large number of different types of warehouses.
- Supplier is a organization or a person who enters into an agreement with the acquirer on the delivery of the product or service.

III. RESULTS

After analyzing the existing process MRO businesses repair –service company, we managed to get the main objects involved in it and their functions.

As a result, we were able to develop a model of multi-agent based system. As it has been added to another control entity "Chief Engineer", which manages the "engineer" of subordinate objects. It monitors their employment and the flow of applications for repair.

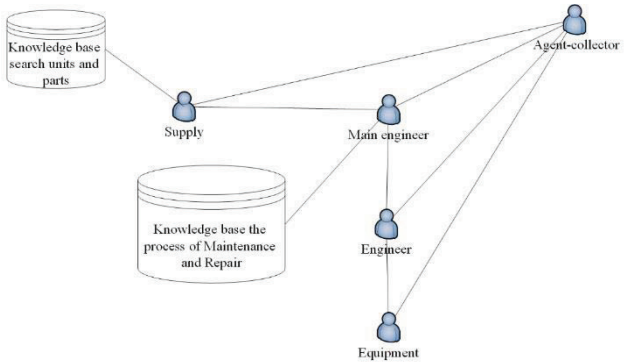


Fig. 2. Multi-agent model

Agent «Equipment» has next conditions: work; plain; break.

It also includes a list of details and connections the equipment consists of. Each element has a number of key parameters and according to these parameters it's possible to determine condition of the equipment.

Agent "Engineer" has next states:

- diagnostic of equipment;
- launching equipment; repair of equipment;
- stop equipment;
- sending status data of equipment to agent "Chief Engineer"; receiving an indication from the agent "Chief Engineer"; receipt of materials for repair of the agent "Chief Engineer".

Agent "Chief Engineer" has next states:

- receiving a report on the state of the equipment from the agent "engineer";
- analysis of the report;
- make a decision on the need of repair;
- order parts needed for repair;
- obtain the necessary parts to repair;
- transmission of recommendations and pieces of repair to agent, "engineer";
- load sharing between agents "Engineer".

In order to make a decision on the need of repair the agent "Chief Engineer" uses a base of knowledge. Nowadays there is no commonly accepted methodology of MRO. That's why in order to model different methodologies of MRO process it's necessary to use a base of knowledge which must be developed for chosen methodology. It makes the process of modelling a bit more difficult.

Agent "Supply" has next states: receiving a list of the details from the agent "Chief Engineer", order the details from provider, receiving details from the provider, sending the details to the agent "Chief Engineer", search and order of such details.

If there is no needed details the agent "Supply" can make a search using a base of knowledge to find details which are identical by design parameters and functions.

For search such details a special base of knowledge is developed. It is bases o and/or graph. It consists of the next elements.

IV. CONCLUSION

In this paper, we have reviewed and analyzed the modern scientific developments in the field of modeling the process of maintenance and repair of equipment. Reviewed and analyzed the process of repair, but MRO - service enterprise. The basic model objects and sets of functions that they perform. Developed conceptual and logical models MAC MRO process.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar.

ACKNOWLEDGMENT

Next, we plan to create the instrumental software in order to provide the realization of multi-agents system prototype, which models MRO process approbations of developed model. This work was supported by financial Grant RFFI № 16-47-340229p_a, Grant RFFI № 16-07-00635_a. This article was financially supported by the Russian Ministry of Education in scope of the base part (project 2586 of task N 2014/16), in scope of the base part (project 2.1917.2014K_2014).

REFERENCES

[1] A.V. Kizim, Establishing the maintenance and repair body of knowledge: Comprehensive approach to ensuring equipment maintenance and repair organization efficiency // Original Research Article Procedia Technology, Volume 9, 2013, pp. 812818.

[2] V. Panteleev A. Kizim, V. Kamaev, O. Shabalina Developing a Model of Multi-Agent System of a Process of a Tech Inspection and Equipment Repair // Knowledge-Based Software Engineering: 11th Joint Conference, JCKBSE 2014, Volgograd, Russia, September 17-20, 2014. Proceedings. – Springer, 2014. – T. 466. – C. 457.

[3] M.V. Denisov, V.A. Kamaev, A. V. Kizim Organization of the repair and maintenance in road sector with ontologies and multi-agent systems/ // Original Research Article Procedia Technology, Volume 9, 2013, pp. 819-825.

[4] V. V. Panteleev, A. V. Kizim, A. V. Matohina, V. A. Kamaev. Intellectual Information Support for Operation of Technical Systems Based on Ontological Engineering. Procedia Computer Science, 64, 808-815, 2015.

[5] V. V. Panteleev, V. A. Kamaev, A. V. Kizim, Developing a Model of Equipment Maintenance and Repair Process at Service Repair Company Using Agent-Based Approach. Procedia Technology, 16, 1072-1079, 2014.

The Subject-Oriented Notation for End-User Data Modelling

Sergey Kucherov, Yuri Rogozov and Alexander Sviridov
Institute of Computer Technology and Information Security
Southern Federal University
Taganrog, Russian Federation
{skucherov, yrogozov, asviridov}@sfedu.ru

***Abstract*—Existing data modeling notations and methods of database development are dedicated to technicians. At the same time, there is a tendency to involve the end-user in the development process. This requires the creation of new modeling notations and development methods for databases area. These tools should be understandable both to end-users and to developers. Because of unified understanding, it is possible to consider the action of the subject (end-user) as a basis of such modelling and designing tools. Subject's action describes the process of creating domain concepts with which end-users are working and which developers should implement in an information system. The paper proposes a subject-oriented data model and its modeling notation, which are devoted to the end user. Minimizing of graphic primitives quantity and universalization of modeling tools are results of using both the unified structure of action mechanism and its variable content.**

***Index Terms*—Database, modelling, notation, subject-oriented approach**

I. Introduction

There is a trend towards individualization of software and information systems, in particular for today. This is due to advent of financial possibilities by small firms to make their own applications, to appearance of agile software development (SCRUM, xP) [1], to creation of numerous frameworks and kits. Along with this the level of end-user involvement in the process of creating systems is increased. It is known that an end-user is the main carrier of domain knowledge [2]. Direct user involvement in the development process allows creating software and databases that are more appropriate to subject area requirements [3,4].

This in turn requires the solving of significant problem of understanding between a customer and a developer [5]. Teaching of an end-user to the level of modern development tools will lead to absence of necessity in technical specialists. It is also labor-intensive and costly process. In this context, development of new notations and concepts of representation, which are based on the perception of domain area and system by an end-user, is promising. The most interest from this point of view is a data modeling process.

Methodologies of software development suggest that data model creation and database design is a prerogative of database designer or data administrator [6]. For these specialists a lot of modelling and designing notations are created. However, all of these notations are based on the concept of object (not object-oriented), in which a database is represented as a set of interrelated data objects. Represented data model can be deprived of a significant proportion of semantics depending on an abstraction level (conceptual, logical or physical).

Existing notations can be divided into: Structural (IDEF1x, E-R, UML) and Streamed (DFD).

Streamed notations are not used directly in a database structure design. They are more focused on the problem of mapping data streams. Such notations are able to demonstrate end-user information about a data conversion on their way from a source to a receiver. Each converter is a separate function with an encapsulated content.

Structural notations, on the contrary, form the basis of modern data modeling tools. Such notations are focused on information entities, interconnected set of which describes a domain data structure. Mechanisms such as normalization, already at a logical level hamper the perception of data model by an end-user. In addition, structural notations require a thorough documentation of a model at all levels of abstraction.

In the research process and the practical work on creation of information systems next important features of a subject area and data modeling have been identified by us. It call into question the applicability of existing notations:

- An end-user does not perceive data separately from processes of their creation, transformation and processing in many cases. The basis of perception by an end-user is subject-oriented concept [7,8]. This concept, in contrast to the object, are focused on user's actions which are leading to generation, transfer and transformation of information, and an information itself as a product of user activity [9].
- Normalized database schemas complicate the perception of information system data model, because along with a description of domain data objects' semantics they require to describe a separation of these objects into a plurality of entities.
- The combination of structural and streamed notations does not provide a complete picture about a domain data, as it does not represent a causal link between a purpose of processing (of creation) data and its result. Processes themselves, in which data can be created or converted are related to business logic of information

systems for today and this is not part of a data model (fig. 1).

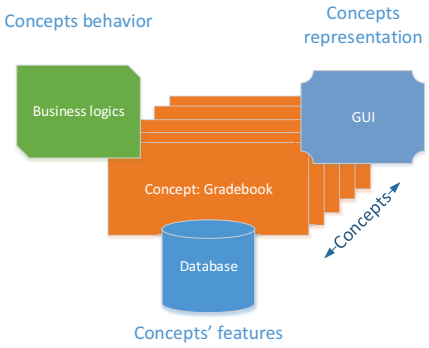


Fig. 1. Representation of concepts in classical systems

Figure 1 shows that the one domain concept, with which end-user works in real life, is represented in an information system as a set of objects of different types: Properties that reflect a concept sense are placed in a database; A visual representation of a concept sense is contained in a graphical user interface; A behavioral aspect of concept is reflected in an application business logic; A concept reflection into an information system are performed non by subject area specialist but by technician, often without knowledge of domain specifics; Different users create each of aspects into that a concept sense is shared.

All this leads to a mismatch both between a domain concept sense and between its realization in an information system, and within a concept sense.

Thus, there is a segmental representation of processes and data, which is also have no means of explicit and formalized binding into a single integrated domain model. This is, firstly, hampering communication between a developer and a customer, and secondly, is preventing to the construction of development methodologies and tools aimed at the end user. In addition, existing methods are not able to fully ensure the dynamism of a database on the same level with a domain area [10].

The key problem, which follows from the foregoing, is the lack of data modeling and design tools, focused on an end-user. Because of this the set of designed flexible databases [11-14] are not commercially successful. They are designed for an end-user but for represent a data model they use technologies and models aimed on a developer. To solve this problem, we have proposed the subject-oriented repository [15,16], which is combining technical and user domain description. The next step is the development of modeling tools, aimed to an end-user. This is due to the fact that a notation that combines the possibility of describing end-user actions and results of these actions is not exist for today.

The aim of this paper is to provide the user-centric notation for the subject-oriented data modeling. The first chapter represents the subject-oriented data model. In the second chapter the graphical notation and description of its implementation are proposed.

II. Subject-Oriented Data Model

From an end-user's (subject's) position data is an integral part of processes and it can not be separated from them to preserve semantics. A potential user until a system introduction performs some work (consisting of individual activities) over a set of domain concepts [17]. Each domain concept has a name, which allows to determine its type and purpose, defining the implementation of this concept in a specific environment.

A concept is a more general term to describe a domain area, but in terms of databases, it can be identified with a data object in the classical approach to development. A data object is characterized by a set of attributes that determine values of its key properties. The concept is characterized by a description of a process, implementation of which leads to its reception. Description of processes is dependent upon a context in which a concept is used. Consider differences between a concept and data object on the example of a book:

- A book as a data object is represented by a set of attributes: title, author, number of pages, type of binding, etc.;
- A book as a concept will be the description of a process of creating a book from cover and pages of text (including a title and an author). In addition, a page of text is a concept, which process of creating contains just a text itself and rules for its printing.

The peculiarity of the subject-oriented approach is that maximum variability is ensured for each concept. At the same time, there are three state of concepts. They are characterized by the content of their sense: As long as a process description does not contain any specific values, the concept has an implicit sense. In the example with a book, we used exactly this type of concept - we do not know in what this book is bound, this book is printed on a paper or electronic, etc. However, we clearly understand that we are talking about a book. A concept sense is filled with concrete characteristics during a process specification (cover type, page size). Because of this process execution, a concept with explicit sense is obtained. It corresponds to a data object, but in contrast, contains not only a description of the expected result (in this example, a book), and the process of its creation. If the specified process will be carried out, its characteristics will have specific values and the concept itself becomes specific. For example, the book about SWEBOK. This corresponds to an instance of a data object (see. Figure 2). Thus, a concept approach is nothing other than the action specification to create a domain object.

The convenient tool for modeling concepts is the mechanism of action [17] - a fixed set of characteristics with variable content. Changing content determines a purpose of an action, while the fixed structure of actions representation allows to use a single abstraction throughout the process of system creating - from a research of domain area, prior to the establishment and operation of a database and information system as a whole.

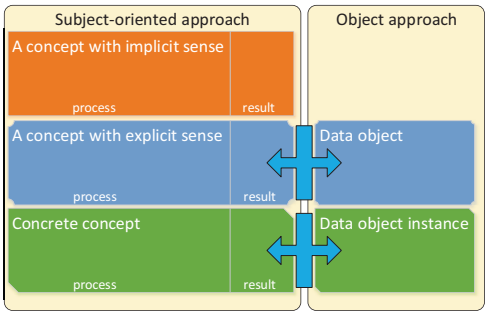


Fig. 2. Correspondence between elements of subject-oriented and object approach

For the description of domain concepts, the mechanism of action uses the simple basic set of features - elements, functions, tools and results. Regardless of a level of detail this base set of features is fixed and is a form, filling if which by content (specification of elements, functions, etc.) is a process of design concepts.

Subject-oriented data model, being adequate to domain as a basic storage unit uses a concept and its forms described above. Presentation of each type of a concept corresponds to a certain type of action.

An action type is the reflection of a concept with implicit sense in the form of named unit of user activity. An action type holds no characteristics (Figure 3).

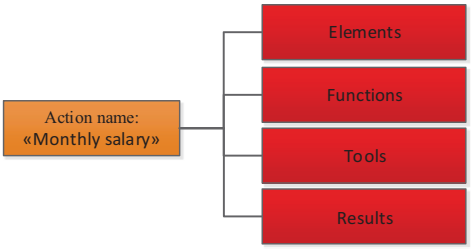


Fig. 3. The model of concept with implicit sense

Elements, tools and functions will largely depend on the form in which the concept will go with specifying its sense. *An action* is the reflection of a concept with explicit sense in the form of the named unit of user activity. An action is expressed by used elements, functions, a tool that regulates the rules for the application of functions on elements and results that can be obtained by performing an action appropriate for the purpose of its implementation (see Figure 4). It should be noted that an action specifies result only as an expected performing targets, while the result value appears after an action implementation.

An action is the basic unit of data storage and presentation of configurable information system. An example of a user's actions can be "payment of monthly wage to worker." As elements it is using information on number of days worked (which in turn is a result of "counting the number of days worked) and information about the cost of one working day. As functions it is using the mathematical operator "multiplication", and as a tool the multiplication rule included in one of the system modules are used.

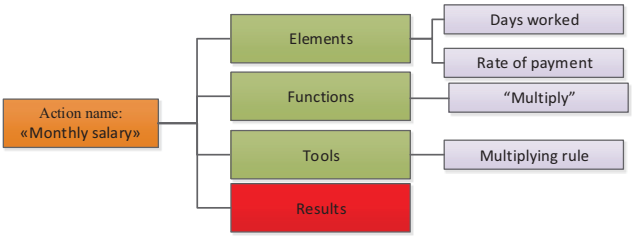


Fig. 4. The model of concept with explicit sense

Perfect action is the result of an action implementation in a particular moment of time, containing specific values of all the characteristics, including the value of result (see Figure 5.).

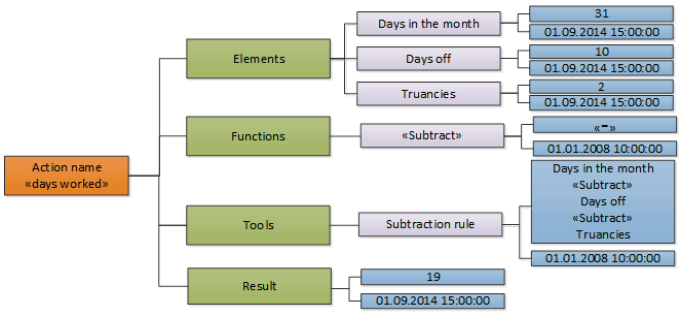


Fig. 5. The model of concrete concept

It corresponds to a specific concept. In the context of the example described above, it may be the value of a particular employee wages in a given month. This will not only preserved the value itself, but also its production process. This greatly expands possibilities of storage in terms of supporting the integrity and historicity changes

Action is declared via the common pattern for the description, which includes an action name, four basic characteristics and a list of their constituent components. The result of action description is degenerate, since an action is described, but not implemented. Perfect action, in contrast to the action has values of characteristics. A time stamp obtaining or setting accompanies each value. In addition, for an action, we store a result of its implementation, and timestamp for values obtaining deemed actions implementing time.

As can be seen from figures, each kind of action is a named tree describing the concept. Inside the node are possible hierarchy, describing the nesting activities in each other and allow to describe the complex subject of activity.

Link in the subject-oriented data model is also action. By analogy with a real world, concepts can be linked within a particular activity. Thus various binding methods are possible:

- Result <-> element. This type of connection is the most common and demonstrates a sequential process of obtaining a data: a result of calculation for number of days worked becomes an element of calculation of wages (see Figure 5);
- Result <-> function. This type of connection demonstrates the possibility of using one of concepts as the description of a function of the other. For

example, the term "increase" is used to describe the concept of "wages" as a function;

- Result <-> tool. This relationship type describes a situation where one concept is a means or a rule to perform a function on elements in the description of other concepts.

III. Notation for Subject-Oriented Data Model

In terms of structure, the subject-oriented data model differs from currently known models and has following features: The dynamic set of characteristics (in the presence of a base constant set) and their values; The lack of a clear separation of the elements by types. Any value of any characteristic may be element, function or tool; Lack of association link as to classical models. All concepts are linked only through action.

It requires alternative tools and methods of implementation. The subject-oriented data model uses only one basic element of submission - action. With this element modeled as concepts of various kinds, and their relationship. Figure 6 shows different ways of action representation.

Figure 6a shows a minimized view of an action. It is consist of «action name» and region of results (for the perfect action, reflecting the specific concepts) «action value».

Figure 6b shows a detailed representation of an action with specified set of characteristics used in its description. As each characteristic is also a concept for its representation minimized type of an action is used (fig. 6a). Figures 6a and 6b are also a representation of a concept with implicit sense (type of action). It does not include any specific elements, functions or tools. Figure 6c shows an action that characterizes a concept with explicit sense from the example in the previous section. For its representation in accordance with the data model is required to transfer characteristics. Modelling concrete concepts is necessary in determining the system constants. Most of the concrete concepts is a specific result of the user's work with information system or database. This goes beyond the issues discussed in the article.

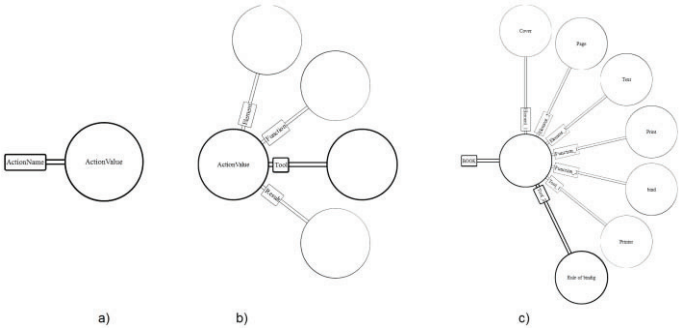


Fig. 6. Graphical primitives of subject-oriented notation

Thus, we can say that the process of describing a data model in accordance with the proposed notation - a consistent description of a hierarchical structure of domain concepts.

In terms of rules and constraints in the proposed modelling notation can be distinguished follows:

- Data modeling begins with a main action, characterizing the scope of an information system;
- The depth of decomposition is dependent on a functional purpose of an information system and potentially not limited;
- In describing characteristics of action bundle «action name»: «action value» must be unique.
- If necessary, a reference to a concept of other hierarchy in field «action value» we specifies the bundle «action name»: «action value» of necessary concept. The communication is a special action, called IS_A_RESULT.
- In one model duplicates of the same action is not permitted with the exception of perfect action.

The proposed set of graphics primitives and rules is minimal and sufficient to represent a domain data model in terms of subject-oriented approach. Figure 7 shows a portion of user data structure, showing the process and results of calculating the monthly salary of each employee.

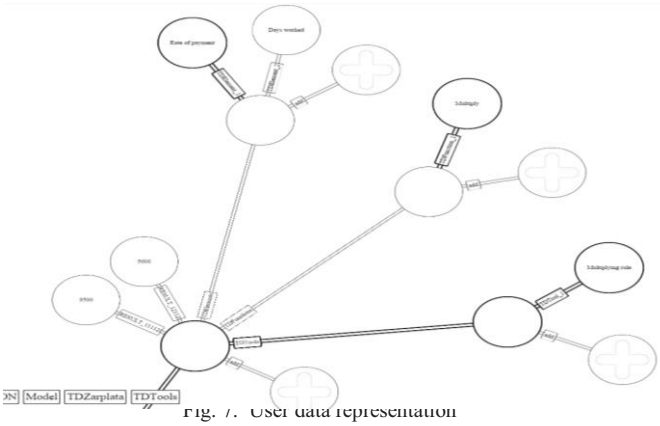


Fig. 7. User data representation

Implementation of storage and modeling tools is carried out on the basis of NoSQL database GlobalsDB [18]. A key feature of this solution is fully supported schema-free approach [19,20]. The storage consists of permanently stored global variables - globals.

The approach to the implementation of storage model is based on the creation of a special data structure. This structure is a multi-dimensional sparse array. Structural elements of the array are elements of the metric chosen to describe a domain. In our case, this metric - the mechanism of action.

This array contains both a specification of subject's works in the form of concepts senses and results of action performing in the form of specific concepts. Configuration layer is described by using globals with the prefix name TD *. Results of user's work with a system are stored in globals with the prefix name RESULT_ *, where * is used as an absolute time, set at the time of obtaining a result.

IV. Conclusions

The paper suggests a notation for the subject-oriented data model. This notation is based on the using of a single representation conception for stored elements and links between them. It is used as a tool to describe a domain area for

a storage of configurable information systems, solving the problem of variability and consistency of application systems and user tasks.

This tool allows to shift modelling to a new level - semantics of a data, nature of their occurrence are clearly documented in a readable form for an end-user. This in turn has a positive effect both on a modelling process and on a process of operation and improving system by end-user.

Subject-oriented modelling notation allows to represent a structure of user data and results of user working with a system in an adequate form with respect to a domain model. Using of subject-oriented approach erases a clear separation of data, logic and interface. Concept described by an action can be repeatedly reproduced with necessary changes in a composition and characteristics of the structure.

The use of NoSQL technologies, in turn, will allow to realize high-performance solutions, devoid of shortcomings of existing flexible databases implementations.

ACKNOWLEDGMENT

The reported study was partially supported by RFBR, research project No.15-07-04033.

REFERENCES

[1] Rubin, Kenneth S.,2013 Essential Scrum: a practical guide to the most popular agile process / Addison-Wesley. - 498 p. ISBN 978-0-13-704329-3

[2] Kucherov, S., 2013. User-configurable information systems as a means to overcome the semantic gap / Journal of Informatization and Communication 5'2013, P. 135-137

[3] Cohn, M., 2009. Succeeding with Agile: Software Development Using Scrum / Addison-Wesley. - 504 p. ISBN 978-0-32-157936-2

[4] Fischer, G,. 2009. End-User Development and Meta-design: Foundations for Cultures of Participation / IS-EUD '09 Proceedings of the 2nd International Symposium on End-User Development, Springer-Verlag Berlin, Heidelberg, p. 3-14.

[5] Lavrishcheva, E., 2013. Software Engineering Computer Systems. Paradigms, techniques and CASE-tools programming / Moscow:Science. Dumka, 2013. - 283 p.

[6] Date, C. J., 2004. An Introduction to Database Systems (8th Edition) / Addison Wesley. – 983 p. – ISBN 978-0321197849.

[7] Rogozov, Y. I., 2013. Approach to the definition of a meta-system as system. Proceedings of ISA RAS. - 2013. - Vol. 63., №4 / 2013. - P. 92-110.

[8] Rogozov Y., 2014a. The general approach to the organization of certain system of concepts based on the principle of generating knowledge. XII All-Russian Conference on Control VSPU-214. Moscow: Proceedings. ISBN 978-5-9145-151-5., pp 7822-7833.

[9] Chernikov B.V., 2014. Specificities of Lexicological Synthesis of Text Documents // Computer Science Procedings. 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014 / Vol. 31. ELSEVIER, p. 431–439

[10] Kucherov S., Rogozov Y., Borisova, E., 2014. Structure–independent databases modelling / Proceedings of IEEE 8th International Conference on Application of Information and Communication Technologies - AICT2014, ISBN: 978-1-4799-4120-9 2, p. 192-196

[11] Kucherov, S., Samoylov, A., Grishchenko, A., 2014. Flexible database for configurable information systems // Proceedings of IEEE 8th International Conference on Application of Information and Communication Technologies - AICT2014. IEEE Catalog Number CFP1456H-PRT, ISBN: 978-1-4799-4120-9 2, p. 374-377

[12] Kucherov S., Sviridov, A., Belousova S., 2014. The formal model of structure-independent databases / Proceedings of 3rd International Conference on Data Management Technologies and Applications DATA2014, Vienna, Austria, Scitepress - Science and Technology Publications. ISBN: 978-989-758-035-2. p. 146-152

[13] Rossi, C., Guevara, A., Enciso. M., 2010. A tool for user-guided database application development. Automatic Design of XML Models using CBD / Proceedings of the Fifth International Conference on Software and Data Technologies, ICSOFT 2010, Volume 2, p.195-200

[14] Palviainen, M., Kuusijärvi, J., Tuomisto, T., Ovaska, E., 2014. Towards a data model of end-user programming of applications / DATA 2014 - Proceedings of 3rd International Conference on Data Management Technologies and Applications, pp. 272-281.

[15] Kucherov, S., Rogozov, Y., Sviridov, A., 2015. NoSQL storage for configurable information system: The Conceptual Model / DATA2015 - Proceedings of 4th International Conference on Data Management Technologies and Applications, ISBN 978 - 989758103 – 8, p. 279-287

[16] Kucherov, S., Rogozov, Y., Sviridov, A., and Rasol, M. 2015. Approach to data warehousing in configurable information systems based on action abstraction / Communications in Computer and Information Science, Volume 535, 2015, Pages 139-148

[17] Rogozov Y., 2014b. Methodology of creation of subject-oriented systems. / Journal of Informatization and Communication 2'2014, P. 6-10

[18] Tweed, R.; James, G., 2010. A Universal NoSQL Engine, Using a Tried and Tested Technology. http://www.mgateway.com/docs/universalNoSQL.pdf

[19] McCreary, D., Kelly, A., 2013. Making Sense of NoSQL: A guide for managers and the rest of us. Manning Publications, 312 p. ISBN 978-1-61729-107-4.

[20] Sadalage, Pramod J., Fowler, M., 2012. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, ISBN-13: 978-0321826626.

Application of Neural Network for Digital Recursive Filter Design

Naila Allakhverdiyeva
Process Automation Engineering Department
Baku Higher Oil School
Baku, Azerbaijan
nailaallakhverdiyeva@yahoo.com

Abstract— **Digital Signal Processing is an advanced technology that will determine the direction of science and technology in the next centuries. One of the main direction of digital signals processing is digital filters, which in the most cases have advantages over analog filters. Currently there are various methods of filter analysis and design. In this work, for synthesis of all types of recursive filters (low pass, high pass, bandwidth, band stop) is used a neural network. The main objective of filter synthesis is to find the filter coefficients. These filter coefficients define the filter transfer function. Using an iterative procedure of the neural network such as Backpropagation algorithm, on base of Visual C++ software was developed the program, which designs recursive filters with required characteristics. This is particularly important for the designing of the new correcting filters characteristics, the purpose of which is to reduce the unwanted noises in the measurement signal.**

Index Terms— **Digital filter, IIR filter, neural network, Backpropagation algorithm, Impulse response, Transfer function.**

I. INTRODUCTION

Digital filters are an important part of Digital Signal Processing (DSP). In fact, the digital signal processing has gained its popularity due to the digital filter. Filter synthesis process is a process of developing the algorithm which produce an output signal in response of input sequence. Filter synthesis can be described as a process of optimization of certain parameters. In particular, for the synthesis of corrective filters such parameter is a minimum of mean square error on the filter output [1].

There are two types of digital filters: filters with Finite Impulse Response - FIR filters and filters with Infinite Impulse Response - IIR filters. IIR filters are the digital equivalent of analog filters. The output of this filter is determined as a linear combination of the previous inputs and outputs. In other words, there is feedback in the implementation of such filters. In FIR filters there is no such feedback. Theoretically, the impulse response of IIR filter never falls to zero, although in practice, taking into account the limited capacity of computer registers it is also possible.

In this work for synthesis of recursive filter structure has been used a Neural Network (NN).

The theory of neural networks uses an approach that is related of processing information by our brain. The idea of artificial neural network was established by analogy with the biological neural networks. The processes occurring in biological neural networks are related to the acquisition, processing, presenting, transmitting and using information. Artificial neural networks try to simulate the human brain function, which has the ability to learn and parallel process the necessary information. A neural network is a complex system (Figure 1), which processes the input signal *x[n]* and generating the output signal *y[n]*.

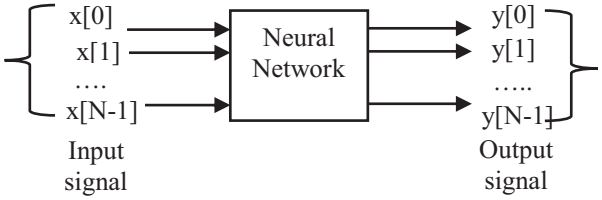


Fig. 1. Common structure of neural network

The common systems operate sequentially in accordance with a predetermined accurate algorithm. Neural networks operate on base of parallel and uncertain algorithm. This is difference between common systems and neural networks. Neural networks is based on an algorithm that results from the process of network training. Network trainning is based on the training data. The features of the artificial neural networks are:

- The network is enriched with information during the training process.
- The layers of the neural network are used to store this information.
- The data are processed in parallel.

Artificial neural networks are fundamentally different from other well-known approaches for system design.

For training of neural networks is used a lot of algorithms. In our work the structure of the neural network was obtained using a gradient descent algorithm, so-called Backpropagation method. For development of user interface has been used Visual C ++ programming environment. In this program, the required frequency response of the designed filter is entered and then using the learning algorithm of the neural network,

the program calculates necessary filter coefficients and displays them and the frequency response of the resulting filter in the window.

II. Neural Network Learning Algorithm

The common convolution equation for any filter kernel (h[n]) and an arbitrary input signal x[n] is defined in [2, 3] as:

$$y[n] = \sum_{k=0}^{M-1} h[k]x[n-k] \quad (1)$$

For IIR filters, in which there is feedback, the output is given as:

$$y[n] = \sum_{k=0}^{M-1} h[k]x[n-k] = \sum_{k=0}^{M-1} a[k]x[n-k] - \sum_{k=1}^{M} b[k]y[n-k] \quad (2)$$

The *a[k]* and *b[k]* coefficients determine filter characteristics. Impulse response of the filter can be expressed in another way:

$$h[n] = \frac{\sum_{k=0}^{M-1} a[k]x[n-k]}{\sum_{k=1}^{M} b[k]y[n-k]} \quad (3)$$

It is possible to obtain frequency response of the filter using z-transform. In this case, z-transform of the input signal is defined as:

$$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n} \quad (4)$$

Where z is a complex variable.

If we apply z-transform to Equation 3, we obtain transfer function (frequency response) of the recursive filter:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{m=0}^{M} a[m]z^{-m}}{\sum_{n=1}^{N} b[n]z^{-n}} \quad (5)$$

Equation 5 is a common equation for recursive filter. It describes the relationship between the coefficients *b[k]* for recursive and the coefficients *a[k]* for nonrecursive part of the IIR filter. *N* characterizes the degree of filter order.

As the weighting coefficients of the neural network has been used a set of recursive filter coefficients. At the beginning of the program all coefficients set to zero except coefficient *a[0]*. Zero sample coefficient is assumed to be one in order to make the characteristic of the filter identity to *δ[n]*. If you allow the program to generate the values of the coefficients for the initial iteration randomly, then design filter may be unstable. Therefore, for next iterations, the neural network weighting coefficients are changed in accordance with the gradient descent algorithm and at each iteration, obtained filter characteristic is compared with the desired characteristic.

Gradient descent algorithm, so-called Backpropagation algorithm, is as follows [4]:

Let denote the weight vector coefficients obtained after *m* iterations as *w(m)*. Then we get:

$$w(m+1)=w(m)+\Delta w(m). \quad (6)$$

Where *Δw* is the change for *w(m)*. This change can be defined as:

$$\Delta w(m)=\mu \nabla J_{w(m)} . \quad (7)$$

Where μ is a parameter characterizing the rate of gradient descent.

$$\nabla J = \left[\frac{\partial J}{\partial w1} \dots \frac{\partial J}{\partial wn}\right]^T \quad (8)$$

Gradient descent method updates the weights *w(m)* at each iteration. The method includes the following steps:

1. Use input vector x_m and desired output vector d_m for the *m*-th training iteration.
2. Calculate e_m^2, where

$$e_m^2 = \left(d_m - w_{(m)}^T x(m)\right)^2 = d_m^2 - 2d_m w^T(m)x(m) + w^T(m)x(m)x^T(m)w(m) \quad (9)$$

3. Calculate the expression:

$$\nabla J = \frac{\partial e_m^2}{\partial w(m)} = 2x(m)w^T(m)x(m) - 2d_m x(m) = -2\left(d(m) - w^T(m)x(m)\right)x(m) = -2e_m x(m) \quad (10)$$

That is, using the expression for e_m^2 you get an approximation for *∇J*:

$$\nabla J \cong -2e_m x(m) . \quad (11)$$

4. In accordance with the Equation 6 and Equation 7 update the coefficients *w(m + 1)*:

$$w(m+1)=w(m)-2\mu e_m x(m) . \quad (12)$$

This process is called Delta Rule for neural network [5].

Obtained on each iteration errors then are summed and are used to update the filter coefficients in order to adjust them to desired frequency response.

The algorithm for the program can be written as the following steps:

- Enter the desired values of impulse response in the form of numeric array.
- Reset all filter coefficients except *a[0]* to zero. Let *a[0]* equal to one.
- Calculate the FFT for identified system.
- Convert obtained values in polar form, and extract magnitude values.
- Calculate the error which is the difference between obtained and desired responses.
- Sum error values and get mean square error.
- Given the importance of the resulting error, go to the next step of iterations.
- Repeat all steps until resulting error becomes negligible.
- Type filter coefficients in the window and draw the resulting filter response.

Flowchart of IIR filter synthesis is shown in Figure 2.

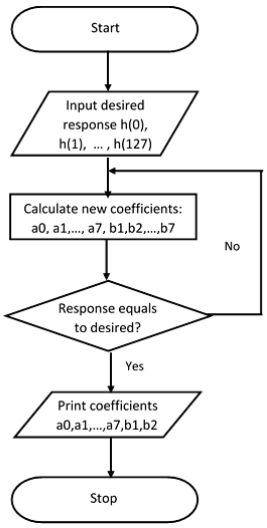


Fig. 2. Flowchart of IIR Filter Design

III. Description Of The Program

For the synthesis of IIR filter has been implemented the program in Visual C ++ programming environment. Program consists of one main window and additional supplementary forms.

Program input is a desired frequency response in the form of numeric array, which contain the desired response amplitude values in the range of [0, 0.5]. This is normalized frequency, which varies from 0 to the Nyquist frequency. The data are read from the Microsoft Access database. There are stored the data for frequency responses of standard filters (low frequency, high frequency, band-pass, band stop, etc.). Database is linked to the program by means of OLEDB tools of Visual Studio programming environment. For quickly access, a copy of the data is stored in DataSet object. The database file is saved in the same directory as the current project. Any changes in the design of the filter after editing are stored in the array, updates will take effect after pressing the <UPDATE RESPONSE> button in the window. To obtain the desired response coefficients of the filter program is performed on based on gradient descent algorithm. By clicking on the form <DESIGN> button, you can get the graphical representation of frequency response of designed filter. In this window, you can also see the obtained filter coefficients and the type of selected filter. In Figure 3 is shown the design window for low pass IIR filter with order equals to 8.

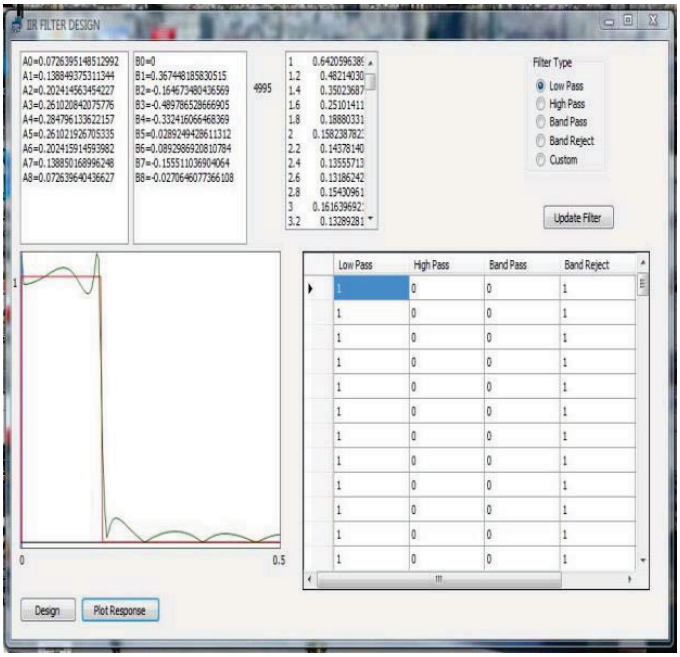


Fig. 3. Filter Design form

Low pass filter coefficients, which have been obtained using the neural network algorithm, are shown in Figure 4.

A(0)=0.073639	B(0)=0
A(1)=0.139849	B(1)=0.36844
A(2)=0.203414	B(2)=-0.17467
A(3)=0.26002	B(3)=-0.48988
A(4)=0.28479	B(4)=-0.32241
A(5)=0.26102	B(5)=0.02892
A(6)=0.20141	B(6)=0.08939
A(7)=0.139849	B(7)=-0.15561
A(8)=0.72649	B(8)=-0.02806

Fig. 4. Coefficients of Designed Filter

For the transfer function of the desired low pass filter with 8-poles obtained coefficients are then substituted into the Equation 5:

$$H(z) = \frac{0.07+0.14z^{-1}+0.2z^{-2}+0.26z^{-3}+0.28z^{-4}+\cdots}{1+0.36z^{-1}-0.16z^{-2}-0.49z^{-3}-0.33z^{-4}+\cdots} \quad (13)$$

From Figure 3 you can see that the resulting filter is stable as all the poles of the filter are inside the unit circle.

To synthesize an arbitrary filter whose parameters are set by the user it is also used neural network and gradient descent algorithm. The filter coefficients are entered into the program window. In this case, from program window you have to select <CUSTOM> type for the filter. Sample window for arbitrary designed filter is shown in Figure 5. In this window, you can see the estimated filter coefficients, as well as the graphical representation of obtained frequency response. Also, consider obtained coefficients of filter becomes clear that the resulting filter is stable.

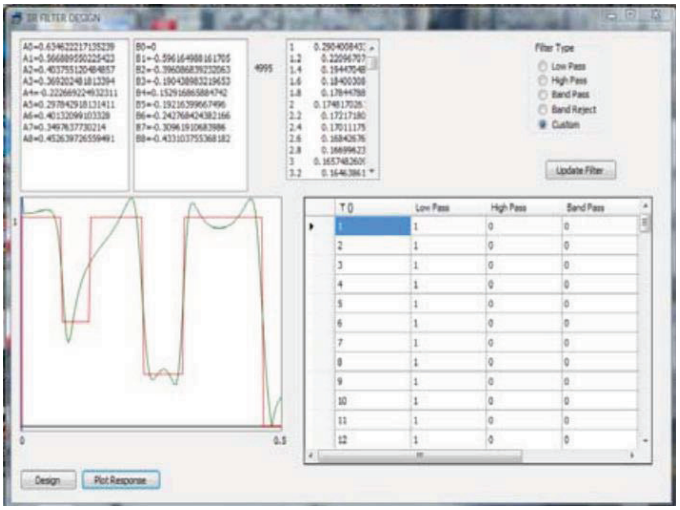


Fig. 5. User-defined filter Design

The program also has a possibility of applying of designed filter for processing some kind of input signal. For this, input signal's data are entered into the program and after pressing <APPLY FILTER> button in the next window signal is filtered on base of designed filter. Result of this window is shown on Figure 6. As seen from the Figure 6, the low-pass filter has been created as a result of network training. This filter reduces the signal/noise ratio for the input signal. Therefore, the synthesized filter has correction characteristics for random noise elimination.

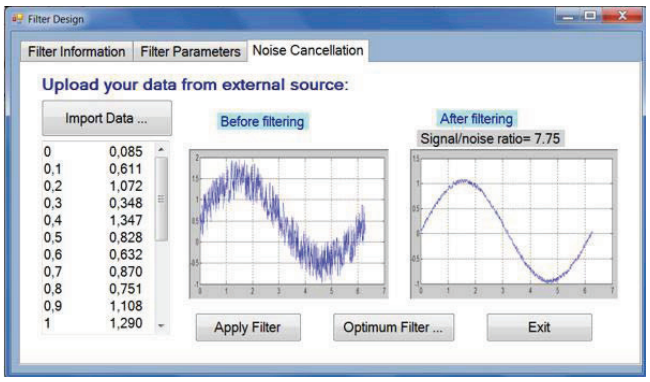


Fig. 6. Applying of designed filter

IV. Conclusion

Before Thus, using this program, you can carry out synthesis of recursive filters based on Backpropagation algorithm of neural networks. A special feature of this program is that it allows you to synthesize arbitrary filters with arbitrary frequency responses. The efficiency and stability of the obtained filter is determined by the fact that the synthesis was done based on minimizing the error at each iteration step. Synthesis of filters with specified characteristics is very important in the measurement technique, when it is necessary to filtering an input signal from the random noise without having any priori data.

V. References

[1] I.Abdullayev, N.Allakhverdiyeva. "Correcting Filtration in Measurement Systems: Monograph", Baku: Chashioglu publ., pp. 50-181, 2005, [I.M.Abdullayev, N.R.Allakhverdiyeva. Korrektiruyushaya filtrasiya v sredstvax izmereniya,].

[2] S.Smith. "Digital Signal Processing: A Practical Guide for Engineers and Scientists", Publishing by Newness, 650 p., 2002.

[3] L.Chaparro. Signals and Systems using Matlab. University of Pittsburgh, PA, USA, Elsevier publ., 855 p., 2015.

[4] T.Kathirvalavakumar. "Neural Networks: FNN Training Algorithms: Simultaneous perturbation, Backpropagation and Tunneling methods", 450 p., 2010.

[5] N.Nikolaev, H.Iba. "Adaptive Learning of Polynomial Networks: Genetic Programming, Backpropagation and Bayesian Methods (Genetic and Evolutionary Computation)", 680 p., 2010.

Computer Modelling of Spatial and Electronic Structures of the Complex of PEG4 with KCl

I.N. Alieva
Baku State University
Z.Khalilov st.23, AZ-1148, Baku, Azerbaijan
iradanur@gmail.com

S.D. Demukhamedova
Institute for Physical Problems, Baku State University
Z.Khalilov st.23, AZ-1148, Baku, Azerbaijan
svetlanabest@mail.ru

Z.I. Hajiyev
Baku State University
Z.Khalilov st.23, AZ-1148, Baku, Azerbaijan
gadjiev_zahid @mail.ru

N.M. Gojayev
Qafqaz University, Baku-Sumqayit Road, 16 km, Khirdalan, AZ-0101, Baku, Azerbaijan
nqocayev@qu.edu.az

***Abstract*—By the method of molecular dynamics the structure of complex of polyethylene glycol (PEG4) with KCl ion pair was investigated. The calculated stable conformation of the PEG4+KCl was used as a base structure for following calculations by quantum chemical ab initio Hartree-Fock method using the 6-31G basis. On the base of calculated geometrical, energy and geometrical parameters of the complex a comparison of the resulting complex structure with the previously studied complexes PEG4+K and PEG4+KCl was carried out.**

Keywords-complex, polyethylene glycol, oligomers, quantum-chemical calculations, spatial and electronic structure.

I. INTRODUCTION

In recent years, using a latest achievements in computer technology, it becomes possible to carried out computer simulations of various molecules and their complexes. Such molecular modeling is necessary to determine the spatial structure, electronic structure, and other characteristics of polyatomic molecules, which play an important role in the study of the properties of these molecules. In many cases they can not be studied by experimental methods in detail The main attention is focuses to the study of systems that are based on polymer films with nanoparticles. In particular, it will be useful in the search for new drugs and materials with predetermined properties. For receiving new materials with specific properties the orderly arrangement of the particles in an original sample were used. Also the design and production of the of drug delivery system directly to the diseased cells are very popular. In this case the specific role assigned for the polymers with cations and anions of different metals. Such complexes also were used for investigation drug delivery system in cancer cells. Experiments testified [1], that polymer+ ligand complexes immediately begin to accumulate in the tumors and increased the efficiency of treatment. Experimentally established [2,3] that polyesters in particular polyethylene glycol, due to its ability to bind with the alkali metal ions to form stable complexes can be used as polymer. Such complexes are of great practical use [2]. Mass spectrometric methods of studying the structure of complexes of polyesters alkali metal were used in [3-6].

PEG4 and PEG5 polyethylene glycol oligomer complex spatial structures with atoms of sodium, potassium, chlorine and NaCl ion pair were earlier investigated by quantum chemistry methods [7-10]. In present paper PEG4 complex structure with KCl ion pair was investigated.

II. METODS

In this study the spatial and electronic structure of the complex which consists of polyethylene glycol (PEG) with repeating four monomeric HO-(CH_2–CH_2–O)$_n$-H units and KCl was investigated. Calculations were carried out in several stages. At first step by using ChemOffice programs package the PEG4+KCL complex structure was constructed. KCl ions pair placed in the center of this structure on distance about 2.8 Å. Molecular dynamics optimization showed that stable structure of the complex formed due to polyethylene glycol oxygen atoms interactions with potassium cation as well as PEG hydrogen atoms interactions with chlorine anions of KCl.

On the next stage ab initio Hartree-Fock method was used for calculation of geometrical, energy and electronic parameters of the investigated PEG4+KCl complex within GAUSSIAN-09 program package.

To further study the theoretical vibrational absorption spectrum of the complex, we took into account that the best way to calculate the frequency of the vibrational spectrum is DFT functional density method. Therefore, we carried out calculation of the PEG4 + KCl complex by DFT / B3LYP in a 6-31G basis (p, d).

III. CALCULATIONS, RESULTS AND CONCLUSIONS

The optimized model of the PEG4 + KCl complex is shown in the Fig.1 (a-b).

The results of the quantum chemistry calculations of the PEG4+KCl geometrical parameters are shown in Table 1.

We also conducted a comparative analysis of the results obtained by the HF/B3LYP method with the results previously studied complexes PEG4 with potassium and chlorine atoms.

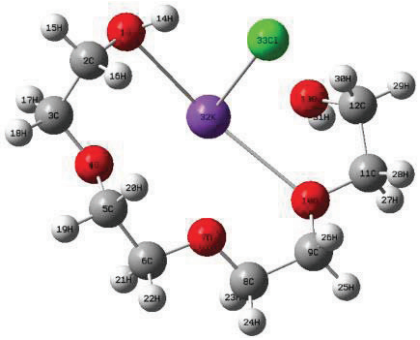
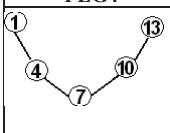
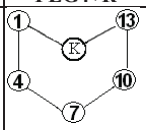
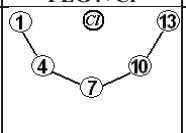
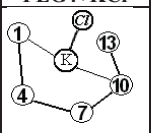


Fig. 1. The calculated model of PEG4 + KCl

Comparative analysis of the interatomic distances between non-bonded atoms of oxygen, potassium and chlorine was shown in Table. 1. As seen from this table the distance between terminal O_1-O_{13} atoms decreases until 2,53 Å in comparison with the pure PEG4, and the inner atoms of oxygen O_4-O_{10} getting closer by 0.26 Å. Placed in the cavity diagonally atoms O_4-O_{13} and O_1-O_{10} are approaching each other at 1,14 Å. Potassium atom is placed on a cavity of PEG4 so that it turns on 0,09 Å closer to the terminal oxygen than to the oxygen placed within the cavity. Thus potassium atom closest to the terminal oxygen atoms at a distance 2,78 Å, forming coordination bonds and is located furthest from the oxygen atom O7 representing mid PEG4 chain. Neighboring oxygen atoms lying within the cavity, approaching each other at 0.1 Å, and the distance between adjacent terminal oxygen O_1-O_4 and O_{10}-O_{13} PEG chains does not change. Formation of the PEG4+Cl complex does not change distances between

TABLE 1
THE INTERATOMIC SPACES (Å) BETWEEN ATOMS OF OXYGEN, POTASSIUM AND CHLORINE WITH OXYGEN ATOMS AND NEIGHBORING HYDROGEN ATOMS IN PEG4 COMPLEXES

PEG4		PEG4+K	PEG4+Cl	PEG4+KCl
				
O_1-O_4	2.78	2.78	2.79	2.91
O_4-O_7	2.87	2.78	2.87	2.85
O_7-O_{10}	2.87	2.78	2.87	2.77
O_{10}-O_{13}	2.78	2.77	2.79	2.75
O_1-O_7	5.38	4.93	5.38	4.70
O_1-O_{10}	6.59	5.46	6.76	5.51
O_1-O_{13}	7.31	4.78	7.63	5.52
O_4-O_{13}	6.59	5.45	6.78	5.71
O_7-O_{13}	5.38	4.93	5.39	4.86
O_4-O_{10}	4.88	4.62	4.98	4.87
K –O_1		2.78		2.81
K –O_4		2.87		2.90
K –O_7		3.11		2.87
K -O_{10}		2.87		2.81
K -O_{13}		2.78		2.90
Cl –O_1			3.90	3.16
Cl –O_4			3.55	4.65
Cl –O_7			3.69	4.50
Cl -O_{10}			3.52	4.28
Cl -O_{13}			3.88	4.38
Cl-H_{14}			3.14	2.31
Cl-H_{30}				2.89
Cl-H_{16}				3.39
Cl-H_{31}			3.11	5.31
K-Cl				3.01

non-bonded neighboring oxygen atoms O_1-O_4, O_4-O_7, O_7-O_{10} and O_{10}-O_{13} within PEG4 while the distance between the terminal oxygen atoms increases by 0,32 Å. Internally non-adjacent oxygen atoms O_4 and O_{10} drifting to 0,12 Å, and diagonally located in the cavity O_4-O_{13} and O_1-O_{10} oxygen diverge at 0.19 Å and 0,17 Å, thereby expanding the complex cavity. Chlorine atom located in the PEG4 cavity so that is closer to the oxygen atoms O_4 and O_{10}. At 0.4 Å further it is from the terminal oxygen atoms O_1 and O_{13} and 0.2 Å away from the O_7 atom. But chlorine atom closest to hydrogen atoms of H_{14} and H_{31} (about ~ 3.1 Å).

Structural adjustment in PEG4, occurring at the formation of the complex PEC4 + KCl are similar to rearrangements occurring during the formation of the complex PEC4+K. However, due to interactions of potassium and chlorine atoms with oxygen and hydrogen atoms of the polyethylene glycol chain, there are determined differences. The distance between the terminal oxygen atoms $O_1 \div O_{13}$ is increased on 0.74Å, between atoms $O_4 \div O_{13}$ and $O_4 \div O_{10}$ on 0.25Å and only the distance between the terminal O_1 and O_7 atoms, located in the middle of the cavity is decreased on 0.23Å. In the complex PEC4+K the potassium atom combines with the neighboring terminal oxygen atoms O_1 and O_{13} by coordination bonds. In the complex PEC4+KCl the terminal O_{13} atom moves away from the potassium atom on 0,09 Å, as a result of a potassium atom is moving closer to the oxygen atoms O_1 and O_{10} (2.81Å), with which it is forming the coordination bonds. Unlike complex PEC4+K, in which the potassium atom is located symmetrically with respect to the lateral oxygen cavity, in the complex PEC4+KCl potassium is in symmetrical position with respect to diagonally located oxygen and it is closer on 0.24Å to the central atom O_7 of polyethylene glycol chain. It is located on 0.06 farther from the O_7 atoms and on 0.09Å from the atoms O_4 and O_{13}. At the formation PEC4+KCl complex compared to PEC4+Cl complex chlorine atom is moving away from all of the oxygen atoms, unlike of the terminal atoms O_1 and O_{13}, to which it is closer on 0.74Å and 0.88Å, respectively. As a result of interaction of chlorine atom with the hydrogen atoms it is approached to H14 hydrogen on 0.83Å. It is located close to distance of 2.31 Å and 2.89 Å from the hydrogen atoms H_{14} and H_{30}, which keep him in the complex. It should be noted that of all the investigated complexes is the most compact complex PEG4 + KCl.

TABLE 2
THE CHARGES ON ATOMS IN OPTIMIZED PEG4 COMPLEXES

Atom	PEG4	PEG4+K	PEG4+Cl	PEG4+KCl
O_1	-0.649	-0.650	-0.651	-0.685
C_2	0.104	0.084	0.103	0.097
C_3	0.101	0.101	0.102	0.086
O_4	-0.653	-0.655	-0.653	-0.669
C_5	0.122	0.115	0.123	0.100
C_6	0.118	0.087	0.114	0.107
O_7	-0.631	-0.642	-0.635	-0.653
C_8	0.118	0.087	0.113	0.103
C_9	0.122	0.115	0.123	0.108
O_{10}	-0.653	-0.655	-0.654	-0.663
C_{11}	0.102	0.101	0.103	0.099
C_{12}	0.104	0.084	0.102	0.040
O_{13}	-0.649	-0.651	-0.651	-0.651
K		-0.152		0.760
Cl			-0.003	-0.854

Tables 2-4 show the results of calculating the electronic structure of complex PEG4 + KCl as well as a comparative analysis of the results of previously studied PEG4 and its complexes PEG4+K, PEG4+Cl. Table 2 presents the charges on the atoms of the complexes calculated by HF / B3LYP method. During PEG4+Cl complex formation chlorine atom is not located in the cavity of the PEG4, but placed at some distance above it. Therefore, we do not see changes in the charge distribution, the maximum changes observed on the atoms of the C_6, C_8 and O_7 which don't exceed 0,005 units of charge. As it is seen from Table 2, at the PEG4 + K complex formation charges on all atoms are reduced as compared with pure PEG4. The greatest change in the distribution of charges observed on the carbon atoms of the C_6 and C_8 (~ 0.03), C_2 and C_{12} (~0,02), as well as on the oxygen atom O_7 (~ 0.01). The formation of PEG4+KCl complex is accompanied by a lowering of charges on all atoms. The most changes are observed on the C_{12} carbon atom (~ 0.064) and on the oxygen atom O_1 (~ 0,036). On the atoms C_5 and O_7 observed a decrease charge on 0.022, and on the atoms of O_4, C_3, C_8 and C_9 by 0.015 units of charge.

It is known that investigated complexes formed due to interactions O and H^+ atoms of the polyethylene glycol (PEG) with potassium and chlorine atoms. So the more stable PEG4+KCl complex structure can be formed because positive charge on the potassium atom and negative charge on the chlorine greatly increased in comparison to PEG4+K and PEG4+Cl complexes.

Calculated by the HF/B3LYP method energies of the highest occupied HOMO and lower LUMO orbitals of PEG4 oligomer and its complexes with potassium, chlorine and ion pair KCl are summarized in Table 3. The energy difference HOMO and LUMO orbitals for the α-electrons decreases at the formation of PEG4complexes with potassium (about 0.125) and the ion pair of KCl (about 0.344) atomic units.

TABLE 3
HOMO AND LUMO ORBITALS IN OPTIMIZED POLYETHYLENE GLYCOL COMPLEXES

Complexes	PEG-4	PEG4+K		PEG4+Cl		PEG4+KCl
Orbitals		α	β	α	β	α
HOMO	-0.419	-0.087	-0.439	-0.421	-0.227	-0.312
LUMO	0.229	0.038	0.050	0.224	0.225	0.033
Δ	0.647	0.125	0.489	0.644	0.452	0.344

Table 4 shows energy characteristics, and the dipole moments of the studied complexes calculated by the HF/B3LYP method. As shown in Table 4, the most energetically stable complex is PEG4+KCl, whose energy was about 1058.7 atomic units lower than the energy of PEG4.

TABLE 4.
ENERGY PARAMETER AND DIPOLE MOMENTS OF PEG4 AND ITS COMPLEXES

Energy and dipole moment	PEG4	PEG4+K	PEG4+Cl	PEG4+KCl
Total energy (RHF, au)	-687.708	-1286.844	-1147.159	-1746.446
nuclear repulsion energy (Hartrees)	855.947	1182.534	1087.141	1457.331
Dipole moment (Debye)	4.778	11.302	4.755	10.971

Formation of PEG4+K and PEG4+Cl complexes lead to the decreasing energy compared with a net PEG4 respectively at 599 and 459 atomic units. Complex PEG4+KCl was energetically favorable and the most compact among the all the investigated PEG4 complexes. Thus, our theoretical calculations testified experimentally observed results.

Finally on the base of the calculated spatial and electronic structure of PEG4+KCl complex the quantum chemical calculations of the vibrational spectrum of this complex by the density functional DFT method using hybrid potential B3LYP with 6-31G (p, d) basis set were carried out. Frequency and intensity of the vibrational bands are calculated. Theoretical absorption spectrum of the PEG4 + KCl complex is shown in Fig. 2.

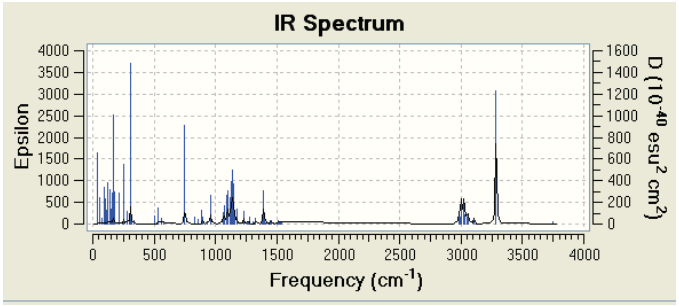


Fig. 2. Theoretical absorption spectrum of the PEG4 + KCl complex calculated by the DFT/B3LYP/6-31G method.

The most intense band in the high range of 3286 cm^{-1} corresponds to the characteristic stretching vibrations of the terminal O_1H and O_{13} H bonds of the PEG4 cavity. The most intense bands corresponding to the stretching vibrations of CH bonds underlying at the 3108-2981 cm^{-1} field are the 3019 and 3005 cm^{-1}. The field between 1539-1502 cm^{-1} corresponds to low intensity of the characteristic bending vibrations HCH bond angles, of which the most intense band is 1513 cm^{-1}.

The following bands are observed corresponding to a mixture of deformation vibrations of the HOC, HCC, HCO angles and non-planar deformation vibrations polyethylene glycol chain HCOH, HCOC, HCOC, the peak of which corresponded to the absorption of 1391 cm^{-1} band. Then there are very intense absorption band at 1150, 1139 and 1131 cm^{-1}, corresponding to the stretching vibrations of OC bonds. Stretching vibrations of CC bonds correspond to the absorption bands in the region 1064-890 cm^{-1}. The absorption bands of 855, 833 and 825 cm-^{1} are also correspond the stretching vibrations of OC bonds. An intense band of 750 cm^{-1} is determined by the non-planar oscillation terminal H_{14} hydrogen atom from the plane of the cavity. Bending vibrations of the CCO and COC are in the range of 556-299 cm^{-1}. The 278 cm^{-1} corresponds to non-planar oscillation of OCCO cavity. An intense band 308 cm^{-1} arises due to deformation vibrations of bond angles COC and CCO. Potassium and chlorine atoms are involved in the appearance of low frequency absorption bands 214.165, 154, 145, 119, 104, 89 and 78 cm^{-1}.

The analysis of the theoretical spectrum of the complex PEG4+KCl can help experimentalists in studying the experimental absorption spectra of similar complexes.

REFERENCE

[1] M. O. Durimanov, A. B. Ulasov, E. A. Belechkaya, N. S. Rodichenko, Yu. B. Rozenkranch, Yu. B. Xramchov, A. A. Rozenkranch, A. S. Sobolev. "Polyplex ligand nanoparticles to deliver genes to melanoma

tumor cells," IVth Russian biophysicists congress. Symposium I "Physicochemical bases of functioning of biopolymers and cells". Nizhny Novgorod, pp. 80, 2012. (In Russian)

[2] J. Chen, S. K. Spear, J. G. Huddleston, R. D. Rogers. "Polyethylene glycol and solutions of polyethylene glycol as green reaction media," Green Chem., vol. 7, No. 2, pp. 64-82, 2005.

[3] M. B. Kosevich, V. G. Zobnina, E. N. Jivotova, I. V. Shigol, O. A. Boryak, V. V. Chagovets, V. V. Chekanova, A. V. Zinchenko, V. A. Pokrovskiy, A. Gomori. "Mass spectrometric study of cryoprotectants on the basis of ethoxylated glycerin derivatives," Mass spectrometry, vol. 6, No. 1, pp.7-20, 2009. (In Russian)

[4] J. Gidden, T. Wyttenbach, A. T. Jackson, J. H. Scrivens, M. T. Bowers. "Gas-Phase Conformations of Synthetic Polymers: Poly(ethylene glycol), Poly(propylene glycol),and Poly(tetramethylene glycol)". J. Am. Chem. Soc., vol. 122, No. 19, pp. 4692–4699, 2000.

[5] T. Wyttenbach, G. von Helden, M. T. Bowers. "Conformations of Alkali Ion Cationized Polyethers in the Gas Phase: Polyethylene Glycol and Bis[(benzo-15-crown-5)-15-ylmethyl] Pimelate," Int. J. Mass Spectrom. Ion Proc., vol. 165/166, pp. 377-390, 1997.

[6] Van L. E. Vlerken, T. K. Vyas, M. A. Mansoor. "Poly(ethylene glycol) modified nanocarriers for tumor targeted and Intracellular Delivery," Pharm. Res., vol. 24, No. 8, pp. 1405–1414, 2007.

[7] S. D. Demukhamedova, Z. I. Hadjiev, I. N. Alieva, N. M. Godjaev, "Spatial structure of the polyethylene glycol oligomers with potassium atoms," Ist International chemistry and chemical engineering conference CCE, Baku, pp.834-837, Aprel 2013. (In Russian)

[8] Z. I. Hadjiev, S. D. Demukhamedova, I. N. Alieva, N. M. Godjaev. "Electronic and spatial structures of the polyethylene glycol PEG4 and PEG5 oligomers with sodium atoms," Journal of Qafqaz University-Physics, vol. 1, No. 1, pp. 40-46, 2013. (In Russian).

[9] S. D. Demukhamedova, Z. I. Hajiyev, I. N. Alieva, N. M. Godjaev. "Theoretical calculation of spatial and electronic structures of the complex of oligomer of polyethylene glycol PEG4 with NaCl ion pair," Fizika. vol. XX, No. 4, pp. 15-22, 2014.

[10] S. D. Demukhamedova, Z. I. Hajiyev. "Electronic structure of the complexes of polyethylene glycol PEG4 and PEG5 oligomers with a chlorine atom," VI International Scientific and Practical Conference "Theoretical and applied aspects of modern science", Part I. Belgorod, pp.18-21,. 2015.(In Russian)

Performance Analysis of e-Archive Invoice Processing on Different Embedded Platforms

Salih Bayar
Research and Development Department
Idea Teknoloji Solutions Research Center
Istanbul, Turkey
salih.bayar@ideateknoloji.com.tr

***Abstract*—In this paper, an embedded Linux-based platform named Digital Invoice ARchiving InSTrument (DIARIST) is proposed. While the DIARIST system is located on the tax payer's side, it is responsible for the processing of e-Archive invoices, it distinguishes e-Archive invoices from e-Invoices, it communicates with the remote server and it allows tax-payers to print their e-Archive invoices for their customers on local printers. In this paper, different embedded platforms for the DIARIST system are compared and the experiences are stated during design and development processes. Experimental results show that the most of the effort is being consumed while rendering the invoice data from HTML format to PDF format. Although HTML to PDF conversion for each invoice is a computation intensive process, we achieved to process an invoice in about two seconds on the tax payer's side.**

***Index Terms*—e-Invoice, e-Archive Invoice, XML, XSLT, XSD, HTML, Schematron.**

I. Introduction

e-Archive Invoice is an application which allows issuing the invoice which is required to be issued, kept and submitted on paper as per Tax Procedural Law (TPL); to be issued on electronic environment and second copy to be kept and submitted on electronic environment pursuant to the conditions contained in TPL General Communique numbered 433 [1].

In this paper, we use our commercial product named Idea VisionPlus e-Archive which is the first authorized e-Archive service by Turkish Revenue Administration (TRA) in Turkey [2]. Taxpayers which sell products and services over internet and having a gross sales revenue of 5 million Turkish Liras and higher in 2014 income statement are obliged to switch to e-Archive application until 1/1/2016 at the latest. Taxpayers which fall within this scope are required to compete their e-Invoice and e-Archive applications and actual transition preparations before the established date.

Companies which switch to e-Archive application can create all of their invoices on the electronic media and keep them throughout the legal period. If their customers are final consumers or commercial customers that use e-Invoice application, corporations can submit invoices through the electronic media.

With the help of e-Archive application, corporations may minimize time, labor, correction, sorting and archiving costs using e-Archive Application which is a significant step to full automation in invoicing. In the case that the paper printout of the invoice issued within scope of e-Archive Application is affixed with stamp and signature and the invoice is issued at the time of delivery of the goods and handed by noting down" Consignment invoice.", the invoice is deemed as way-bill.

II. The System

The whole e-Archive system with multiple DIARIST systems is given in Figure 1. In the whole system there are multiple DIARIST systems, multiple sellers/ buyers, a single remote server and a single TRA.

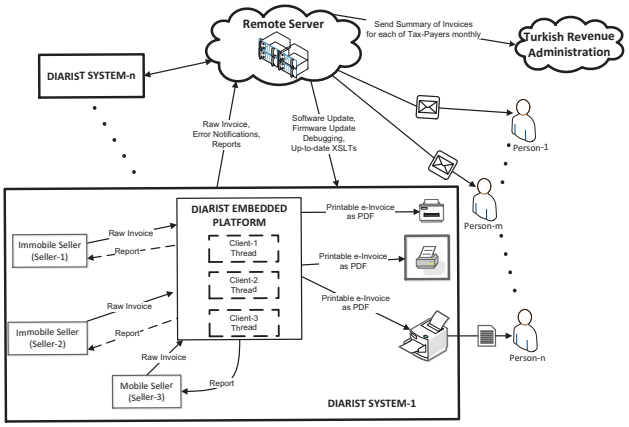


Fig. 1. Whole e-Archive System with Multiple DIARIST Systems

In this architecture, each DIARIST system is located on the tax-payer's location. There might be multiple sellers and printers, which are directly connected to a single DIARIST system. While each DIARIST system is independent from each other, each of them has a regular connection to the remote server. Here, the remote server is a private integrator. It is the approach used by taxpayers that are granted with private integration permit by the TRA, to benefit from e-Archive application through information system.

Taxpayers that require to use e-Archive application through private integrators can participate in the application by applying to private integrators which are certified by the Administration for e-Archive application. Our company (İdea

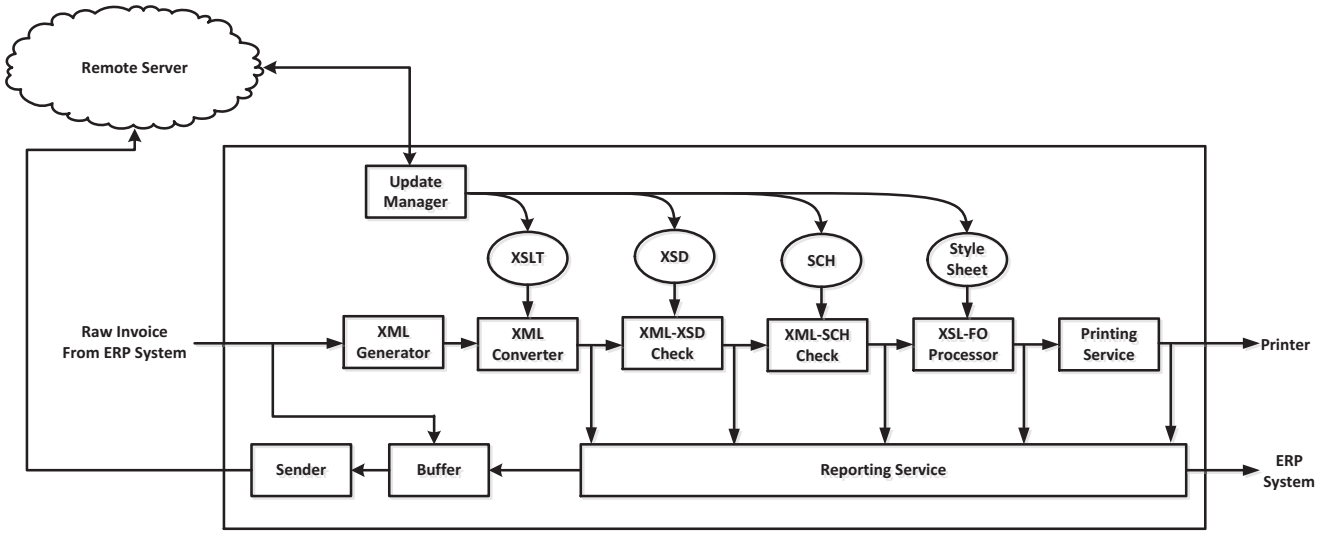


Fig. 2. Software Block Diagram of a Single DIARIST System

Teknoloji Çözümleri) is the first service provider receiving e-Archive service authorization from TRA in Turkey.

The DIARIST system is used by tax-payers, who are registered to e-Invoice system. In this approach, if the buyer is also registered to e-Invoice system, then the invoice can send to the buyer thorough the DIARIST system, private integrator (i.e. remote server) and TRA respectively. However, if the buyer is not in the e-Invoice system (e.g. personal customer or small enterprise that is not registered to e-Invoice system), then the type of the invoice must be e-Archive invoice and it must be given to the buyer as a printout. At this point DIARIST system decides the invoice type that is it distinguishes e-Archive invoices from the e-Invoices. If the invoice is an e-Archive invoice, then the invoice is printed by a printer which is connected to the DIARIST system on the tax-payer's site. As the invoices are generated from ERP systems for tax-payers, they are raw invoice data and must be processed on the taxpayer's site. Here, the DIARIST system is mainly responsible for the processing of e-Invoices.

The software block diagram of the whole invoice processing steps is shown in Figure 2. Each DIARIST system is connected to an ERP system, which is used as a source of Invoices. As the DIARIST system is compatible most of the popular ERP systems, it can be directly communicated to an ERP system using socket programming technology. To achieve this, ERP system has a small library of DIARIST in order to send raw invoices to the embedded DIARIST platform.

The raw invoice data, which is generated by the ERP system on the tax-payer's site, can be in the form of idoc, text, csv, excel file or an XML file. After raw invoice data is received from the ERP system, it is processed by the DIARIST system. As it can be seen from the Figure 2, in the first optional step, the raw invoice data is transformed into a predefined simple XML if the raw invoice data is not XML. After generation of predefined XML file for e-Invoice, it is converted to the UBL-TR XML format [3] with the help of its specific XSLT implementation for the corresponding customer. Extensible Style-sheet Language Transformations (XSLT) is a specific language for XML that can add/remove elements, sort or rearrange them and a lot more. However, an XSLT file implemented for one XML file could not be used for other XML file. Therefore, every XML file has its own XSLT implementation and this XSLT file is the secondary input file of XSLT transform.

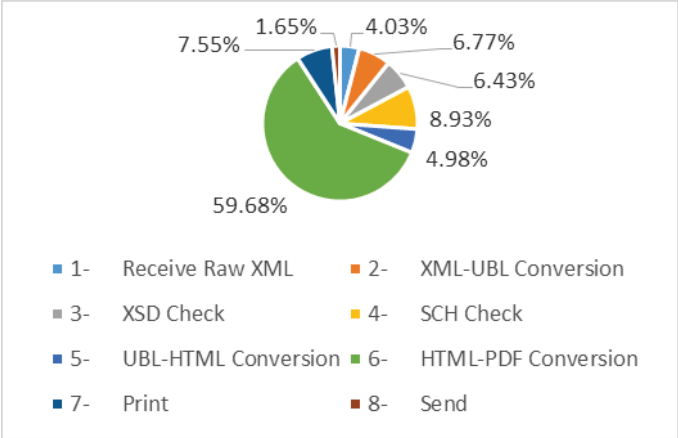


Fig. 3. Processing Time Percentages of a Single e-Invoice on BeagleBone Black (Rev B) Embedded Platform

After XML-Converter step, the UBL-TR XML file is checked synthetically by using XSD (XML Schema Definition). If the invoice is checked successfully by XSD, then is checked for content consistency by using Schematron, which is a rule-based validation language for making assertions about the presence or absence of patterns in XML trees.

After XML transformations and XSD, Schematron checks are completed successfully, the invoice data is converted to the HTML format as a printable invoice by using XSL-FO processor within the DIARIST system. While creating HTML, XSLT is not sufficient enough to accomplish all objectives. For

TABLE I. PERFORMANCE COMPARISON OF ODROID XU4 AND BEAGLEBONE EMBEDDED PLATFORMS FOR E-INVOICE PROCESSING

e-Invoice Processing Steps on the Embedded Platform	BeagleBone Black Rev B		Odroid XU4	
	Average Execution Time [ms]	*Percentage(%)*	*Average Execution Time [ms]*	*Percentage(%)*
1-Receive Raw XML	328	4,03	226,1	12,24
2-XML-UBL Conversion	550,5	6,77	67,5	3,65
3-XSD Check	522,8	6,43	59,6	3,23
4-SCH Check	726,7	8,93	70,2	3,80
5-UBL-HTML Conversion	405,2	4,98	36,2	1,96
6-HTML-PDF Conversion	4855,7	59,68	1272,1	68,87
7-Print	614	7,55	95,1	5,15
8-Send	134	1,65	20,4	1,10
Total Execution Time	8136,9	100	1847,2	100

TABLE II. DETAILED PERFORMANCE RESULTS OF TEN INVOICES ON BEAGLEBONE BLACK (REV B) [MS]

Invoice Number	Receive Raw XML	XML-UBL Conversion	XSD Check	SCH Check	UBL Check	HTML-PDF Conversion	Print	Send	Total
1	131	900	757	907	243	6290	811	321	10360
2	246	360	268	551	282	4608	447	60	6822
3	361	614	463	955	784	4728	80	67	8052
4	234	599	555	621	346	4601	498	86	7540
5	350	654	894	681	389	4775	421	67	8231
6	310	567	499	1275	317	5144	1500	440	10052
7	290	298	260	504	269	4632	446	94	6793
8	736	611	279	520	877	4582	1040	50	8695
9	283	334	811	480	271	4524	442	69	7214
10	339	568	442	773	274	4673	455	86	7610
Average	328	550,5	522,8	726,7	405,2	4855,7	614	134	8136,9
Percentage (%)	4,03	6,77	6,43	8,93	4,98	59,68	7,55	1,65	-

TABLE III. DETAILED PERFORMANCE RESULTS OF TEN INVOICES ON ODROID-XU4 [MS]

Invoice Number	Receive Raw XML	XML-UBL Conversion	XSD Check	SCH Check	UBL Check	HTML-PDF Conversion	Print	Send	Total
1	43	174	71	191	70	1936	135	74	2694
2	232	61	59	63	40	1198	89	13	1755
3	256	59	66	86	37	1204	89	12	1809
4	234	58	61	56	37	1203	100	12	1761
5	231	60	55	51	31	1194	86	15	1723
6	286	56	59	50	35	1206	94	14	1800
7	300	54	73	51	33	1192	89	21	1813
8	223	51	54	53	28	1188	93	15	1705
9	231	50	48	47	26	1193	89	15	1699
10	225	52	50	54	25	1207	87	13	1713
Average	226,1	67,5	59,6	70,2	36,2	1272,1	95,1	20,4	1847,2
Percentage (%)	12,24	3,65	3,23	3,8	1,96	68,87	5,15	1,1	-

this reason, Cascading Style Sheets (CSS) and JavaScript are also used. As a final step, obtained e-Invoice as HTML is converted to the PDF format before printing. After that it is sent to the printer to give e-Invoice to the buyer as a printout. During each step, necessary log files are generated by the DIARIST system. These log files are sent both to the ERP system and to the remote server for controlling.

III. EXPERIMENTAL RESULTS

In order to achieve a real-time system on the tax-payer's site, different embedded Linux-based platforms are used. Within the scope of this work, we firstly used Raspberry-I and -II embedded platforms [4]. However, the performance of these platforms were insufficient for the processing of invoice data real-time. After trying these platforms, we also used Beagle-Bone Black [5] and Odroid-XU4 [6] platforms. The performance of last two platforms seem to be sufficient for real-time e-invoice processing.

Figures 3 and 4 show the processing time percentages of a single invoice on BeagleBone Black (Rev B) and Odroid XU4 embedded platforms respectively. As it is obvious from both of these figures, the most of the processing time is consumed in the conversion from HTML to PDF.

In Table I, a detailed analysis is shown for performance comparison of Odroid-XU4 and BeagleBone Black embedded platforms. Here, we can see that processing time of a single

invoice on BeagleBone Black embedded platform is about 8 seconds, while it is about only 2 seconds for Odroid-XU4 platform. In our previous trials we used PDFKit and PhantomJS for HTML to PDF conversion. But both methods were too slow. Hence, at the end we decided to use wkHTMLtoPDF shell tool for this purpose. Although wkHTMLtoPDF shell tool is much faster than both PDFKit and PhantomJS, HMTL to PDF conversion seems to be still the bottleneck of the system.

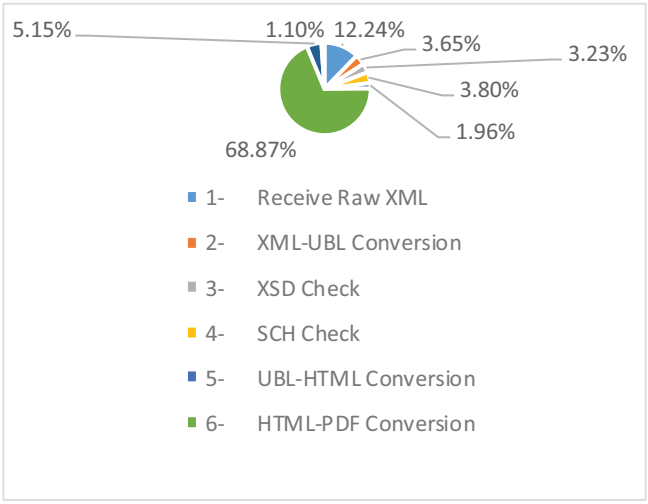


Fig. 4. Processing Time Percentages of a Single e-Invoice on Odroid XU4 Embedded Platform

In addition to the Table I, we also tested both embedded platforms for ten number of invoices and take the average execution time for each processing steps. The detailed performance results of ten invoices on both BeagleBone Black (Rev B) and Odroid-XU4 can be seen in Tables II and III. Here, it is obvious that execution times of some steps may change for different invoices. This is dependent on the invoice length and the system load at that time. Hence, for the calculation of performances, we took the average values of each processing step.

IV. Conclusion

This paper introduces development process of processing e-Invoice and e-Archive invoices with VisionPlus e-Archive application. The processing of e-Invoices and e-Archive invoices are done on the tax-payer's site without affecting the performance of private integrator performance in the manner of distributed processing system. As each tax-payer uses the same system, each of them has the same performance. Hence, the maintainability of these systems on the tax-payer's site will be very easy by remote desktop connections. As the communication between each DIARIST system is achieved by 3G/4G USB modem dongles, there will be no conflict during communication with the remote DIARIST systems.

V. Future Work

As the DIARIST system is sufficient for the real-time e-Invoice processing (2 sec. for processing a single e-Invoice), we do not plan any major change in the whole system. However, we are aware of that the HTML-PDF conversion is still the bottleneck of the system. In our future work, we will try different tools for HTML rendering process in order to speed up HTML to PDF conversion.

References

[1] Tax Procedural Law General Communique numbered 433, http://www.gib.gov.tr/node/87891

[2] Turkey's First Private Integrator Receiving E-archive Service Authorization from Turkish Revenue Administration, www.earsiv.visionplus.com.tr/home-page.html

[3] Universal Business Language Turkish Customization, www.en.wikipedia.org/wiki/Universal-Business-Language

[4] Raspberry Pi 1 Model B+ And Raspberry Pi 2 Model B, www.raspberrypi.org

[5] BeagleBone Black Low-Cost, Community-Supported Linux Embedded Platform, www.beagleboard.org/black

[6] Odroid-XU4 Linux Embedded Platform, www.hardkernel.com

The Synthesis of Reliable Solutions of the Logistics Problems Using Geographic Information Systems

Stanislav Belyakov
Southern Federal University
Taganrog, Russia

Igor Rozenberg
Public Corporation "Research and Development Institute of Railway Engineers",Taganrog, Russia

Marina Belyakova
Southern Federal University
Taganrog, Russia

Marina Savelyeva
Southern Federal University
Taganrog, Russia

***Abstract*— The paper proposes a method of constructing solutions based on the experience represented by cartographic images - centers and transformations of situations that do not change their meaning. The method is based on the metatransformation of images. The goal of metatransformation is to construct new situations by known precedents. The usage of metatransformations allows to evaluate the semantic proximity of situations and adapt the earlier decisions in a special way. The factors that determine the reliability of the generated solutions are analyzed.**

Index Terms— **Decision making, knowledge, precedent analysis, intelligent systems, cartographic visualization, geographic information systems, logistics.**

I. Introduction

Reliable decision-making using geographic information systems (GIS) is essential for many applications that use spatial data. Maps and charts show the real world and make it possible to take into account many factors of the environment that affect the outcome of planned actions. It should be noted that GIS not only provide important information for decision making, but also are able to maintain the experience of solving problems for later use. This feature is implemented by the intelligent geographic information systems and network services. A distinctive feature of this type of systems is the use of knowledge. We strive to achieve a high level of reliability due to knowledge – correspondence of decisions to the real state of the area of the external world, in which this decision will be implemented. In case of inconsistency the damage arises as a consequence of using inadequate decisions [1].

II. Case-based Reasoning In the GIS

The basis of receiving solutions based on the experience is precedent analysis. In the analysis of precedents observed earlier in accordance with the hypothesis of compactness we suggest that the situations with similar meaning have decisions with similar sense. Consequently, the task of assessing the proximity of precedent plays an important role in the selection of potentially applicable solutions. The difficulty of this task is the selection of valuation metrics of precedents' proximity. The metric reflects properties of the space of the observed situations and world view of the expert, which fixes the precedents. Shape of metrics must be selected and evaluated by subject area, model of its presentation and logic output reports.

The precedent analysis in relation to the use of material flows' management experience in logistics systems is considered in this paper [2]. The problem of material flow's management is to provide moving of material objects from one area to another area with a predetermined level of quality indicators. Movement of goods, products, materials, raw materials involves the construction of logistics projects and monitoring of their performance with the help of GIS. The logistics project is information model of the specific cargo delivery problem. The material flow in the logistics project can be described by a model.

$$< S_L, P_S, T >,$$

in which S_L is a set of spatially localized logistic centers P_S - the set of transport routes between the logistics centers, T - the schedule of cargos's delivery. Generally the logistics center is the spot of the performance of logistics operations with the cargo (packing, unpacking, loading, unloading, sorting, recycling, etc.). The mappable description of the flow reflects the referencing S_L and P_S to the conditions of the real world using spatial data for strategic planning and the operational control of the flow.

The mappable description of the logistics project can be built as a description of the situations change process [2]. Then the main decision making should be seen as a result of the cartographic analysis of situations, routes of transportation and schedules restrictions. Cartographic analysis can be performed in a conventional manner [3]. However, experience has shown

that this does not give the expected result for the following reasons:

- the GIS tools are for card, schemes and plans processing as figurative and symbolic models of the real world. The semantic space of situations and solutions are not directly displayed and requires the development of special models of representation of meaning in a particular interpretation. In particular, it is for materials management tasks;
- the concept of "sense of the situation" is interpreted by GIS analysts subjectively. Because of this, it is impossible to create a single semantic space of situations and achieve the reuse of the experience of decision-making. It takes the concept to unify the meaning of the situation in the context of decision-making;
- the concept of reliability is associated exclusively with the properties of spatial data. In practice, this is not enough, because the spatial data reflects the reality statically. The knowledge that will allow to predict dynamics of reality, appearance and consequences of decisions is necessary.

These problems can be solved by introducing a new concept of presentation and use of maps, charts and plans for the semantic description of situations.

The specific approach to the presentation of precedents in geoinformation systems is the presentation of situations by images [4]. The image of the situation

$$I_s =< c, H(c) >$$

is described by the image center (c) and the set of its transformations $H(c) = (h_1(c), h_2(c), ..., h_M(c))$. Meaningfully center is real the situation observed previously and the transformations describe possible transformations of situation's parameters from the point of view of the expert that do not change its meaning. Conversions concern spatial, temporal and semantic characteristics of the situation. Comparing of situations (evaluation of proximity) is based on a comparison of the mutual arrangement of the centers of images and their transformations. Thus, the creative thinking of humans is simulated. The logic of figurative comparison is described and the metric of evaluation of proximity of situations as the degree of generality of transformations h_i and two images I_1 and I_2 is offered in this paper

$$\alpha = 2S(h_i^{(I1)} \cap h_i^{(I2)}) / S(h_i^{(I1)}) + S(h_i^{(I2)}), i = \overline{1, M},$$

where $S(x)$ is the area of the field x. Value $\alpha = 1$ takes place at the coincidence of possible transformation $\alpha = 0$ otherwise. It can be seen that the possibility to compare images of situations solely in the common area maps is laid in this expression. If this is not the case,

$$S(h_i^{(I1)} \cap h_i^{(I2)}) = 0$$

The situations are considered as different in meaning. This feature guarantee the reliability of solutions through the use of deduction. However, this limitation makes it impossible to use the experience of decision-making in geographically remote situations. At the same time, the practice of the expert shows the need and the opportunity to compare "the same thing in different places". From the point of view of logic reasoning, in this case the deduction should be replaced by the abduction, aimed at generating additional facts and statements for reliable output. Therefore it is necessary to improve the principle of comparing the situation changes.

The aim of this work is to improve the reliability of the generated GIS solutions based on the application of the method of converting the images of situations.

III. Setting of the Problem of Reliable Solutions' Search

In general, the problem of finding the most reliable solutions using GIS is formulated as follows: there is a set of precedents of decision making in the form of images

$$I = \{I_n\}, n = \overline{1, |I|},$$

each of which includes a component of description of the situation and decision making

$$I_n =< I_{s_n}, I_{d_n} >.$$

Suppose there is the image of a problematic situation $\tilde{I} =< \tilde{I}_s, \tilde{I}_d >$ for which you want to find a solution. Let I^* the image of the solution which corresponds to a problem situation, and has a significant decision. The fact that such a solution does exist, can only be tested experimentally. Therefore, we assume that the decision corresponds to reality, if it is compatible with the experience of implementing solutions and consistently depicted on the map Ω, in other words, when you build it does not violate the rules, reflecting the connections and relationships between map items $R(I_d^*, \Omega) = true.$

Then the search for a fair solution of the problem situation is to build the image of the solution $\tilde{I}_d$ that is the closest to reality:

$$\begin{aligned} &\| I_d^* - \tilde{I}_d \| \to \min, \\ &\| I_{s_n} - I_s^* \| < d, n = \overline{1, |I|}, R(\tilde{I}_d, \Omega) = true, \end{aligned} \quad (1)$$

where $\| \bullet \|$ is the metric of distance between the images, d - is the restriction to the value of the distance. Analyzing the ways of solving the problem, the following should be noted:

- as the image I_d^* is not known, the problem should be solved constructively: scanning known precedents build them on the basis of their possible solutions and evaluate their reality through the correct cartographic representation;
- the procedure of searching the solution of the problem is unstable. Each case is a set of objects and relations, probability of exact repetition of that in other regions of space and time is almost zero. The reduction of accuracy of comparisons will not give the desired effect, as it unpredictably affects the accuracy of the solution. Therefore, when you use case descriptions images, which indicate not only the situation, but also their possible changes, the probability of finding a similar situation increases, the decision process becomes stable;
- the metric of proximity of situations should be determined through a conversion map objects as constructive search for a solution of problem is specified by Eq. 1 requires the display of images of precedents in a problematic spatio-temporal and semantic domain;
- the rules of consistency $R(I_d^*,\Omega)$ in the formulation of the problem solution of problem is specified by Eq. 1 are clearly marked, which creates certain difficulties of finding solutions: generation time of the alternatives would be useless after the validation rules. It seems appropriate to enable the restrictions to the consistency of cartographic mapping procedure.

IV. Metatransformation of Images

Statement of the problem solution of problem is specified by Eq. 1 has disadvantages which are overcome in the case of the introduction metatransformations that display images of situations and decisions in random areas of the map. Let

$$I_2 = H(I_1, c_2)$$

to be metatransformation of the image I_1 to image I_2 with the center c_2. Metatransformations' result is a set of map objects, which should be built in the area of the map given by the center c_2 . Metatransformation requirement is that all the objects and the relationships of the image I_1 have been kept in the image I_2 . This ensures the preservation of meaning and consistency.

For example, let us consider metatransformation of the vehicle of parking space. Transport parking is considered as the simplest logistics center, which is the part of the logistics chain. Parking image is described as follows: the image center is the point on the map, showing the location of the parked vehicle, conversions are zone of its possible location that do not affects the meaning of the parking operation. Let $\hat{P}_1$ to be the image of the parking precedent, $\hat{P}_2$ to be image of a problematic situation. Figure 1 displays the image $\hat{P}_1$ by a polygon, the image center is displayed by the point. The center of image $\hat{P}_2$ is set by coordinates of the point (x_1, y_1).

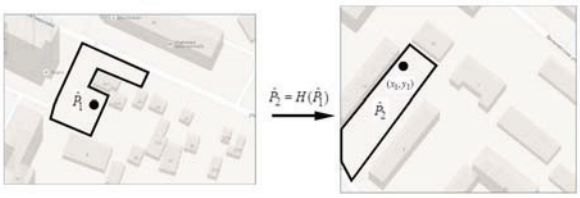


Fig1. The example of the metatransformation "Parking the car"

Metatransformation of "Parking the car" $\hat{P}_2 = H(\hat{P}_1, c_2)$, which displays the image $\hat{P}_1$ in the area defined by the center of the image c_2 at the point (x, y) builds a polygon object $\hat{P}_2$ of class "parking place" in the case:

- a point (x, y) located on the relation "LieWithin" with the object of one of the classes "parking place", "paid parking", "Free parking", "Asphalt platform", "road curb", and the object $\hat{P}_2$;
- the object is not in the relation "Cross" with the objects of one class "Buildings", "Facilities", "Utilities", "Pond", "Gardening";
- the object $\hat{P}_2$ has an area different from the area $\hat{P}_1$ not more than 30%, the dimensions $\hat{P}_2$ should not deviate from the size $\hat{P}_1$ of more than 10%;
- object $\hat{P}_2$ should be on the relation "Attached" to the object "Road" or "Driveway" or "Lane" or "Highway";
- the object $\hat{P}_2$ should not be located closer than 200 m from the object "railway-crossing";
- the object $\hat{P}_2$ should not be in the "No Stopping" zone of action of a road sign.

Figure 1 shows the result of the display $\hat{P}_1$ to the given point (x_1, y_1).You can see that metatransformation consists of geometrical construction of a convex polygon and verification of compliance of the restrictions listed above.

Metatransformation's internal structure is written as follows:

$$H: < I_{s_1}, I_{d_1} > \rightarrow < I_{s_2}, I_{d_2} >, \; H =< H_s, H_d >,$$
$$H_s: \; I_{s_1} \rightarrow I_{s_2}, \; H_d: \; I_{d_1} \rightarrow I_{d_2}.$$

We emphasize that metatransformation of image includes two components: metatransformation of the situation (H_s) and the metatransformation of solution (H_d). This type of metatransformation corresponds to the hypothesis of logical independence of situations and decisions made in these situations. This makes it possible to take into account the four cases in the process of search:

- $H_s(I_{s_1}, c_{s_2}) \neq \varnothing, H_d(I_{d_1}, c_{d_2}) \neq \varnothing$ means that the precedent with the image I_1 may occur elsewhere and the earlier decision may be used repeatedly;
- $H_s(I_{s_1}, c_{s_2}) = \varnothing, H_d(I_{d_1}, c_{d_2}) \neq \varnothing$ means that the precedent with the image I_1 cannot occur elsewhere, although the earlier decision is very applicable in this case;
- $H_s(I_{s_1}, c_{s_2}) \neq \varnothing, H_d(I_{d_1}, c_{d_2}) = \varnothing$ means that the precedent with the image I_1 can occur elsewhere, but its decision cannot be reused;
- $H_s(I_{s_1}, c_{s_2}) = \varnothing, H_d(I_{d_1}, c_{d_2}) = \varnothing$ means that the precedent with the image I_1 is unique so it cannot be in another place, and decided to use again.

These cases cover practically useful options for reuse of experience.

We assume that metatransformation exists if a non-trivial image from the map objects can be constructed:

$$H(I_1, c_2) \neq \varnothing \Leftrightarrow \ I_1 \neq \varnothing, I_1 \in \Omega, I_2 \neq \varnothing, I_2 \in \Omega.$$

An important role in the evaluation of the results of metatransformation plays the fact that initial conversations of images' center are described by vector of transformations $(h_1(c), h_2(c), ..., h_M(c))$. Each individual transformation $h_i(c)$ is the mapping object. Metatransformation's result can be considered as the restriction (for example, the polygon object is displayed as a linear one or a linear one as a point) or as an extension (linear object is displayed as a polygon or a point is displayed in the line). The neutral correspondence should be considered as a full correspondence of types of map objects. Then metatransformation rate can be defined as

$$\| H(\overline{h}(c), c) \| = \sum_i f(h_i(c)),$$

$$f(h_i(c)) = \begin{cases} f_0, \text{if the image objects are abscent,} \\ f_1, \text{if the objects narrow down the image,} \\ f_2, \text{if the image is neutral,} \\ f_3, \text{if the objects widen the image.} \end{cases}$$

The search of reliable solutions presented by the problem is specified by Eq. 1 can be reformulated as finding the image $\widetilde{I} = << \widetilde{c}_s, H_s(\widetilde{c}_s) >, < \widetilde{c}_d, H_d(\widetilde{c}_d) >>$ such as

$$\begin{aligned} &\| H_d(I_{d_n}, \widetilde{c}_d) \| \to \max, \\ &\| H_s(\widetilde{I}_{s_n}, \widetilde{c}_s \ \| < d, n = \overline{1, |I|}. \end{aligned} \tag{2}$$

The statement of the problem is specified by Eq. 2 characterized in the fact that the reliability measure is the number of conversions that displayed by metatransformations of images. The advantage of this approach is shown in the following.

Firstly, real-world conditions which will be applied the formed solution are continuously changed. The implementation of logistics projects with the same initial data is never exactly repeats itself. As a consequence, solutions are not repeated and their methods of modifying are not repeated. The more dynamic external environment and higher the sensitivity of decisions to the change these conditions are, the higher the probability to construct the unreliable solution is. The logic of case-based analysis is not compromised in this case. The reason for the unreliability of the solution is in the absence of the forecast of its use. After turning into a mechanism the case analysis of the evaluation of the alleged effect the accuracy of made decisions increases.

Secondly, the geographic information systems naturally support the prediction of the state of the real world. Geographical maps are never a snapshot of reality. There are generalizations laid in them that are the result of cartographic generalization. For this reason, the accuracy of the solutions found are continuously supported by updating the GIS cartographic basis.

Thirdly, the use of metatransformations increases the robustness of solutions. Knowledge about the stability of the solutions, their application invariants has great practical importance for logistics systems, avoiding large losses for small changes in transport conditions.

The search procedure of reliable solutions in accordance with the statement of the problem is specified by Eq. 2 consists of the following:

- we give category of images and metatransformation that will be used to search for solutions;
- we give the center of the problem situation $\widetilde{I}$, which is a set of map objects. For the tasks of management of material flows the center is the starting and ending point of the transportation;
- precedents of given category are selected of GIS knowledge base. If those are not available, we should change the category and take the next attempt of selection again. If there are no precedents the search procedure is terminated without result, owing to the lack of experience;
- we form a set of possible solutions of the problem. To do this, we run metatransformations of images of the situations to the center of the image of the problem situation and the value of metatransformation standards is controlled. If there are no images with relative meaning, the procedure is terminated without result;
- we search for the most reliable solution. For this purpose we make metatransformations of the images of the solution to the center of the image of the problem situation and select the variant with a maximum value of metatransformation standards.

V. Example of realization

Consider an example that illustrates the proposed procedure. Suppose you want to transport cargo from point **M** to point **N,** displayed in Fig. 2. The GIS-based knowledge has precedents for transportation $I = \{I_{AB}, I_{CD}, I_{EF}\}$. The map shows the centers of precedents images that have a form of a pair of points and arrows corresponding to the direction of transportation. The center of image shows the position on the map of the source and the destination transport point.

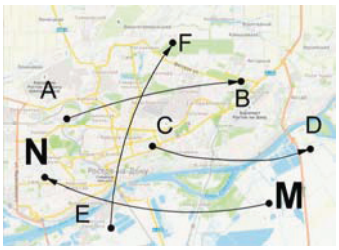


Fig. 2. The centers of images for transport tasks

Figure 3 shows the conversion of images of situations I_{AB_S}, I_{MN_S} that are relevant to the map fragments. Converting for I_{MN_S} is received by metatransformation for "Loading refuges" of point **A** to the **M** by metatransformation "Loadout area" of point **B** to point **N**. Each metatransformation retains the dimensions and area of map objects, relations with the surrounding objects, similar to those in the previous example. The transformations are shown by shaded areas. Let us assume that a metric of conversion vector is calculated by the formula

$$\| H(\tilde{h}(c), c) \| = \sum_i f(h_i(c)),$$

$$f(h_i(c)) = \begin{cases} 0, & \text{if the displayed objects do not exist,} \\ 1, & \text{in all other cases.} \end{cases}$$

Then the distance between the situations is 2, and when $d = 1$ the image I_{AB} is a close within the meaning of the image I_{MN}.

The solution for a precedent with the image I_{AB} includes 4 logistics centers $\{S_{L_1}, S_{L_2}, S_{L_3}, S_{L_4}\}$ and 3 transport sections $P_{S_1}, P_{S_2}, P_{S_3}$. On the Fig. 4 they are shown as elements of the image conversion schema. Conversion of logistics centers have the form of shaded polygons, conversions of the each transportation's section are presented by a set of broken lines, each of which indicates a possible path of the cargo.

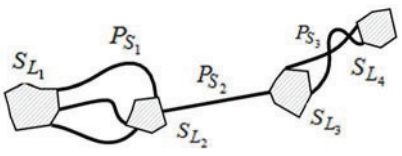


Fig. 3. The conversion of images of situations

Fig. 5. Metatransformation of the image's solution

The metatransformation of logistics center is performed in a manner specified above; the metatransformation of path's trajectory builds the shortest path between the given logistics centers. The chart of the result of metatransformation of the solution is shown in Fig. 5. The dotted line shows the trajectory of transportation, shaded polygons show logistics centers.

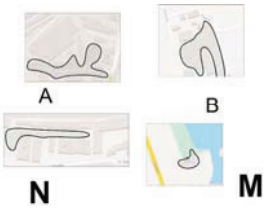


Fig. 4. Conversion of the image's solution I_{AB}

Likewise the solutions for the remaining images I_{CD} and I_{EF} are built. It is possible that each of the images will give a significant decision.

VI. Conclusion

The approach set out in this paper summarizes the mechanism of using the experience of decision-making in an environment of geographic information systems. Using the approach in practice will require the development of conceptual models of cartographic knowledge of a higher level. At this level the results of critical thinking of accepted solutions by the expert are accumulated. In the long term this should lead to an extension in the areas adjacent to the management of material flows. Further studies may be aimed at the effective implementation of metatransformations and the study of their behavior in the GIS cartographic base dynamics.

Acknowledgment

The reported study was funded by RFBR according to the research project № 15-01-00149a and 16-01-00090a.

References

[1] Y. Fang, V. Dhandas , E. Arriaga Spatial Thinking in Planning Practice . Portland: Portland State University, 2014.

[2] K. Pangbourne, S. Alvanides "Towards intelligent transport geography in Journal of Transport Geography," № 34, pp. 231–232, 2015.

[3] J. Malczewski "GIS-based land-use suitability analysis: a critical overview," Progress in Planning, vol. 62, pp. 3-65, 2004.

[4] S.L. Belyakov, A.V. Bozhenyuk, M.L. Belykova, I.N. Rozenberg "Model Of Intellectual Visualization Of Geoinformation Service," in proc. 28th European Conference on Modelling and Simulation ECMS 2014, pp.326-333, 2014.

Recommender System Providing Recommendations for Unidentified Users of a Commerial Website

Igor N. Gluhih, Ivan Y. Karyakin
Information System Department
Tyumen State University
Tyumen city, Russia
igluhih@utmn.ru, ivan_nyc@mail.ru

Lyudmila V. Sizova
Department of Foreign Languages and Intercultural Professional Communication in Natural Sciences
Tyumen State University
Tyumen city, Russia
milasizova@yandex.ru

Abstract— **Recommender systems are a popular trend in recent research in Internet technologies. The models and algorithms of these systems are based on applying information of users and website content as well as their interconnections. However, there is a problem of applying these models and algorithms when the users are unidentified and there is an information gap to give recommendations. Each study case appears as a pair of a situation and a set of possible recommendations with their characteristics. The paper offers to solve the problem of generating recommendations through the adaptation process of the initial set of recommendations into which the recommendations are included on basis of criteria of similarity of the main and recommended contents. The adaptation process uses the iteration formula optimizing the recommendation utility function.**

Keywords **- recommender systems; unidentified user; search systems; recommendation utility; cold-start.**

I. INTRODUCTION

Recommender systems have become one of the recent important trends in Internet technologies [1-3]. Properly selected product, service or content recommendations make the website more attractive for its users. It is beneficial for a website owner because it provides user loyalty to the website, improve its ranking in search systems, ensures growth in income. To generate the recommendations meeting users' current demands and preferences is an important task for developers of recommender systems.

Various methods are developed to solve this problem. They are classified as follows: collaborative filtering, content-based filtering, and hybrids [1]. Collaborative filtering uses the information about user activity and their social relations, takes into account recommendations and opinions of different people belonging to the same group. Content-based filtering is grounded on the knowledge of the content and the user himself/herself without regard to the opinions of other people. Hybrid methods combining collaborative and content-based filtering are used for giving recommendations.

In all these cases, the information about a user plays a key role. The lack of this information at the early stages of the operation engenders one of the main problems in this area. That is a cold-start [1]. There is another problem receiving little attention in research. It is the problem of giving recommendations to unidentified users [4]. The matter is that major online retailers (search systems, social networks, advertising platforms, analytical systems, big e-shops) have a wide knowledge of users, their behavior and relationship. The access to this information allows identifying or recognizing any new user of the website as the one belonging to a certain group of people. On basis of this information, he/she can be offered a personalized content or recommendations for the products meeting his/her current interests and preferences.

However, as a rule ordinary site owners have no access to global knowledge of users. In particular, it is relevant to commercial corporate sites used when promoting and selling goods or services in small and medium-sized business. Considerable proportion of e-commerce website population comprises the people who visit the site for the first time and in some special way convey no information about their interests, preferences, social relations, etc. to it. The sources of such visitors are search systems and online advertising service providing up to 70-80% of inbound traffic [4]. When dealing with such users, e-commerce website enhancement continues to be relevant.

This paper offers a solution to the problem of generating recommendations for unidentified users of e-commerce site. This solution is based on the use of recommendation utility and serial adaptation of many offered recommendations.

II. CONCEPTUAL MODEL

The utility shows the website owner's view of the benefit provided by the recommendation for the purposes of the tasks allotted. The recommendations virtually present the description of products, articles, video or other content that appear in a special block of recommendations. If a user is interested in the recommendation, he/she moves to the recommended page. In the simple case, moving to the recommended page is beneficial for the website owner as it makes positive impact on behavioral factors of site ranking in search systems. In more complex cases, the recommendation utility is confirmed by planned actions on the page (completing an order form, signing-up, following the link, etc.).

In this paper, we consider a base case with a single-criterion utility function of the i-th V(Ri) recommendation that is calculated as ratio of ni to Ni: V(Ri) = ni/Ni, where Ni is the total number of views of the recommended page; ni is the number of page transitions meeting the task-based constraints. In the base case, this restriction is the time for page viewing which must be higher than the threshold.

We introduce the concept of the initial set of recommendations R = {R1, R2, R3, ...}. This set includes the pages of the site meeting the criteria of similarity to the page where the user gets recommendations. To determine the similarity of pages we use the similarity estimation methods described in [5].

The next task is to separate a subset of the relevant recommendations of the maximum utility. This problem is solved by optimizing the aggregate recommendation utility V(R) = Σ V(Ri) →max, which is calculated according to the results of the recommendations presented to users in the previous period. Thus, the users on the page with a specific content vote for this or that recommendation they have received after analysis of the similarity of the content between main and recommended pages. Following the results of this voting, a package of recommendations is formed. It becomes clear in the following steps-iterations. Figure 1 shows a general scheme of giving recommendations.

The difficulty is that the preferences of the users, to whom the recommendation package is presented, are changeable. This means that after completing the optimization task the result is unfixed in final form and the set (R) must be adapted taking into account both these changes and the emergence of new content on the site. Further, we offer a mathematical model and procedure of such an adaptation.

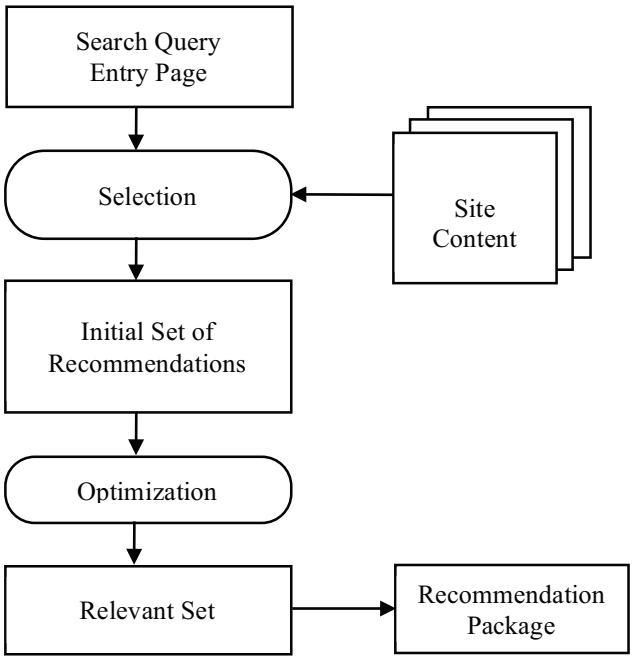


Figure 1. General diagram of giving recommendations

III. Mathematical Model

Let us match each i-th recommendation to its relative frequency: λi = [0; 1], Σ iλ= 1. Then it is possible to present a set of recommendations as many pairs of R = {<Ri, λi> | i = 1, ..., N }, where N is the power of R .

The adaptation process is divided into steps-iterations, j = 0.1, 2, ..., M, where j = 0 – the iteration of "cold-start" which has the input of the initial set of recommendations - R^0, and $\lambda^0 i$ is equal to 1/N for all i.

In the output of each j-th iteration, relative frequencies are recalculated and R^{j+1} = {<Ri, λj+1i >} is formed.

The iteration formula of frequencies recalculation is derived from the known gradient optimization formula:

$$\lambda_i'^{j+1} = \lambda_i^j + t\left(n_i^j / N_i^j\right) \Big/ \left\| n^j / N^j \right\| \quad (1)$$

where t is the coefficient of adaptation speed, ||…|| is the vector norm.

We offer to calculate t in the following way:

$$t = 0 \text{ при } \left(n_i^j / N_i^j\right) < w \quad (2)$$

$$t = \left(n_i^j / N_i^j - w\right) \Big/ (1-w)(r-1) + 1 \text{ при } \left(n_i^j / N_i^j\right) \geq w \quad (3)$$

Here, w is the threshold of uncertainty which initially set value is 0.5; r ≥ 1 – the parameter setting the t rate of change when the utility function $V^j(R_i)$ exceeds the threshold of uncertainty. It has an impact on the sensitivity of adaptation process to the change in users' preferences expressed in changing of utility at the j-th iteration. The normalized vector λ is passed to the input of next iteration:

$$\lambda^{j+1} = \lambda'^{j+1} \Big/ \sum_i \left(\lambda'^{j+1}\right) \quad (4)$$

Recommendations with the nonzero frequency λ, calculated at the output of the previous iteration, are considered relevant. They form a recommendations unit for the next (j +1)-th iteration, in which R is included at the calculated frequency of views.

When the recommender system operates, the process of adaptation is in operation, too. It monitors the selection of users and responds to a change of their interests or emergence of new website content, i.e. new elements in set ***R***. It becomes possible due to recomputation rules of w, r parameters on the basis of V^j (R) and $V_{thr.}$ comparison in last iterations.

IV. Application

The module of the recommender system (MRS) becomes integrated in the content management system (CMS) to apply this model practically. The code of MRS is put into the specified pages of the website. It generates a set of recommendations and collects data on behavior of users on the website. When the website page loads for the first time, the initial set of recommendations ***R*** is formed. It is performed automatically by

a request from database (DB) for the pages, headers, and meta descriptions meeting the requirements of the recommender system.

Further, the procedure of adaptation starts. As far as it collects information, it specifies content and metric values of a set of recommendations.

Using the MRS module editor allows monitoring the model parameters, specifying an eigenfunction of the target, implementing criteria and rules to form ***R*** as well as identifying kinds of the elements from the website database that may act as recommendations (information content, goods of the company, commercial proposals, etc.). Tracing and statistics analysis of the recommendation selection become another important resource.

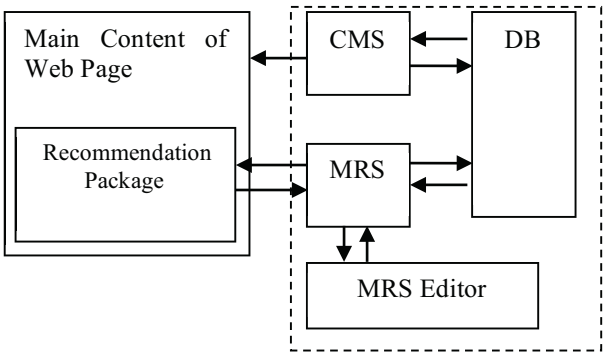


Figure 2. Application scheme of work

V. CONCLUSIONS

The paper reviews the problem of generating recommendations for unidentified website users, the information of whom is insufficient for applying the traditional methods of recommender systems.

We offer a solution of this problem through the adaptation process of the initial set of recommendations. The recommendations are included in the set on basis of criteria of similarity of the main and recommended contents.

The adaptation procedure uses the iteration formula that optimizes the utility function of recommendations. By this formula, at the output of each iteration the values of relative frequencies of recommendations showing are recalculated at the next stage.

The offered mathematical model of adaption procedure allows you to use it both at a cold-start stage and at tracking changes in users' preferences. This provides inclusion of the new recommendations in the recommendation package which previously were not demanded or that emerged as a result of the new additions to the site content.

We suggest creating the functional module RMS integrated in the website content management system to implement this model. Thus, it provides a possibility to manage an individual recommender system of the commercial corporate website that is independent of external user information sources and takes into account the preferences of the website owner.

Further development of this approach involves taking into account multiple criteria when determining the function of recommendation utility. Another area of the study becomes the improvement of the methods for selecting the initial set of recommendations for which we plan to use the case based reasoning [6, 7].

REFERENCES

[1] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation", Egyptian Informatics Journal, vol. 16, 2015, pp. 261–273.

[2] Sung-Woo Byun, So-Min Lee, Seok-Pil Lee, "A recommendation System Based on user's Content Consumption Patterns and Objects of the Interest", International Journal of Applied Engineering Research ISSN 0973-4562, vol. 11, num. 9, pp. 6457-6461, 2016.

[3] Kourosh Modarresi, "Computation of Recommender System Using Localized Regularization", ICCS 2015 International Conference On Computational Science. Procedia Computer Science, vol. 51, 2015, pp. 2407–2416

[4] I.N. Gluhih, M.V. Gubin, "Optimizazia vydachi contenta v dinamicheskom internet-proekte na osnove analiza vyborov neidentificirovannyh polzovateley", Vestnik Tyumen state university, vol. 6, pp. 182-188, 2010.

[5] V.N. Zaharov, A.A. Horoshilov, "Avtomaticheskaja ocenka podobia tematicheskogo sodergania tekstov na osnove sravnenia fomalizovannyh smyslovyh opisaniy", International Conference RCDL'2012, 2012, pp. 135-142.

[6] J.L. Kolodner, "An Introduction to Case-Based Reasoning", Artificial Intelligence Review, vol. 6, 1992, pp. 3-34.

[7] I.D. Watson, F. Marir, "Case-based reasoning: A review", The Knowledge Engineering Review, vol. 9, num. 4, 1994, pp. 355-381.

Intelligent Control System For The Rail Transportations

F.F.Pashchenko
V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences
Moscow, Russia
feodor@ipu.rssi.ru

N.A.Kuznetsov, I.K.Minashina, N.G.Ryabykh, E.M.Zakharova
Moscow Institute of Physics and Technology (State University)
MIPT
Moscow, Russia

***Abstract*— One of the problems to be solved in the management of rail traffic is the problem of attaching the traction resources to the trains. In this paper, we will examine in detail two algorithms that are used for performing this task - stochastic algorithm and genetic algorithm. The second problem is schedule optimization problem. This paper presents auction method for solving this problem.** *(Abstract)*

***Index Terms* — Rail transport control system, freight scheduling problem, assignment problem, auction method, global extreme, optimization problem.** *(key words)*

I. Introduction

Intellectualization of the transport systems of today is the most important and forward-looking activities in the railroad freight activity and resource management. This approach is in accordance with solving one of the most difficult and time-consuming tasks as construction of optimal schedules. The complexity of this problem is depends on the system parameters and technological constraints. In this work the best schedules or solutions finding was reduced to the problem of finding the global extremum. At the same time to find the most optimal schedule is the difficult problem because of necessary ensuring only one existence of this extremum. The task of attaching the traction resources can be divided into the following sub-tasks: the assignment trains to the timeslots and the assignment traction resources to those trains.

II. Problem description

A. Train scheduling problem

The main difficulty of trains to timeslots assignment is a limited pool of timeslots and travel periods. According to rail traffic scheduling technology, trains can be assigned to specific time periods - timeslots, which are considered input data to the solving problem. The input system serves all of the available timeslots for the next day (24 hours). This problem has been reduced to the assignment problem, i.e. finding the optimal schedule of the trains assignment. Each train has initial location and route - the stations through which the train must pass. The main task is to find the most available timeslots to trains, so that this choice was the most optimal. In addition, each train can be assigned to a time slot partly, several trains may occupy one time slot for various segments, and the train can use multiple slots to reach their destination.

B. Locomotive scheduling problem

This problem can be formulated as follows: at the entrance serves locomotives, train, train routes that are attached to them, time intervals. Train route contains information about all stations which should be visited with exact times of departure and arrival for each station. For locomotives passed their characteristics, the station at which it is at the moment, as well as its state. The solution to this problem is to build a locomotive train destination plan that provides the most efficient use of traction resources

C. Assignment problem description

One of the classical mathematical problems is the assignment problem, which is an example of a combinatorial problem of optimization.

In general form this problem can be formulated in following way:

- A given number of workers and a number (M) of tasks (N). Each worker can perform a work piece (P) ($0 \leq P \leq N$), where the performance of each employee depends on the type of work performed by them - it makes some tasks more efficiently, some - other way around. The task is to distribute the work a way that maximizes the overall efficiency of all assignments. For each pair <task, worker> given the utility function, this allows to quantify the effectiveness of the appointment of the employee for this task. Thus, the optimization problem is to maximize the overall effectiveness of ΣUij all purposes, taking into account the fact that one work one person can be appointed.

Train and locomotive scheduling problems can be modeled as an assignment problem in a similar way. By way of example, let's consider how to formulate the problem of assigning trains to slots as a classical assignment problem:

- As workers are considered trains that you want to assign to the time intervals. The main requirement for the decision will ensure that all employees were assigned to work, i.e., for each train would have found such a segment.
- In addition, each employee can perform several tasks, i.e. Train can be assigned to several time intervals in order to proceed to its destination Moreover, some < train, slot > pairs are impossible, for example, when route of a slot is totally different from train route and they have

no common partial route within their routes. Each train can be assigned to several time intervals or on different parts of the same length.

- The assignment problem is solved for all trains leaving at some point with a given station.
- The input data for this problem is a set of trains, which are should be assigned. At the first stage, this is a lot of trains that can departure at the moment, as well as trains for which the station is an intermediate point of the route.
- After the decision of the first planning sub tasks, we can determine the time of arrival at the next station for each train. Therefore at the next iteration all trains assigned at the previous iteration would turn into "trains previously partially assigned and soon arriving at a given station" at the next station of train routes.
- Thus, at the beginning of each iteration of the algorithm is calculated number of trains to be set for a certain time interval at each station. After solving the assignment problem for each station and for each time interval, which is sent at least one train, we get a final decision.

III. Utility function calculation

Quantitative estimating of optimal decision is the most difficult in this problem formulation. For this purpose the utility function of each given pair *of <train slot>* or *<train locomotive>* was introduced.

We will not dwell in detail on the identifying of assignment criteria, such criteria were determined and introduced for train assignment problem and locomotive assignment problem in the papers [4], [5].

Each criteria is put in correspondence with a normalized value u_k, then the utility function value for each < *train*, *slot* > pair is calculated as $U_{ij} = \sum_k c_k u_k$, where c_k weight of k-th criterion. Weights can either be chosen experimentally or be assigned a specific value for each fragment.

The calculation of this function is useful only for a particular set of timeslots and locomotives, which theoretically can be attached to train. Other "impossible" timeslots (locomotives) utilities remain unassigned. The slot can be considered "impossible" for the train if the train route coincides with a timeslot route even partially, or slots designed to train the category into which the train does not fall. A set of criteria can be assigned separately for each fragment.

IV. Model Structure

The process of train control was divided into two phases. The first is the solution of the assignment problem, which was solved by the auction method [3].

This method can find a global extremum function, but is less effective than other optimization algorithms. That is why the development of the objective function extremum seeking algorithms required for this type of problem.

The following steps were performed for the solution of these problems:

- Developing of a mathematical model of the railway traffic management system.
- Solving the assignment problem using auctions
- Developing of searching algorithm for the most optimal locomotive assignment.
- Comparison of the global optimization techniques with reference to indicated problem.
- Simulation developed freight management system on Far Eastern Railway and analysis of the efficiency of proposed searching algorithms.

V. Solving the assignment problem using auctions

The assignment problem is a special case of the transportation problem, which in turn is a special case of linear programming problem. This type of problem is solved by the classical simplex method, however, the use of a specialized method provides greater convergence rate.

The best known way to solve the assignment problem is a "Hungarian algorithm" - one of combinatorial optimization algorithms. It was proposed in 1955, its complexity is $O(n^4)$, where n is the number of jobs and workers (the number of jobs must be equal to the number of employees). The algorithm can be modified to achieve $O(n^3)$ complexity. The Hungarian method has been implemented for the slot assignment problem for the trains, but it has slow convergence rate. Therefore, an auction method was developed another algorithm for solving the assignment problem.

An assignment problem solving via auctions was proposed D. Bertsekas in 1989 and then in 1992 it was changed Bertsekas and Castañón [2]. Its initial purpose was for solving skewed assignment problems where the number of jobs is not equal to the number of employees. The main concepts of the auction algorithm applied for the train assignment problem are listed in the paper [4].

Mathematical proof of convergence of auction algorithm, as well as a guideline for choosing values for λ and ε is given in [2].

VI. Developing of optimization algorithm

The problem of finding optimal schedule may be approximated by the global optimization problem, whose complexity of the number of system parameters and process limitations. Search global extremum is a common problem in various areas of science. There are a large number of different optimization algorithms to solve real problems of searching for the best solution. However, since it is necessary to take into account the specifics of the technological criteria for analyzing the effectiveness of each approach, the task of train control optimization has gained complexity of spatial variables phase and its limitations. Therefore, most of the known algorithms are not effective for the construction of complex systems with a large number of constraints, specific criteria and fuzzy structure [1]. In this work we were considered some of the algorithms search for the optimal solution for this problem. One of them relates to the methods of mathematical programming, the second - to the soft computing theory. These algorithms can be used at various

planning stages. The hybrid optimization algorithm based on the methods of stochastic approximation and coordinate descent. Genetic algorithm refers to soft computing techniques. GAs are direct, parallel, stochastic method for global search and optimization which imitates the evolution of the living beings described by Charles Darwin [5]. The system was built in such a way that in the first phase to search the local extremum. This is the starting point for further optimization, and search has global extreme.

The problem of the appointment of the example the destination of locomotives on the train, as well as both of the optimization algorithms.

A. Locomotive assignment optimization problem

The task assignment locomotives for trains

To quantify the optimality of the solution of the assignment problem of locomotives on the train was also determined the total utility function:

$F(x) = \sum_{ij} Uij = \sum_i Ui(xi)$ (1)

This feature is a multidimensional feature vector argument $X = [x_1, x_2, \ldots, x_m]^T$ where vector X is the solution to the problem of the appointment of the locomotives on the train. X is represented as $(loco_1, loco_2, \ldots, loco_m)^T$, where $loco_i$ is a unique locomotive number assigned to the i-th plan and m - the total number of such plans. Because of this, this problem is reduced to the problem of finding the extremum of function with appropriate restrictions:

$F(X) = \Sigma_i U_i(x_i) \rightarrow extr(max)$; $X^* = argextrF$; $g_j(X^*) \le b_j$, $j = 1, \ldots, k$ (2)

where X* is the best solution; $g_j(X^*) \le b_j$ - constraints of the problem; K - number of limitations. In general extremum search algorithm presented in the following form:

$X^{n+1} = ||A||\, X^n$, $||A|| < 1$ (3)

where A - operator iteration.

This algorithm allows you to find the extremum of the given function on the set of technological and operational constraints. It is necessary to find such form of iteration of A, by which the process (3) converge to the best solution, ie, to the extremum, which is in turn represented as $X = [x_1, x_2, \ldots, x_m]^T$ with the limitations of the form

$g_j(x_1, x_2, \ldots, x_m) \le b_j$ for all $j = 1, \ldots, k$ (4)

under which the maximum of the function $F(X)$. Valid is a solution that satisfies all constraints (4) for all $j = 1, \ldots, k$.

Totality of admissible decision is called domain D. The final solution is a pair $(x^*, f^*(x^*))$ of the optimal solutions and value functions.

B. Hybrid optimization algorithm

In the course of a detailed study of the problem revealed that the search function of the form $F(X)$ is a non-trivial task. This difficulty arises because of the strong dependence of the vector components X apart. Due to the lack of completeness limitations as $g_i(x_1, x_2, \ldots, x_m) \le b_i$ would use well-known methods for the local or global optimization is not possible. Therefore, we developed a special hybrid algorithm based on coordinate descent methods and stochastic optimization, which is presented in detail in [5]. This approach allows you to find the extremum of the objective function $F(X)$ on the given constraints trivial kind.

C. Genetic algorithm

GA is one of the methods of soft computing theory and adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. This approach is based on the idea of the theory of natural selection, "survival of the fittest" or the fittest. In this paper we consider the adaptation of this method to the problem of the appointment of locomotives on the train. This research was worked by the Russian Science Foundation (grant no. 14-19-01772) project. The input of the algorithm is served: the number of individuals in one generation, the maximum number of generations, and the number of deaths in a single generation. In more detail the algorithm is considered in [5]. The output of the algorithm selects the fittest, is responsible for the appointment of the best option.

VII. RESULTS

A. Results of the auction algorithm work

To compare and select the most effective algorithm was implemented Hungarian algorithm and auction method using AgentSpeak language to implement multi-agent approach in both cases. Basic agents are slots agents, stations agents, and the scheduler agent serves to synchronize all the elements of a central system and the formation of the final solutions. The difficulty of this approach is caused by a large number of partners with conflicting interests, a variety of criteria for decision making, uncertainty, dynamics and other factors. Adapting and finding a solution that satisfies both general and interests of each party an advantage multi-agent approach. When tested on the data size of 200 * 200 revealed that the algorithm based auction method converges to the solution faster than Hungarian. 20 seconds vs. 3-4 minutes. The given time estimate shows that auction algorithm converges in Nlog(N) time, which is a significant improvement compared to $O(n_4)$ of Hungarian method.

Later an auction algorithm was implemented using Java as part of Intelligent Control Systems in Railway Transportation (ISUZHT) project that deals with the freight scheduling problem. The system is used for scheduling rail transportation East Siberian Railway [5] . Planning was performed in a network of 350 stations. Train routes typically consisted of 5 to 100 stations. The overall number of trains to be assigned is circa 2000. The performance of the java application with the data described above is~3 minutes that is an acceptable amount of time.

B. Results of the hybrid optimization algorithm work

To test the implementation of the hybrid optimization algorithm following test data were obtained: 6-hour period, with 199 scheduled trains, the region with 220 stations, 240 train plans, and 163 locomotives with their parameters. The first stage was obtained the initial solution purpose locomotives on trains, which was subsequently used as a starting point for similar hybrid optimization algorithm. When

testing the operation of the hybrid algorithm we analyzed the optimization procedure with different parameters to optimize satisfy convergence conditions [5]. The dependency function of the normalized summary objective function F n Norm depending on the number of iteration was analyzed:

$F^{n}_{Norm} = F(X^{n})/N_{F}m$ (5)

where N_F is normalizing coefficient of summary objective function.

Also, analysis of the convergence of the global extremum was carried out depending on the number N of iterations and optimal steam parameters were obtained through this to solve this problem. Hybrid optimization algorithm has improved the initial solution to the problem of appointments of 7%, which in turn indicates higher efficiency in comparison with the original decision. Also, some of the qualitative evaluation of the effectiveness of the developed system based on the quality indicators of the use of the locomotive fleet was given. As a result, the resulting multi-agent control system provides efficient and high-quality solution for rail traffic planning problem.

C. Results of the genetic optimization algorithm work

The GA gives better results than the starting GA "greedy" algorithm on the test data in the independence of the optimization criteria. The efficiency of using resources traction also significantly higher than that of the initial "greedy" solution

VIII. Conclusion

Railway is one of the most important elements for the solution of transportation problems. For most efficient use intellectualization of the existing control system in this area is the most important problem. In this paper we considered multiagent approach using the auction method to develop the most optimal solution of the assignment problem. Hungarian algorithm and the auction method were compared. As a result of this research the auction method has faster convergence rate and has been developed as a part of intelligent management system ISUZHT.

However, the optimization problem is to find the optimal solutions. In this area, the task is characterized by a large number of parameters and technological limitations of railroad freight activity management. In order to optimize the initial decision hybrid and genetic algorithms have been implemented, both of which showed improvement in convergence rate and performance.

Acknowledgment

Research was supported by grant of The Ministry of Education and Science of Russia, unique identifier - RFMEFI58214X0003.

References

[1] F.F. Pashchenko, Vvedenie v sostoiatel'nye metody modelirovaniia sistem; Ucheb. posobie: V 2-h ch. Ch. 1. Matematicheskie osnovy modelirovaniia sistem. Ch. 2. Identifikatsiia nelineınyh sistem, M: Finansy i statistika, 2007.

[2] Bertsekas, D., Castanon, D. A forward/reverse auction algorithm for asymmetric assignment problems. Computational Optimization and Applications, 1992, 277 – 297.

[3] Zavlanos M. M., Spesivtsev L., Pappas G. J. A Distributed Auction Algorithm for the Assignment Problem. In Decision and Control, 2008, 47th IEEE Conference, 1212 — 1217.

[4] N.A.Kuznetsov, F.F.Pashchenko, N.G.Ryabykh, I.K.Minashina, E.M.Zakharova, O.A.Tsvetkova / Procedia Computer Science 00 (2015) 000–000

[5] N.A.Kuznetsov, I.K.Minashina, F.F.Pashchenko, N.G.Ryabykh, E.M.Zakharova Design and analysis of optimization algorithm for multi-agent rail control system, 2015, Informatsionnye Protsessy, 2014, Vol. 14, No. 3, 256 — 274.

Bioinspired approach for 3D packaging problem

Vladimir Kureichik, Alexander Glushchenko, Vladimir Kureichik Jr., Anton Orlov
Department of CAD
Southern Federal University
Rostov-on-Don, Russia
vkur@sfedu.ru, alex-14-93@mail.ru, kureichik@yandex.ru, ky92@mail.ru

***Abstract*— The article deals with one of the most important optimization problems - the problem of three-dimensional packaging (3DP) of various sized elements. It belongs to the class of NP- hard and complex problems. The formulation and restrictions of the 3DPP are considered in the article. Bioinspired approach that uses a multi-level evolution and partially allows to avoid a preliminary convergence of algorithms. Genetic and evolutionary algorithms which obtain sets of quasi-optimal solutions in polynomial time were developed. To carry out computational experiments on test examples (benchmarks) the authors developed a software on the basis of bioinspired approach for the 3DPP. Conducted tests and experiments allow possible to clarify the theoretical estimations of algorithm time complexity. In the best case algorithms the time complexity is represented as O (n^2), in the worst case - O (n^3).**

Index Terms*— Three-dimensional packaging; containers packaging; bioinspired approach; genetic algorithm; genetic operators; evolutionary algorithm

I. Introduction

3DP of various sized elements is a problem that occurs in many areas of industry, associated with the solution of such problems as the optimal filling of the empty volume, loading of the ships holds, freight trains, pallets, transport aircrafts, control of finished products warehouse, etc. In the modern world in terms of developing industries and manufacturing a delivery of products and goods in the distribution sphere become more urgent (moving, reloading, storage of goods, formation of transport batches etc). So, it is necessary to analyze and improve transport and cargo information processes, as well as the development of new methods of support and optimization of information processes in container terminals [1]. These problems are NP-complete [2]. This means that even with a small amount of cargo solution of thee problems is not possible in polynomial time, even with the use of modern supercomputers. On the other hand, there are additional constraints during the placing of the load in predetermined space. Therefore, an urgent and important task is the development of new approaches, methods and algorithms based on bioinspired search [3-5].

II. Dimension Packing Problem Formulation

The 3DPP can be formulated as follows: we have area of 3D space with width *W*, length *L* and height *H*. Also, it is given a set of blocks $A=\{a_n\}$, a number of blocks is *N*. The set of blocks is divided into subsets of types $T=\{tj\}$. Each type determines the set of parameters for each block, such as length, width, height, weight and the default orientation in space [6,7]. It is necessary to place a set of blocks in a given dimension. Output data are a plan of blocks packing in the space, value the objective function and time of problem solving.

Each type of cargo is characterized by the 4 tuple $<l_i, w_i, h_i, m_i>$, where l_i, w_i, h_i are dimensions of block, m_i is block weight. In the course of problem solving position of specified type elements, defined by the set $S=\{s_i=(<x_{1i}, y_{1i}, z_{1i}>, <x_{2i}, y_{2i}, z_{2i}>) \mid i=1, 2, \ldots, n\}$ in space, where $<x_{1i}, y_{1i}, z_{1i}>$, $<x_{2i}, y_{2i}, z_{2i}>$ are corners coordinates - the closest to the beginning of coordinate axes, and the most distant corner of element respectively. Also a box (parallelepiped) M={L, W, H} is determined, where L, W, H are dimensions of the field [6].

The authors introduce restrictions taking in to account a criterion. V that means a ratio of packed blocks volume to box, V_{con} volume required for packing of all blocks [7]. There are:

1) the height of all blocks is the same;
2) none of the element can not extend beyond the predetermined volume, i.e. satisfy the system of inequalities:

$$\begin{cases} x_{1i} \geq 0; \\ y_{1i} \geq 0; \\ z_{1i} \geq 0; \\ x_{2i} \leq m_x; \\ y_{2i} \leq m_y; \\ z_{2i} \leq m_z. \end{cases}$$

3) the total amount of the elements can not exceed the packaging field:

$$\sum_{i=1}^{n} (l_i \cdot w_i \cdot h_i) \leq L_x \cdot W_y \cdot H_z.$$

4) absence of intersections (it is impossible to put two objects in one area):

$$(x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \ \& \ z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \ \& \ z_{2i} \geq z_{1j}) \vee$$
$$(x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee$$
$$(x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \ \& \ z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \ \& \ z_{2i} \geq z_{1j}) \vee$$
$$(x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \ \& \ z_{2i} \geq z_{1j}) = 1 \forall i \leq n, j \leq (i \neq j).$$

5) the total weight of all blocks cannot exceed the specified limit

$$\sum G \leq G_{max},$$

where G is a block weight;
G_{max} is maximum weight of all blocks.

6) each block must locate on the surface of another block or box;

7) elements must be perpendicular to predetermined area.

Optimization criterion is a volume occupied by blocks. The objective function is represented as follows:

$$F = \frac{\sum V_{bl}}{V_{con}} \rightarrow 1,$$

where V_{bl} is a volume of each block, V_{con} is a container volume.

This means that it is necessary to reduce free space in packing area. The maximum efficient result of problem solving is reached in case, when the objective function is equal to 1, i.e. packaging area is filled to 100%.

III. Bioinspired Approach For Blocks Packaging

Bioinspired approach presented in Fig. 1 combines special procedures of genetic and evolutionary search [3,4,8-11].

In the developed approach optimal solutions search uses several random directed search algorithms (genetic and evolutionary algorithm. Solutions search usually is continued while at least one of the stopping criterion is reached. In this paper, the stopping criterion is an achieving a certain quality of solutions or exceeding of the algorithm running time. Let us consider proposed bioinspired approach based on multi-level evolution [3,4].

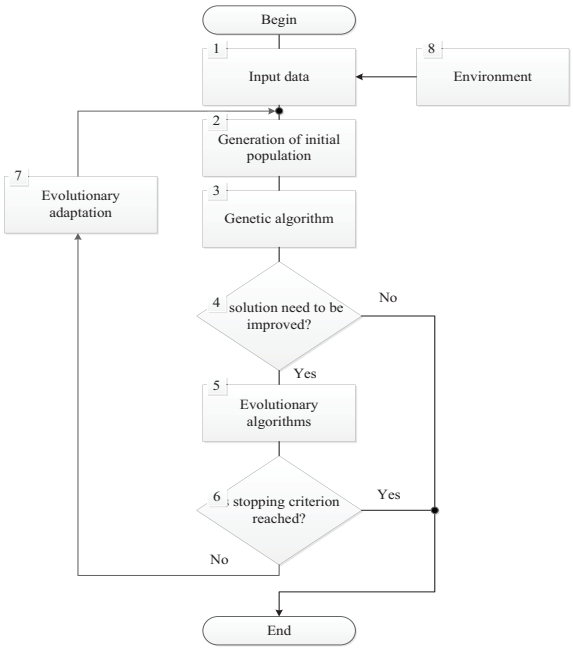


Fig.1. The integrated bioinspired search algorithm

In the first stage there are input such data and parameters, as size of the packaging field, a number of blocks, a probability of operators using, size of population, a number of iterations, stopping criteria.

The environment (decision making person) affects to the “Input data” block. Decision making person (engineers, operator) is based on personal experience and statistic data as well predetermined restrictions of an element base. Decision making person (DMP) chooses values of input data and parameters which control hybrid search based ob bionspired algorithm. DMP inputs such data and parameters as: 3D packing size; block parameters; a number of blocks; a probability of genetic operators using; size of population; a number of iterations; stopping criteria.

Further, the "Generation of the initial population" Block initialize a set of alternatives of the packing problem. The initial population is generated with the help of heuristic, based on a sequential algorithm. After that it is estimated average value of OF for the population, as well as the value of OF for each alternative solution.

Suggested bioinspired approach is as follows. There is implemented an idea of a multi-level evolution with the two-level hybrid approach. At the first level, the genetic algorithm is performed until it reaches local optimum. In the genetic algorithm new solutions in the population formed by various genetic operators: a crossover, a mutation and a reduction [8]. This algorithm has large computational complexity, however, guarantee the emergence of a qualitatively new solutions with common features that allows us to implement efficient search within a specific area. Then, there is created a new population on the basis of the best solutions. This population is a second-level object for rapid evolutionary search. In the evolutionary algorithm new solutions by the GA, are generated with the use of just one mutation operator [10].

After the evolutionary algorithm it is checked on the work completion according with the criterion of exceeding the maximum allowable time of the algorithm. In case of obtaining a new high quality solution and the stopping criterion is reached, the best result is saved. Otherwise, evolutionary adaptation is applied. Evolutionary adaptation is to adjust search process, that continues iteratively until, the stopping criterion is reached.

Stop criteria of bioinspired search are the follows:

$$\Delta > \Delta Z,$$

where Δ is a variable that determines convergence OF values, ΔZ is given by a decision maker person;

$$t < T_{max},$$

where t is running time of the algorithm, T_{max} is maximum allowable time of the algorithm.

Note, that when receiving effective solutions are received by the genetic algorithm, the search process can be finished.

Application of the proposed bioinspired approach produces effective solutions in polynomial time, and through the use of two-level approach it allow to solve the problem of pre-convergence algorithms, as well as to consider a more promising solutions and remove pre-unpromising ones.

IV. Experimental Results

To solve the 3DPP a software environment was developed. The authors used Visual C ++, Borland C ++, Builder. Debugging and testing of the developed algorithms was performed on the IBM PC with core i5 processor with 6GB of RAM. Computational experiments were carried out.

To determine the effectiveness of the developed bioinspired approach been it was studied time and quality of solutions for different set of test cases (benchmarks), different number of blocks [12]. The results of these studies are presented in Table. 1 and Fig. 2.

TABLE. 1. COMPARISON OF THE HYBRID ALGORITHM WITH PEERS

A number of blocks	Ngoi et al.	Bishoff et al.	Gehring et al.	Developed Bionspired Search
шт.	%	%	%	%
100	61,9	63,1	61,8	62,9
200	68,8	66,9	67,7	70,1
300	74,6	73,8	75,7	77,9
400	80,5	84,1	83,1	85,1
500	89,1	88,9	90,6	91,9
Average value	74,98	75,36	75,78	**77,58**

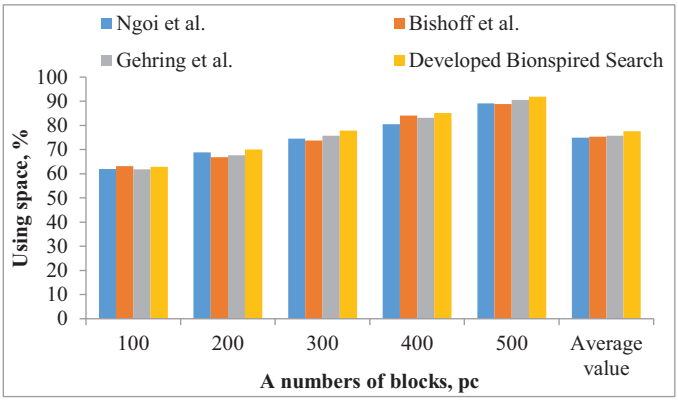


Fig. 2. Comparison bar chart of the packing blocks efficiency on the basis of different test circuits by different algorithms

Quality of packing obtained on the basis of the developed bioinspired approach is higher on average 2.2% than packing results obtained using known algorithms suggested by Ngoi et al., Bishoff et al., Gehring et al., which demonstrates the effectiveness of the proposed approach.

Experimental calculations have shown that the use of the bioinspired approach allows to obtain a set of local optimal solutions in polynomial time.

V. CONCLUSION

The bioinspired approach for solving the 3DPP were developed. A distinctive feature of this approach is the use of multi-level evolution in which different algorithms are used at different levels. At the first level the classic genetic algorithm uses different genetic operators, and at the second level - fast evolutionary search. This approach allows to obtain sets of optimal and quasi-optimal solutions in polynomial time, and partially solve the algorithms pre-convergence problem. A software environment in C ++ were developed. Computational experiments were carried out. Conducted series of tests and experiments made possible to clarify the theoretical estimations of the algorithms time complexity. In the best case, algorithms time complexity is O (n^2), in the worst case - O (n^3).

ACKNOWLEDGMENT

This research is supported by grants of the Ministry of Education and Science of the Russian Federation, the project #8.823.2014.

REFERENCES

[1] Lutsan, M.V., Nuzhnov, E.V., Kureichik, V.V. “Self-learning of the containers service coordinator agent in multi-agent automation environment of transit cargo terminal (2015).” Advances in Intelligent Systems and Computing, 347, pp. 109-117.

[2] Kacprzyk, J., Kureichik, V.M., Malioukov, S.P., Kureichik, V.V., Malioukov, A.S. “General questions of automated design and engineering (2009).” Studies in Computational Intelligence, 212, pp. 1-22.

[3] Yurevich Zaporozhets, D., Victorovna Zaruba, D., Kureichik, V.V. “Hybrid bionic algorithms for solving problems of parametric optimization (2013)” World Applied Sciences Journal, 23 (8), pp. 1032-1036.

[4] Bortfeldt A., Gehring H. “A hybrid genetic algorithm for the container loading problem” / A. Bortfeldt, H. Gehring // European Journal of Operational Research. – 2001. – Vol. 131, Iss. 1. – P. 143–161.

[5] Kureichik, V.V., Zaruba, D.V. “The bioinspired algorithm of electronic computing equipment schemes elements placement (2015)” Advances in Intelligent Systems and Computing, 347, pp. 51-58.

[6] Chauny F. “A Bloc Heuristic for the Container Loading Problem” / F. Chauny // Grouped' étudeset de recherche en analyse des decisions, Montréal. – 2005. – P. 1–18.

[7] Mustafa Kucuk, Murat Ermis “A new hybrid evolutionary algorithm for three - dimensional packing problems (2010)” IEEE International Conference, pp. 4029 – 4034.

[8] Kureichik, V.V., Kureichik, V.M. “Genetic search-based control (2001)” Automation and Remote Control, 62 (10), pp. 1698-1710.

[9] Koide S., Suzuki S., Degawa S. “A Palletize-Planning System for Multiple Kinds of Loads using GA Search and Traditional Search” / S. Koide, S. Suzuki, S. Degawa // Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. – IEEE. –1995. – Vol. 3. – P. 510–515.

[10] Chan F. T. S., Kumar N., Wong T. C. “Three-Dimensional Air-Cargo Loading Problem: An Evolutionary Algorithm Based Approach” / Proceedings of the 7th Asia Pacific Industrial Engineering and Management Systems Conference. – 2006. – P. 758-765.

[11] Gehring H., Bortfeldt A. “A genetic algorithm for solving the container loading problem” / H. Gehring, A. Bortfeldt // International Transactions in Operational Research. – 1997. – Vol. 4, Iss. 5–6. – P. 401–418.

[12] Kacprzyk, J., Kureichik, V.M., Malioukov, S.P., Kureichik, V.V., Malioukov, A.S. “Experimental investigation of algorithms developed (2009).” Studies in Computational Intelligence, 212, pp. 211-223+227-236.

Analysis of bioclimatic structure of animals' habitats on the base of the heat balance simulation

Vladimir Mikhailov
Leading researcher
Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences
Saint Petersburg, Russia
mwwcari@gmail,com

Konstantin Klokov
Professor
Dept. Geography
Saint Petersburg State University
Saint Petersburg, Russia
k.b.klokov@mail.ru

Anna Pestereva
Programmer
Aerial Services, Inc.,
Cedar Falls, IA USA.
antrin@yandex.ru

***Abstract*— The article covers questions of bioclimatic habitat analysis with outlining of thermoneutral zones favorable to the existence of animals. The technique of determining the structure and boundaries thermoneutral zones using the animal heat balance model and GIS technologies. We present examples of bioclimatic distribution over reindeer habitat territory and pastures in northern Central Siberia.**

***Index Terms*— bioclimatic distribution, thermoneutral zone, the thermal resistance, heat balance simulation, GIS-technology**

I. Introduction

The bioclimatic approach relies on the basic assumption that any living species can live comfortably in the climatic conditions in which the thermal balance with the environment is maintained thanks to the energy produced in the process of "normal" life activities (standard metabolism energy, digestion, physical activity) when the physiological thermoregulation system is at work [1, 2]. The habitat zones which satisfy these conditions are called thermoneutral. Within the boundaries of the thermoneutral zone there is no extra energy consumption or no limitations for the energy received from food which serves exclusively to maintain the organism's thermal balance. It is extremely important both from the scientific and practical viewpoints to know the boundaries of climatically comfortable zones. It allows one to determine migration routes and season locations of wild animals, as well as the preferred pastures of domestic animals.

II. Material and Methods

The different regions of a thermoneutral zone have different weather and climate conditions to which animals' bodies also react differently. Near the lower boundary of the zone the heat conductivity of fur and skin tissues are minimal, as well as the heat losses due to respiration and perspiration. However, a small fall of the ambient air temperature might lead to heat balance failure and hypothermia. The situation is totally different near the upper boundary – fur and skin tissues heat conductivity and perspiration are nearly maximum, the respiration system starts functioning in the polypnoea mode which provide the highest possible heat rejection mode. Within the boundaries of the thermoneutral zone the heat loss level is relatively constant and equals the "normal" heat production of the body, while the thermal flows and the thermal balance of the animal's body are kept constant due to the changes in the thermal resistance of fur, skin tissues and the work of other physiological thermoregulation mechanisms.

On the whole, the condition of the physiological thermoregulation system can be assessed by the value *R o*f the generalized thermal resistance of the animal's body. The value of this resistance is determined as:

$$R=(tc-tb)/Q$$

where *tc* is rectal temperature, *tb* is ambient air temperature and *Q* is total heat loss of the body.

At the lower boundary of the thermoneutral zone the resistance reaches its maximum, limited by the potential of the mechanisms of the physiological thermoregulation system with the actual values of meteo- and actinometric factors, while at the upper boundary the resistance is minimal. Let's take the normalized value of the generalized thermal resistance *K* as the stress factor of weather and climate conditions for reindeer (*R*-index):

$$K=(R-Rmin)/(Rmax-Rmin).$$

The minimum and the maximum resistance values for normalizing must be calculated using the actual values of weather and climate factors. Within the thermoneutral zone $0<K<1$, at the upper zone boundary $K=0$, at the lower - $K=1$. The deviation of the stress value from ceiling values will characterize the stability margin of the animal's thermoregulation system for overheating or hypothermia in the current weather and climate conditions.

Beyond the thermoneutral zone boundary the animal's body thermal balance is regulated by changing heat production with the fixed values of generalized thermal resistance, which have achieved their critical level. The stress factor as we knew it is of no use here. Therefore we will estimate the stress of weather and climate conditions beyond the thermoneutral zone boundary using the value of fictitious generalized resistance *Rf,* calculated when there are no limitations of the thermoregulation physiological system according to the following formula with *P* denoting the "normal" heat production of the organism:

$$Rf=(tc-tb)/P.$$

Normalizing is performed using true threshold values of thermal resistance.

$$K=(Rf-Rmin)/(Rmax-Rmin).$$

When the organism is overheated, stress factor $K<0$, in case of hypothermia $K>1$. It corresponds to such relative increase or decrease of threshold values of thermal resistance which would provide the thermal balance recovery with the "normal" value of heat production. The higher relative to one and the lower relative to zero the stress factor is, the more severe weather and climate conditions are, the faster will hypothermia and overheating of the organism take place and the more energy will the animal have to spend in order to recover its thermal balance. This energy corresponds to extra energy expenditure which is to compensate heat losses during hypothermia and the energy deducted from the "receipts section" of the organism energy budget in order to decrease heat production and eliminate overheating. Energy disbalance causes the deterioration of defense and reproductive abilities of animals. As a result, they cannot survive in such conditions for a long time.

There are a lot of experimental data concerning the influence of environment on the thermal state of animals and humans. For example, A.Ya. Sokolov and A.V. Kushnir's monograph [2] contains a list of about 500 references on the issues of bioenergetics and thermal regulation only of reindeer. However, it is impossible to conduct real experiments for arbitrary values of environmental factors, the condition of the animals' organisms and their behavior patterns. The problem could be solved by substituting real experiments with computer simulations with the corresponding animal model. To do it, we developed a special model of the animal's thermal balance, which provides us with the data required for estimating the stress of weather and climate conditions and analysis of the bioclimatic structure of the habitat. The model is different from the known models of such type [3, 4] and our earlier versions [5, 6] because the concept of thermoneutrality was realized in it - relative stability of the animal's body heat losses in the wide range of changing weather and climate factors due to the work of the physiological thermoregulation system.

The main principles of the developed model are based on the following empirical biological generalizations:

- the animal's body is considered to consist of "the core" with the relatively constant temperature and "the shell" with the changing temperature;
- energy expenditure on the work of the physiological thermoregulation system (changing the shell and the fur heat conductivity, adaptive changes of the respiratory system, perspiration) are negligibly small;
- the thermoregulation system reacts to the disbalance of heat production and heat losses;
- all mechanisms of the physiological thermoregulation system work simultaneously.

The active system of regulating heat flows is realized in the model. However, the external factors influence the thermoregulation system indirectly by changing the total heat in the animal's body. The thermoregulation mechanisms include:

- indumentums piloerection,
- change of the shell tissues thermal resistance,
- change of heat losses with breathing,
- heat production change (grazing decrease in hot weather which leads to less energy consumption for specific dynamic food action and reindeer biomass increase).

The input data used for calculations are: air temperature, wind speed, cloudiness, humidity, depth and thickness of snow, as well as the animal's weight and the type of its activity. See [7-9] for the detailed description of the model.

Three tasks have to be solved in order to build the bioclimatic field of the habitat using the thermal balance model. The first one is collecting the body of input data for making calculations in the model. The body contains the time-series of necessary actinometric and meteorological data: air temperature, wind velocity, cloud amount, depth and thickness of snow, direct and scattered solar radiation and sun's altitude. These data may be provided by the network of meteorology stations in the area, weather satellites and meteorological databases. The problem consists in selecting the data with the required time averaging, their adjustment and formatting. The second task is conducting calculations on the thermal balance model using the collected body of data in order to estimate the stress of animal thermoregulation system performance in certain localities of the habitat. The third task is to build bioclimatic fields using GIS technologies and to determine the boundaries of thermoneutral zones [10].

III. Results and Discussion

As the preliminary calculation showed the most dangerous periods for reindeer are summer (July-August) when overheating is possible and winter (December-February) when hypothermia might occur. In the frost-free season the limiting group includes the most bulky reindeer (strong male) with minimal heat loss. The heat loss was maximal in the "rest in the standing position." In winter the limiting group consists of calves – animals with little weight and maximum relative heat loss.

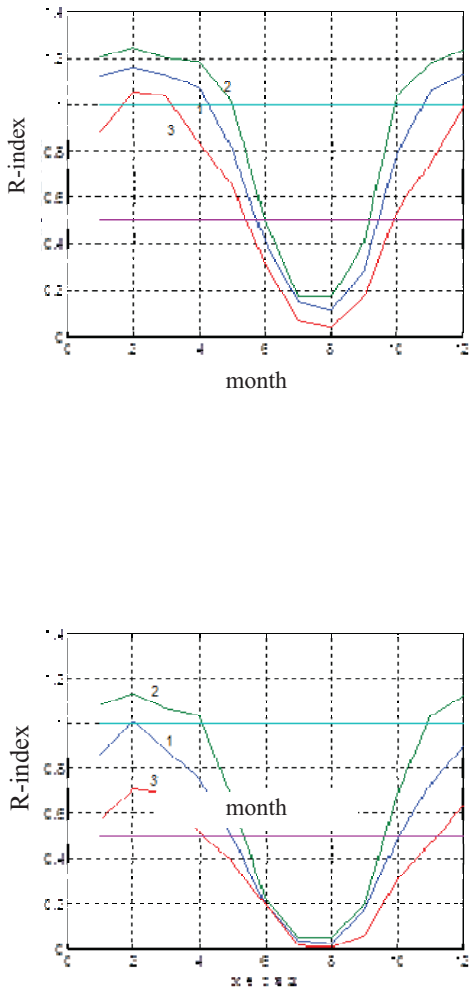


Fig. 1. R-index for calves. Rest in a standing position (upper), rest in a couching position (lower). 1-norml year, 2-cold year, 3-warm year, Dixon district (Taimyr). Vertical scale – R-index, horizontal scale – months (1 – January).

The heat loss of the body is minimal in the "rest in a couching position". In May-June (the period of calving of wild and domestic reindeer) and in September – October (autumn grapier and preparation for rutting season) practically the whole territory of Northern Russia is a favorable thermoneutral zone for reindeer (Fig. 1).

The bioclimatic calculations method was used for analyzing wild reindeer habitats in Taimyr and domestic reindeer breeding in the North of the Russian Federation [9, 10].

To analyze the habitats of the Taimyr reindeer population we used the data from 25 meteorology stations located on this territory for 1986 - the year with "average" weather and climate factors in the interval 1970-1990. The stress fields of weather and climate conditions of the area were constructed by interpolating the values connected with the location of meteorology stations. The interpolation was performed using the algorithm of inversely weighted distances. When the temperature decreases the boundary of the zone shifts to the Dixon - Ust-Tareya. Let's consider some analysis results. Figure 2 shows the stress fields and the location of the boundary of the overheating zones for reindeer calves in July in

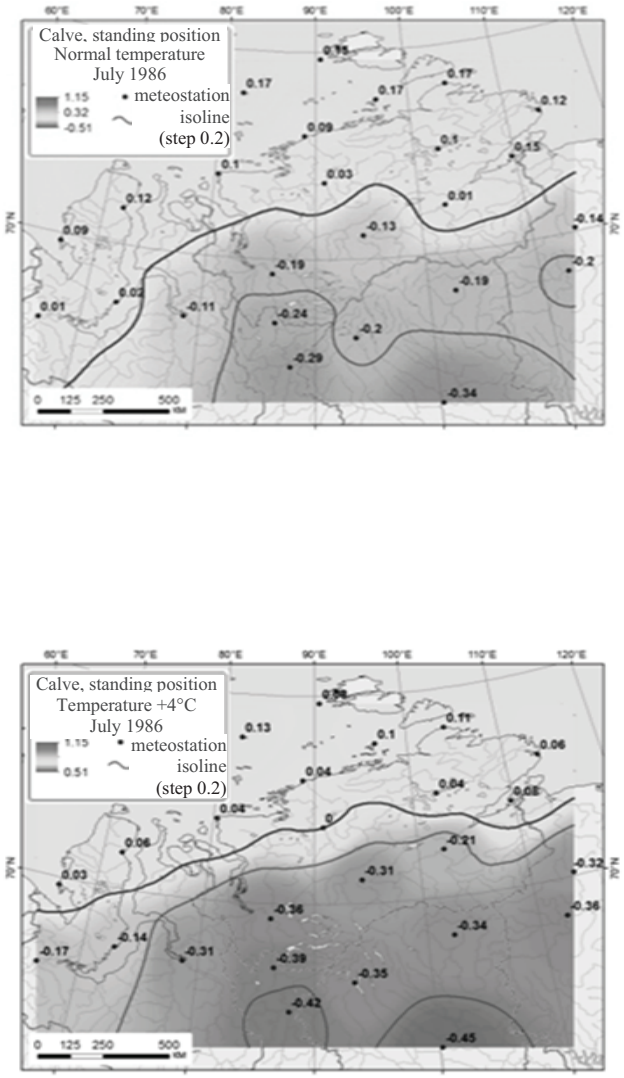


Fig. 2. The stress fields and the location of the boundary of the overheating zones (bold) for reindeer calves. July, 1986, (upper). Weather condition as July, 1986, but ambient air temperature was increased by 4°C (lower).

normal weather conditions of 1986 (upper) and when the ambient air temperature was increased by 4°C (lower). Similar data are provided for male (Fig. 3). As can be seen from the figures, the increase of the average monthly ambient air

temperature causes the significant shift northwards of the regions of comfortable living for reindeer, and this is especially true for male species.

In July-August large overheating areas in the region appear. In Taimyr and in the Evenk country these include the regions of Dudinka, Volochanka, Khatanga and the whole territory further to the South. In August the overheating zone shifts southwards to the regions of the Yessey, the Tura, the Agata and the Igarka. The obtained results agree perfectly with the biological data about the seasonal location of Taimyr wild reindeer groups.

Traditional reindeer herding is the basis of the way of living of the indigenous population in many regions of Northern

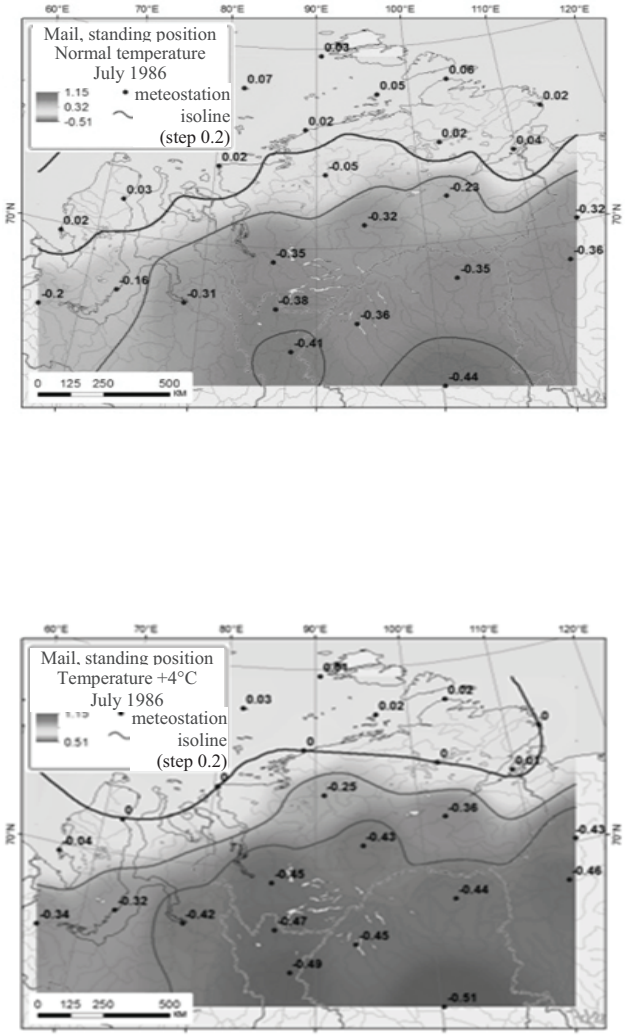


Fig. 3. The stress fields and the location of the boundary of the overheating zones (bold) for reindeer male. July, 1986 (up). Wheather condition as July, 1986, but ambient air temperature was increased by 4ºC (lower).

Russia. The essential physical factor, determining the perspectives of reindeer breeding development is the fodder resources of the pastures. However, this point of view contradicts the newly obtained data. For example, the last several decades have seen the growth of reindeer population in the tundra of the Yamal-Nenets Autonomous District (YNAO) which is the region with the strong lack of fodder resources. At the same time in many other regions where the pasture resources are not fully used, the reindeer population is decreasing or remains quite small despite the financial support provided by the government [11, 12].

Using the objective data we have to assess which regions are more suitable for reindeer herding, and which ones – less suitable. This information is essential for planning the development of this economic sector and taking decisions concerning its state support as it forms the traditional basis of the way of living of indigenous groups.

Using the thermal balance model based on the data from 25 meteorology stations, we analyzed how the territories of the most intensive summer reindeer grazing in YNAO depend on the climate [13]. The obtained results help to explain the historically developed peculiarities of the location of traditional reindeer herding.

The heat balance factor is most important in summer, therefore Yamal reindeer herders, using the limited period of spring migration, move to the North, as far as it is possible and accumulate their herds along the Western coast of Yamal where heat loss increases due to stronger winds. This strategy reduces to the minimum the threat of crossing the upper boundary of the thermoneutral zone. It seems quite possible that the above mentioned climatic peculiarities are reflected in the traditional knowledge of reindeer herders, and therefore for summer pastures they choose the most comfortable for reindeer areas from the point of view of their thermal balance, in spite of the fodder resources depletion.

The calculations made on the model also allow us to predict how possible climate change might influence reindeer herding in Yamal. If the average monthly temperature rises by 2ºC the boundaries of the thermally neutral zone will shift northwards by about 100 km. This will lead to the deterioration of reindeer herding conditions, especially in warm years. As a result the Southern part of Yamal might become the zone of "risky reindeer herding."

Besides, the thermal balance model was used for the approximate assessment of the suitability of the climate of different regions of Russia for domestic reindeer herding. The calculations showed that in the regions which are perfect from the point of view of ecological and economic conditions of their location (for example in the tundra of YNAO) it is possible to have profitable and cost-effective reindeer herding. In the other regions with less comfortable temperature regime for reindeer labor costs of reindeer herding are much higher.

Reindeer herding in these regions brings no appreciable profits, but is supported by regional governments in order to give employment to indigenous population.

IV. Conclusion

The computer simulation results of the reindeer thermal balance allowed to point out the areas of optimal and pessimal climatic conditions for wild reindeer populations and domestic reindeer herding. The model, presented in this work, is adjusted using the data for reindeer. However, with the corresponding modification such an approach can be used for studying the influence of weather and climate factors on other wild and domestic animal species.

Acknowledgment

This article is based upon researches supported by Russian Foundation for Basic Research (projects № 15-06-041-95 and № 16-08-00510).

References

[1] K. P. Ivanov. "Bioenergy and temperature homeostasis", (in Russian), Leningrad, Nauka, 1972,

[2] A. Ya. Sokolov, A.R. Kushnir, "Thermoregulation and bioenergy of reindeer", (in Russian), Novosibirsk, 1997.

[3] B. A. Aizenstatm, "Method of calculation of radiation and heat balance of animals" (in Russian), Biometeorology, Vol. 20 (101), pp. 27-48, 1974.

[4] R. E. White, D. E. Russell, C. J. Daniel, "Simulation of maintenance, growth and reproduction of caribou and reindeer as influenced by ecological aspects of nutrition, climate change, and industrial development using an energy-protein model", Rangifer, Special Issue, No. 22, 2014.

[5] V. V. Mikhailov, V. Yu. Mordovin, "Animals' heat losses model and climate", (in Russian), Proceedings of SPIIRAS, St. Petersburg, Nauka, Issue 2, Vol. 2, p. 407 – 417, 2005,

[6] V. Yu. Mordovin, V. V. Mikhailov, L. A. Kolpashchikov, "Bioclimatic model of the range of wild reindeer (guidelines for the construction and practical use)", (in Russian), Research Institute of agriculture of Extreme North, 2005.

[7] V. V. Mikhailov, "Simulation of Animal's Heat Balance". Trans. of IV Int. Conf. Problems of Cybernetics and Informatics (PCI'2012), Baku, p 47-63, 2012.

[8] V. V. Mikhailov, "Model of regulation of the heat balance of the reindeer, taking into account seasonal changes in radiation and meteorological factors", (in Russian), Proceedings of SPIIRAS, vol. 13, p. 255-276, 2013.

[9] V M. Makeev, K. B. Klokov, L A Kolpashchikov, V. V. Mikhailov, "Reindeer in the face of climate change", (in Russian.), SPb.: Lemma, pp. 244, 2014.

[10] V. V. Mikhailov, A V. Pesterev, "Zoo-climate monitoring based on heat balance model and GIS technology", (in Russian.), Proceedings SPIIRAN, Vol 13, pp. 276-291, 2013.

[11] B. K. Klokov, "Fuctuations in national and regional variations in the number of domestic reindeer in the Russian North: some possible explanations, ordinary", Vol. 10, Issue 1, UK, pp 23-47, 2011.

[12] B. K. Klokov, "Changes in the reindeer population in Russia: impact of political environment or climate?" Northern, 32 (1), Norway, pp. 19-33, 2012.

[13] .K. B. Klokov, V. V. Mikhailov, "Identify locations of climatic optimum for traditional reindeer herding of the indigenous peoples of the Yamalo-Nenets Autonomous district", (in Russian), Proceedings of the state agrarian University, 2015, No. 40, pp. 105-108.

Usage of fuzzy extractors in a handwritten-signature based technology of protecting a hybrid document management system

Pavel Lozhnikov, Alexey Sulavko, Danil Volkov
Omsk State Technical University
Omsk, Russia
lozhnikov@mail.ru, sulavich@mail.ru, volkovdanil91@gmail.com

***Abstract*—The paper covers the issue of protecting private keys of a digital signature in a hybrid document management system. It proposes a method of generating private keys based on a handwritten signature using a method of building a reference signature description. The method demonstrates the following indicators of reliability: FRR=0.148, FAR=0.05 when the length of the key is 304 bits.**

Index Terms*—*a hybrid document, authentication, encryption key generation, user's signature, biometrics.

I. Introduction

Nowadays many documents are created using IT. But in many countries, in Russia in particular, the total transition to the "paperless technologies" will not occur in the nearest future. Legislative requirements and government regulations state the most important business documents to be drawn up in paper only. Besides, many managers prefer to work with paper documents to make legally valid solutions. Therefore, we can tell about the transit to a compound or hybrid document management system. An attribute of a hybrid document is a digital signature that is built using biometrical data, a hash-function of a document and a handwritten signature [1]. The limitation of a typical digital signature is its ability to be alienated from the owner. Another limitation appears if a document has to be translated from digital to a paper form, the standard digital signature becomes useless in this case.

A problem of providing authenticity of open keys is solved by means of the certification authority [2]. A common practice of protecting private keys is the usage of data storage equipments or secure data banks [3]. In this direction new methods of protecting temporary memory and communication line during accessing the storage equipment or a data bank are being developed.

There are known ways of protecting encryption keys and passwords by means of biometric data (a fingerprint, a voice, a face, an iris) [4]. The paper [5] suggests an algorithm of protecting a key and biometric data themselves by means of an owner's fingerprint using the error-control coding to correct errors derived from the essential impossibility to obtain a pair of completely identical images of a fingerprint in the process of another scanning. Another method of protecting private keys are known as well, the majority of them are based on using an owner's fingerprint [6-7] and a fuzzy extractor method. According to this method a vector of biometric parameters and some additional information – an open string that is dependent on reference biometric characteristics but is not a template itself are necessary to generate the key [8]. However, there is no efficient methods of restoring templates from the open string. There are known methods of protecting a private key via "fuzzy storage" [9-10]. The difference between the fuzzy storage and a fuzzy extractor is in the fact that in the first case an open string is a random set, in the second case it is a finite-dimensional vector.

The attributes stated in papers [4-7] are visible therefore their biometric images may be stolen and forged. In comparison with static biometric attributes the dynamic characteristics of the manner of handwriting a signature are much more difficult to capture, that is a signature privilege. A signature may be replaced with a secret handwritten password. But the stability of obtaining the signature characteristics is significantly lower.

This paper proposes a technology of protecting a hybrid document management system that differs from others by using information about the parameters of the signature handwriting manner to generate a private digital key, and by creating a special analog of a digital signature in a form of a barcode. The paper covers the first stage of this technology – the private key generation based on a signature. A method of fuzzy extractors is proposed to use to bind a traditional handwritten signature with a private key [8].

Several approaches are known to aim at the increase of the reliability of the key generation and the signature-based authentication: the search of new information-bearing attributes, the development of more efficient algorithms of decision-making and comparing a presented sample with a template, the consideration of the signer's psychophysiological state while creating a signature template [11]. This paper suggests decreasing the probability of errors of key generation by finding and excluding incorrect signature samples in a moment of open string generation and by using the most stable biometric attributes only.

II. Past achievements

The reliability of the biometric protection for a key is defined by the probability (or a percentage) of the mismatch of a key generated from attributes of the same user (FRR – False Rejection Rate) and the probability (or a percentage) for the mismatch of a key generated from attributes of different users (FAR – False Accept Rate). When FRR=FAR a coefficient of equal probability (EER) is used. The following results of generating keys via signatures are known. The paper [12] describes the research of signatures belonging to 126 test persons (the research was supported by the Spanish Ministry for Science and Technology under the project MCYT TIC2003-08382-C05-01 and the European Commission for Science and Technology under the IST-2002-507634 Biosecure NoE project) where the following result is achieved: FRR=57.30% along with FAR=1.18% for professionally done forgeries, and with FAR=0.32% for forgeries made without monitoring the process of handwriting the original signature. The paper [13] provides the following result: EER≈6.7% for 40 test persons. The paper [14] demonstrates EER≈9% when the length of the key is 100 bits. Another result is FRR=28% along with FAR=1.2% presented in the paper [15].

Let us consider several results of generating a key using static biometric attributes. It should be noted that the number of these studies is much larger. For methods using a fingerprint there are: FRR=0.9 along with FAR≈0, a key length is 256 bits [16]; FRR=12.6% along with FAR≈0%, a key length is 327 bits [17]. For methods using an iris there are: FRR=0.47% along with FAR≈0%, a key length is 140 bits [18]; FRR=4.64 along with FAR≈0, a key length is 128 bits [19]. As is seen from above key generators based on static biometric attributes significantly exceed systems of key generation based on signature attributes. Nowadays reaching the same or at least comparative indicators for signature attributes is a subject of studies. Their relevance is determined by the advantages that a traditional signature used as an authenticator provides.

III. Space of signature attributes

A digital tablet or a tablet computer and a stylus may be used to input signatures. A signature is characterized by a stylus position function $x(t)$, $y(t)$ and a stylus pressure function $p(t)$; they contain information about the dynamics of stylus movement and a signature image, where t is the time. It is necessary to find attributes that means the values describing the owner. The attribute values obtained from signatures of the same owner are similar (the values are close) and obtained from signatures of different owners should be different (values significantly differ).

Signature samples differ in the duration (a number of readings). Firstly, samples should be normalized by the equal duration. For this purpose, the functions are passed through the direct Fourier transform and inverse Fourier transform with the same duration, this procedure is called resampling, and the duration is equal an average duration of scaled signals [11]. The normalization consists of the following steps:

- The initial and final values of all dots with a zero pressure are discarded.
- The functions $x(t)$, $y(t)$ and $p(t)$ are passed through the one-dimensional Fourier transform.
- This functions are passed through the inverse Fourier transform, the output dimension should match with the number that is the nearest lowest integer multiple of the second power.

Different approaches to defining attributes are reported, they describe both static signature images and dynamic characteristics of the handwriting manner. The present paper uses all approaches and selects the best number and a combination of attributes in a process of a calculating experiment.

One of the approaches to the making the attribute space is based on building a matrix of distances between readings of the signature. The process of calculating the matrix of distances may be divided into the following stages:

- A calculation of the step $s=N/d$, where N is a number of dots resulted from the inverse Fourier transform, d is a desirable dimension of the matrix multiple to the power of 2. The calculation results in many elements, and it requires significant calculation resources that is why calculations should be done with the step s.
- A calculation of the matrix of distances R in 3-dimensional space (pressure is the 3rd dimension) for the complete set of coordinates with the step s. Distances between the i^{th} and the j^{th} coordinates that may be calculated using the formula 1 are the elements of the matrix.

$$r_{ij} = \sqrt{[(x_i-x_j)^2+(y_i-y_j)^2+(p_i-p_j)^2]}. \quad (1)$$

- Normalization of the matrix by the signature length:

$$r'_{ij} = r_{ij}/(r_{12}+r_{23}+\ldots+r_{(N-1)N}).$$

Elements of this matrix are biometric attributes. The matrix of distances in 2-dimensional space (p is not taken into account) may be calculated in a similar way.

The next approach covers attributes that characterize the signature image:

- A signature length-to-width ratio.
- A center of the signature described with the coordinates C_x, C_y and C_p.
- An angle of slope for the signature calculated by the formula 2. An angle of slope is a cosine of a mean spiral angle for a signature polygonal path to the X axis:

$$\theta = (1/N-1)\Sigma_{i=1}^{N}((x_{i+1} - x_i)/\sqrt{[(x_{i+1} - x_i)^2+(y_{i+1} - y_i)^2]}). \quad (2)$$

- An angle of slope between centers of the signature halves. After the center of the signature C_x has been found let us divide the set $(X,Y,Z)=\{(x_i,y_i,p_i)\}$ into two subsets $L=\{(x_i,y_i,p_i)|x_i \le C_X\}$. Let us find the coordinate of the left half C_{xL} of the signature using the formula 3. Other coordinates of the corresponding halves of the signature are calculated in the similar way.

$$C_{xL} = (1/|L|)\Sigma_{xi \in L}^{N}(x_i). \quad (3)$$

The next approach is a simple way to measure the dynamic attributes of the signature based on Fourier transform. The functions of pressure and velocity of a stylus on a tablet [20] are expanded into Fourier series. If the function of velocity for a stylus on a tablet surface is used, then the dependence between the angle of a tablet and a position of a user's hand disappears. Further the amplitudes of the obtained harmonics are normalized by the energy of the functions (values of the amplitudes are divided into a sum of squared amplitudes of all harmonics for the corresponding signal).

Another approach to the calculation of attributes is based on the correlative analysis of the signature functions *x(t)*, *y(t)* and *p(t)* [11, 20]. Values of the correlation coefficients between the signature functions *x(t)*, *y(t)* and *p(t),* and their first-order derivatives are used as attributes.

IV. The Proposing Method of Key Generation

To make an open string a user enters several samples of his signature using a special device (a digital tablet or a tablet computer that supports pressure force measurement). Any signature is converted into a vector of values of the attributes mentioned above. Further mean values of sample attributes are calculated. Any sample attribute value is converted into a 8-bit vector. A mapping of a range of possible attribute values onto a set {0, 1, 3, 7, 15, 31, 63, 127, 255, 254, 252, 248, 240, 224, 192, 128} is used, the elements of the set are presented in a binary form. The sequence of bits is set in a random manner for every user individually. Further the elements of the bit vectors are concatenated into a final bit sequence. A private key is encrypted with the Hadamard error-correcting code. The advantage of the Hadamard codes is a high speed of processing.

Characterizing by a large code distance the Hadamard codes allow correcting a large number of error as well. It is achieved through high redundancy. Sometimes it is worth it. In order to apply the Hadamard coding it is necessary to build the the Hadamard matrix of the corresponding order. For any integer $n>0$ a squared matrix $H=h_{ij}$ of the n^{th} order is called the Hadamard matrix if $h_{ij}\in\{+1; -1\}\forall i,j$ и $HH^t=nI$ where I is a unity matrix. Let us create a code using all lines of the matrix I and their complementary relations. This code (or its {0,1} correspondence) is called the Hadamard code. Let us select code words starting with c+1 in the Hadamard code with the length of n, discard the first coordinates and translate into the {0,1}-code. The resulting equidistant binary code with the length of n-1, the dimension of n and the distance of n/2 is the shortened Hadamard code. A fuzzy extractor based on the Hadamard code will encrypt a bit representation of a random string (a private key) taking a block size as a parameter. The error-correcting ability of the code depends on the block size; the best block size may be found in a process of a calculating experiment. The length of the generating key is the biggest integer of the difference $8M/b$, where M is a number of attributes, b is a block size. The encrypted sequence is combined with the obtained bit sequence using the XOR operation. The result of the combination is an open string. To obtain a private key a user inputs a new signature sample that is processed in a corresponding manner and "subtracted" of the open string to "detach" biometric data. Further the error-correcting code is applied to this string, and if the presented biometric image matches the template image to a high degree, the private key will be obtained.

This approach does not require storing a template of a user's signature, only an open string should be stored. As the sequence of attributes for any user differs, the order of attributes in the presenting vector will not match if the biometric data are detached of another user's open string. Therefore, the probability of an error is significantly lower, at this time the probability of an error of the private key mismatch for 2 signature samples of the same user (FRR) is permanent for sure (a sequence of attributes in the "Self" image is the same). The sequence of attributes may be stored in a certain storage device or a dedicated server. This information is worth being kept in secret. If it is disclosed the probability of the error of the II type increases.

V. Discard of an Incorrect Sample During Generating an Open String

A personal signature is a dynamic biometric image that changes throughout life. The changes may be natural, the reason is the changes in the manner of handwriting in the course of life. Nowadays methods of updating a template in a case of successful authentication are used to take into account these variations. It is recommended to create a new template (a period is defined based on the character of these changes) from time to time. But unexpected shifts may be caused by the inaccuracy of the equipment and the signature input. Users often makes mistakes while inputting a signature, the reasons may be different: a signer's hand jerked, a user diverted attention away, etc. As a result, many attributes derived from these incorrect signatures have values that are not typical for this user. Therefore, it is important to consider only correct signature samples to build template attribute values.

The simplest way would be the evaluation of the correlation relationships of the signatures after the normalization by the duration. But the initial functions *x(t)*, *y(t),* and their derivates (velocity functions of the x and y coordinates) depend on the signature angle of slope. The signer's hand jerks are captured by the velocity and the pen pressure functions as a rule that leads to the decrease in a number of correlation relations between the corresponding functions of incorrect and correct signatures of the same user. But experiments show that some test persons have low values of correlation coefficients between correct samples of these functions (below 0.5, see Table 1). Some test persons have signatures that do not possess salient values of the pressure function (see Table 2).

TABLE I. LOW CORRELATION DEPENDENCE OF THE VELOCITY FUNCTION FOR A STYLUS ON A TABLET

	Test person 1					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	0.1958	-0.14	0.2321	0.0445	0.0263
2	0.1958	1	0.3052	0.1644	0.2662	-0.323
3	-0.140	0.3052	1	0.3611	0.5941	0.1769
4	0.2321	0.1644	0.3611	1	-0.068	0.2878
5	0.0445	0.2662	0.5941	-0.068	1	0.1843
6	0.0263	-0.323	0.1769	0.2878	0.1843	1
	Test person 2					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	-0.03	-0.373	-0.167	-0.036	-0.3881
2	-0.03	1	0.2261	0.2339	0.2246	0.1095
3	-0.373	0.2261	1	0.4161	0.0241	0.1681
4	-0.167	0.2339	0.4161	1	0.7196	0.62
5	-0.036	0.2246	0.0241	0.7196	1	0.3798
6	-0.388	0.1095	0.1681	0.62	0.3798	1

TABLE II. LOW CORRELATION DEPENDENCE OF THE PRESSURE FUNCTION FOR A STYLUS ON A TABLET

	Test person 1					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	0.1757	0.2098	0.3311	0.0065	0.6785
2	0.1757	1	0.5028	0.4952	0.299	0.5286
3	0.2098	0.5028	1	0.0928	0.4808	0.3574
4	0.3311	0.4952	0.0928	1	0.1816	0.6677
5	0.0065	0.299	0.4808	0.1816	1	-0.075
6	0.6785	0.5286	0.3574	0.6677	-0.075	1
	Test person 2					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	0.5053	0.0101	0.4088	0.278	0.1633
2	0.5053	1	0.1638	0.7963	0.8215	0.3897
3	0.0101	0.1638	1	0.6181	0.4945	0.7722
4	0.4088	0.7963	0.6181	1	0.9384	0.7229
5	0.278	0.8215	0.4945	0.9384	1	0.7576
6	0.1633	0.3897	0.7722	0.7229	0.7576	1

Therefore, an estimation of the correlation relations is proposed to be made between the vectors of the attribute values derived from the initial signature samples to search incorrect samples. These correlation coefficients are more steady values for all test persons (see Table 3).

TABLE III. CORRELATION DEPENDENCE FOR THE VECTORS OF THE SIGNATURE ATTRIBUTE VALUES

	Values of the elements q_{ij} for the test person 1					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	0.8319	0.6850	0.8222	0.5331	0.6465
2	0.8319	1	0.7592	0.8023	0.5028	0.6491
3	0.685	0.7592	1	0.6768	0.6114	0.5291
4	0.8222	0.8023	0.6768	1	0.5029	0.7802
5	0.5331	0.5028	0.6114	0.5029	1	0.518
6	0.6465	0.6491	0.5291	0.7802	0.518	1
	Values of the elements q_{ij} for the test person 2					
No of a signature sample	***1***	***2***	***3***	***4***	***5***	***6***
1	1	.,4637	0.4048	0.5066	0.5451	0.4091
2	0.4637	1	0.5985	0.884	0.9278	0.6636
3	0.4048	0.5985	1	0.7344	0.6813	0.8983
4	0.5066	0.884	0.7344	1	0.9556	0.7148
5	0.5451	0.9278	0.6813	0.9556	1	0.7508
6	0.4091	0.6636	0.8983	0.7148	0.7508	1

The proposing method of searching incorrect samples while generating an open string is realized after the translation of the entered user's signatures into the vectors of attribute values and before the calculation of the mean (master) values of the attributes, and is as follows:

- A matrix of correlation coefficients q_{ij} (Formula 4) for the vector of attribute values for k entered signature samples (see Table 3) is built.
- A multiple correlation coefficient $c_{i;1,2,..,i-1,..,i+1,..,k}$ for every i^{th} attribute with other vectors of attribute values derived from the previous step is calculated. Thus, an estimation of the correlation dependence for every signature sample and other entered signatures. As is known, a coefficient of multiple correlation may be calculated by the formula 5.
- The calculating coefficients of correlation are compared with some threshold value. If the obtained coefficient is less than this value, the signature corresponding to this coefficient is considered to be incorrect and is not taken into account for generating an open string. Instead, a user is offered to input another signature, and the procedure is repeated for the next signature.

$$q_{ij} = \Sigma_{v=1}^{M}[(x_{iv}-\mu_i)(x_{jv}-\mu_j)]/\sqrt{[(x_{iv}-\mu_i)^2]}\sqrt{[(x_{jv}-\mu_j)^2]}, \qquad (4)$$

where M is a number of attributes, x_{iv} is a value of the i^{th} attribute derived form the v^{th} signature sample, μ_i is an expectation (a mean value) of the i^{th} attribute

$$c_{i;1,2,..,i-1,..,i+1,..,k} = \sqrt{[1-(|Q_k|/Q_{ii})]}, \quad (5)$$

where $|Q_k|$ is a determinant of the matrix of the correlation coefficients Q_k for k vectors of the attribute values, Q_{ij} is an algebraic complement of the element q_{ii} of the matrix Q_k.

It is empirically stated that the optimal threshold value in the space of attributes under examination may slightly vary for every user, but its value ranges within the interval [0,5; 0,7].

VI. Definition of the Optimal Number of Attributes and an Estimation of the Efficiency of the Suggested Method

To perform a calculation experiment 65 test persons were selected. Any user entered 50 signatures minimum using WACOM tablets of the latest models. Every signature was translated into a bit sequence by the method described above. A proportional frequency of occurrence for on-bits in a bit presentation of the attribute values was calculated for all entered samples. An integral estimation of permanency of bits for a converted attribute value as a multiplication of relative frequencies of occurrence of on and zero bits in all digits (elements of a bit vector) is performed. The relative frequencies of 0 and 1 are replaced with the nearest figures (for a product not to be equal 0).

The more digits the frequencies close to 0 or 1 have, the less the final product and the higher an integral estimation of the attribute permanency. Further, a certain number of the most permanent attributes is selected, the others are discarded. The optimal number of attributes and a block size for the Hadamard code is defined based on the calculation experiment.

In the course of the calculation experiment a half of samples entered by the users were used to generate open strings. The other samples were used to generate keys. The experiment was repeated several times, moreover the learning selection and the experimental selection of signatures changed in a random manner. The major results of the experiment are presented at the figures 1-2. The test validity is more than 0.99 under the confidence interval equal 0.03. As the figures show the optimal block size is 6 bits. Under the smaller block size, the error probability is high if the key length is smaller. If the size is larger, the length of the key is larger, but the probability of the error of generation is too significant.

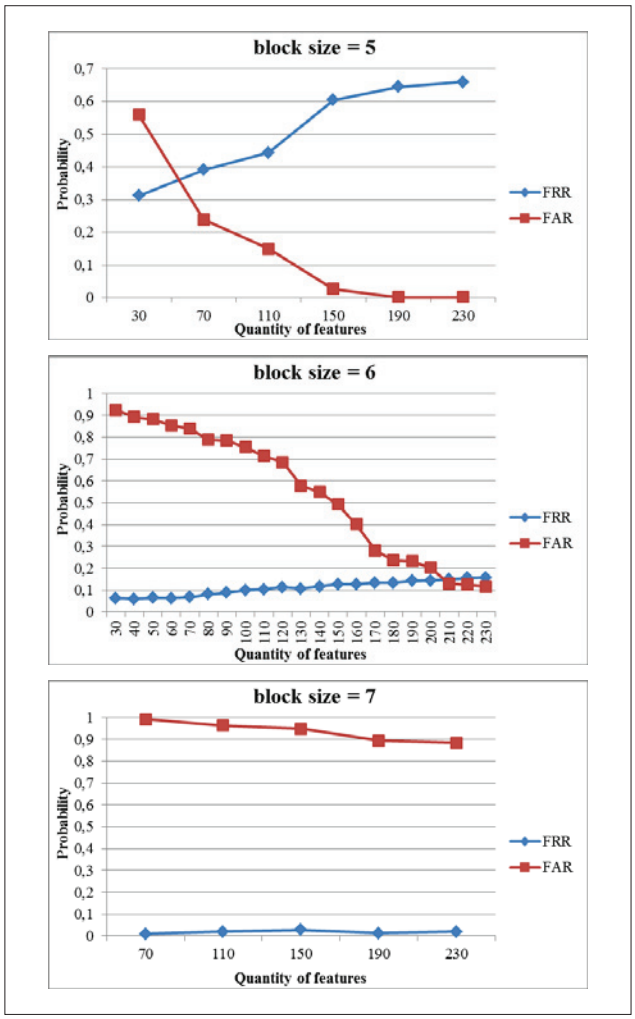


Fig. 1. Probabilities for errors of key generation without using a method of discarding incorrect samples along with different block sizes of the Hadarard code

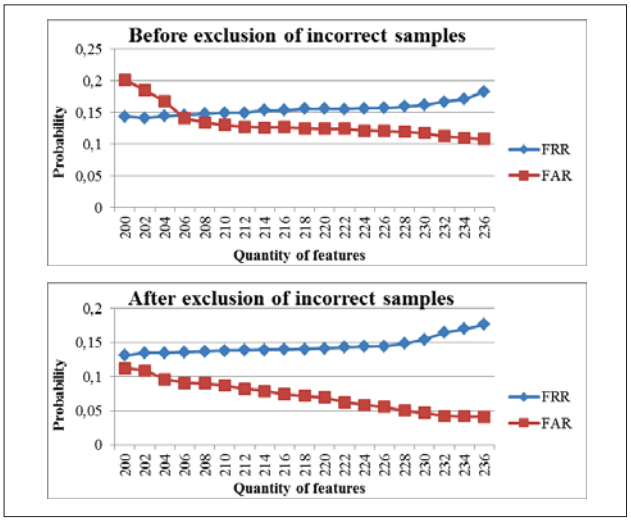


Fig. 2. Probabilities for errors of key generation before and after using the method of discarding incorrect samples when the block size of the Hadamard code equal 6 bits (explicitation of the result)

VII. Conclusion

Based on the results the optimal number of attributes from all reviewed in these paper is 228. The method of discarding incorrect samples decreases a number of errors of the type I (FRR) by 6.5% on average, and of the type II (FAR) by 46%. The best result in generating keys based on the Hadamard code in a given space of attributes is the following: FRR=0.148, FAR=0.05 along with the length of key equal 304 bits. The study validity is more that 0.99 along with the confidence interval equal 0.03. The result was obtained when incorrect samples were discarded during generating an open string and the size of the block of the encrypting message was 6 bits. The number of errors of the I and the II type was minimal when these parameters were used. The study demonstrated that the advantage of the extractors based on the Hadamard code in comparison with BCH (Bose-Chaudhuri-Hocquenghem) code, for example, is a high speed of generating a key.

The further studies on upgrading the methods of protecting private keys are planned to pursue in the decrease of redundancy for an open string. One of the variant of the upgrade is a partial carrying of error syndromes out of an open string that gives an opportunity to increase the key length while decreasing error probabilities. BCH codes, Reed-Solomon codes in particular, a subset of BCH codes allow this.

Acknowledgment

The reported study was partially supported by RFBR, research project No. 16-07-01204 A.

References

[1] Ivanov A.I., Lozhnikov P.S., Samotuga A.E. Technology of formation of hybrid instruments // Cybernetics and Systems Analysis, 2014, T. 50, № 6. pp. 152-156.

[2] C. K. Williams. Configuring enterprise public key infrastructures to permit integrated deployment of signature, encryption and access control systems. MILCOM 2005 - 2005 IEEE Military Communications Conference, Vol. 4, 2005, pp. 2172 – 2175.

[3] Le Guan, Jingqiang Lin, Bo Luo, Jiwu Jing, Jing Wan. Protecting Private Keys against Memory Disclosure Attacks using Hardware Transactional Memory. 2015 IEEE Symposium on Security and Privacy, 2015, pp. 3-19.

[4] Indu Verma, Sanjay Kumar Jain. Biometrics security system: A review of multimodal biometrics based techniques for generating crypto-key. Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference, 2015, pp. 1189 - 1192.

[5] Nguyen Thi Hoang Lan, Tran Quang Duc, Nguyen Tu Hoan. A Biometrics Encryption Key algorithm to protect private key in BioPKI based security system. Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference, 2009, pp. 1 – 5.

[6] Cheol-Joo Chae, Kwang-Nam Choi, Kiseok Choi, Jae-Soo Kim, YongJu Shin. Enhanced biometric encryption algorithm for private key protection in BioPKI system. Journal of Central South University, November 2014, Volume 21, Issue 11, pp. 4286-4290.

[7] Thi Hoang Lan Nguyen, Thi Thu Hang Nguyen. An approach to protect Private Key using fingerprint Biometric Encryption Key in BioPKI based security system. Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference, 2008, pp. 1595 – 1599.

[8] Dodis, Y., Reyzin, L., Smith, A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data // Proceedings from Advances in Cryptology. EuroCrypt. – 2004. – P. 79-100.

[9] Juels, A., Sudan, M. A fuzzy vault scheme // Proc. IEEE Intil. Symp. Inf. Theory. – 2002. – 408 p.

[10] Vu Hiep Dao, Quang Duc Tran, Thi Hoang Lan Nguyen. A Multibiometric Encryption Key Algorithm Using Fuzzy Vault to Protect Private Key in BioPKI Based Security System. Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference, 2010, pp. 1 – 6.

[11] Lozhnikov P.S., Sulavko A.E., Samotuga A.E. Personal Identification and the Assessment of the Psychophysiological State While Writing a Signature. Information. 2015, № 6, 454-466.

[12] Santos, M. F., Aguilar, J. F., Garcia, J. O. Cryptographic key generation using handwritten signature // Proceedings of SPIE, Orlando, Fla, USA, Apr. 2006. – 2006. - Vol. 6202. - P. 225-231.

[13] Yip, K. W., Goh A., Ling D. N. C., Jin A. T. B. Generation of replaceable cryptographic keys from dynamic handwritten signatures // In Proc. ICB, Lecture Notes in Computer Science 3832, Springer. – 2006. - P. 509–515.

[14] Maiorana, E., Campisi, P. Fuzzy commitment for function based signature template protection // IEEE Signal Processing Letters. – 2010. - № 17. – P. 249–252.

[15] Hao, F., Chan, C. W. Private key generation from on-line handwritten signatures // Information Management & Computer Security, Issue 10. – 2002. – №. 2. – P. 159-164.

[16] Teoh, A., Kim, J. Secure biometric template protection in fuzzy commitment scheme // IEICE Electron. Express. – 2007. – № 4(23). – P. 724-730.

[17] Nandakumar, K. A fingerprint cryptosystem based on minutiae phase spectrum // In Proc. of IEEE Workshop on Information Forensics and Security (WIFS). – 2010. – P. 1-6.

[18] Hao F., Anderson R. and Daugman J. Combining Cryptography with Biometrics Effectively // IEEE Transactions on Computers. – 2006. – № 55(9). – P. 1081-1088.

[19] Rathgeb, C., Uhl, A. Adaptive fuzzy commitment scheme based on iris-code error analysis // In Proc. of the 2nd European Workshop on Visual Information Processing (EUVIP'10). – 2010. – P. 41-44.

[20] Lozhnikov P.S., Sulavko, A.E., Volkov D.A. Application of noise tolerant code to biometric data to verify the authenticity of transmitting information / Control and Communications (SIBCON), 21-23 May 2015, Omsk, Russia – p.1-3.

A Memory Efficient Distributed Fuzzy Joint Points Clustering Algorithm

Can Atilgan
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
can.atilgan@deu.edu.tr

Efendi Nasibov
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
efendi.nasibov@deu.edu.tr

***Abstract*—The fuzzy joint points (FJP) is a method that uses a fuzzy neighborhood notion to deal with neighborhood parameter selection issue of classical density-based clustering and offers an unsupervised clustering tool. Recent works improved the method in terms of speed to enable the method for big data applications. However, space efficiency of the method is still a limiting factor. In this work, we discuss techniques to improve the space efficiency of the method, so that FJP is applicable regardless of the size of data and offer a distributed version of the algorithm.**

***Index Terms*—Clustering, Fuzzy joint points (FJP), Memory efficiency.**

I. Introduction

Clustering is one of the most common tools for data analysis. Basically, clustering algorithms partition data into groups called clusters, such that intra-cluster similarity is maximized and inter-cluster similarity is minimized. A clustering method is mainly characterized by its choice of similarity measure.

The k-means algorithm [1] is a simple and yet an effective, and a very reputed algorithm, that defines similarity directly by a distance function. The k-means and its many variants typically need a level of supervision for determining the number of clusters and the initial partition. Moreover, using merely a distance function to measure similarity implies that the clusters have similar and regular shapes.

Density-based (also known as neighborhood-based) clustering methods have the ability to discover clusters in data regardless of their shapes and DBSCAN [2] is considered to be the pioneer of such methods. Also, density-based algorithms do not require number of clusters to be determined in advance. On the other hand, DBSCAN and its many variants require one or more neighborhood parameters to be set by user. Adjusting the parameter values naturally needs experimenting and is usually a subtle task.

The fuzzy joint points (FJP) method [3] makes us of fuzzy neighborhood notion to deal with parameter dependency of classical density-based approach and offers a fully unsupervised clustering tool. It is shown in [4] that DBSCAN is very sensitive to 2 input parameters and needs delicate choice of parameters to achieve quality clustering, whereas FJP achieves similar results in automatic fashion.

In its early versions [3,5] FJP appeared to be quite slow, making it impractical to use for even small data sets, such that the experiments in those works could be done on data sets up to only 1053 data points. An optimal time version of FJP presented in [6] enabled the approach to be used for rather big data sets. An implementation of the optimal time algorithm was capable of clustering 10000 data points only in a few seconds. Moreover, a heuristic improvement of FJP called αScan was also proposed in [6] that accelerates the algorithm by a constant factor.

Even though the FJP methods offer the advantages of the density-based clustering without need of supervision of user and their speed have reached to a point where they are applicable to large data sets, they suffer from rather high memory usage. In fact, it was the memory that limited the size of the data sets used in the previous practical experiments rather than the calculation time. The space complexity of FJP was not addressed in the past and it is intended in this paper to handle this important issue.

II. The FJP Methods

Several definitions that are essential to FJP are given in the following [7].

Definition 1. A conical fuzzy point $P = (p, R) \in F(E^m)$ is a fuzzy set whose membership function is given by

$$\mu(x) = \begin{cases} 1 - \frac{d(x,p)}{R}, & d(x,p) \leq R \\ 0, & \text{otherwise} \end{cases},$$

where $p \in E^m$ is the center of the conical fuzzy point P, $d(x,p)$ is the distance between the point $x \in E^m$ and the center p, and $R \in E^1$ is the radius of the support set $supp\ P$, where

$$\text{supp P} = \{x \in E^m \mid \mu_P(x) > 0\}.$$

Definition 2. The α-level set of a conical fuzzy point P is defined as the follows

$$P_\alpha = \{x \in E^m \mid \mu_P(x) \geq \alpha\}.$$

Definition 3. A fuzzy neighborhood relation $T: X \times X \to [0,1]$ on a set X is defined by

$$T(X_1, X_2) = 1 - \frac{d(x_1,x_2)}{2R},$$

where $x_1, x_2 \in E^m$ are the centers of the conical fuzzy points $X_1, X_2 \in X \subset F(E^m)$, respectively. For $|X| = n$, the radius is chosen as

$$R = \frac{\max d(x_i,x_j)}{2} = \frac{d_{max}}{2}, \ \ i,j = \overline{1,n},$$

in the methods. This implies

$$T(X_i, X_j) = 1 - \frac{d(x_i,x_j)}{d_{max}}.$$

Definition 4. The α-neighborhood of a conical fuzzy point P is the set of conical fuzzy points, whose values of fuzzy neighborhood relations to P are larger than or equal to a particular α value. In other words, conical fuzzy points $X_1 = (x_1, R)$ and $X_2 = (x_2, R)$ are α-neighbors, if

$$T(X_1, X_2) \geq \alpha, \ \ \alpha \in (0,1],$$

and it is denoted by $X_1 \sim_\alpha X_2$.

Definition 5. Let $T: X \times X \to [0,1]$ be a fuzzy neighborhood relation. A relation $T' = (t'_{ij})_{n\times n}$ that is calculated as

$$t'_{ij} = \max_k \min\{t_{ik}, t_{kj}\}, \ \ i,j = 1, \dots, n,$$

is the max-min composition of the relation matrix $T = (t_{ij})_{n\times n}$, and is denoted by $T \circ T$.

Definition 6. Let $T: X \times X \to [0,1]$ be a fuzzy neighborhood relation. A relation $\hat{T}$ that is defined by

$$\hat{T} = T \cup T^2 \cup \dots \cup T^n \cup T^{n+1} \cup \dots,$$

where $T^k = T \circ T^{k-1}, \ k \in \mathbb{Z}, \ k \geq 2$, is the max-min transitive closure of the relation T.

Definition 7. Conical fuzzy points X_1 and X_2 are fuzzy α-joint points, if there is a sequence of α-neighbors between them, such that

$$X_1 \sim_\alpha Y_1, Y_1 \sim_\alpha Y_2, \dots, Y_{k-1} \sim_\alpha Y_k, Y_k \sim_\alpha X_2, \ \ k \geq 0.$$

It is proved in [3] that any points X_1, X_2 are fuzzy α-joint points, if and only if

$$\hat{T}(X_1, X_2) \geq \alpha,$$

where $\hat{T}: X \times X \to [0,1]$ is the max-min transitive closure of the relation $T: X \times X \to [0,1]$.

Given a set $X = \{x_1, x_2, \dots, x_n\}$ a pseudo-code of the FJP algorithm is as follows.

The underlying idea of the FJP algorithm is to calculate the transitive closure of fuzzy neighborhood relation between points and, according to transitive closure, to handle the equivalence classes as the clustering partitions. The fundamental observation here is that the different α_i-level sets, where α_i are the unique values in the transitive closure matrix result in unique clustering partitions [3].

Given the ideas above, for an input set $X = \{x_1, x_2, \dots, x_n\}$ a pseudo-code of the FJP algorithm is as follows.

Algorithm 1: FJP
Input: Dataset X;

Step 1. Calculate the distances: $d_{ij} = d(x_i, x_j), \ i,j = 1, \dots, n$; $d_{max} = \max d_{ij}, \ i,j = 1, \dots, n$;

Step 2. Calculate the fuzzy neighborhood relation T: $t_{ij} = 1 - \frac{d_{ij}}{d_{max}}, \ i,j = 1, \dots, n$;

Step 3. Calculate the transitive closure matrix $\hat{T}$;

Step 4. Constitute the array $V = (a_i)_{i=1}^{n^2}$ consisting of the elements $\alpha_i, i = 1, \dots, n^2 - 1$ of $\hat{T} = (\hat{t}_{ij})_{n\times n}$ matrix sorted in descending order: $\alpha_i \geq \alpha_{i+1}, \ i = 1, \dots, n^2 - 1$, Set:
$\Delta\alpha_i = \alpha_i - \alpha_{i+1}, \ i = 1, \dots, n^2 - 1,$
$z = \arg \max_{i=1,\dots,n^2-1} \Delta\alpha_i$;

Step 5. Determine the connected components of the α_z-level set of $\hat{T}$ to obtain the resulting clustering partition and output the partition.

Stop.

Step 4 of Algorithm 1 involves finding a critical value named α_z, whose level set provides the final clustering partition. This value serves as the final neighborhood parameter, providing the autonomy of the method.

The algorithm requires a transitive closure calculation, which is a computational bottleneck. αScan is a heuristic approach that avoids the transitive closure calculation by trying different α_i-levels, where $\alpha_i - \alpha_{i-1} = \Delta\alpha$ and $\Delta\alpha$ is a parameter given by user. At first glance, having a parameter dependency contradicts the autonomy of FJP, and yet the parameter is easy to select so that it achieves the same

clustering quality. Parameter selection issue is studied in detail in [7]. αScan algorithm is also essential for our work in this paper, and so its pseudo-code is given in the following.

Algorithm 2: αScan
Input: Dataset X, and scanning unit $\Delta\alpha$;

Step 1. Calculate the distances: $d_{ij} = d(x_i, x_j),\ i,j = 1,\dots,n$; $d_{max} = \max d_{ij},\ i,j = 1,\dots,n$;

Step 2. Calculate the fuzzy neighborhood relation T: $t_{ij} = 1 - \frac{d_{ij}}{d_{max}},\ i,j = 1,\dots,n$;

Step 3: Set:
current α value $\alpha_c = 1$,
number of clusters of the last partition $k_l = 0$,
number of repetitions of the current partition $r_c = 0$,
number of repetitions of the most repeated partition $r_{max} = 0$,
the best partition obtained so far $S^* = \emptyset$;

Step 4. Calculate α_c-level set of T and determine its connected components to obtain the current clustering partition. Set results as:
$k =$ Number of the clusters in the current clustering partition,
$S =$ Resulting current clustering partition;

Step 5.
if $k_l = k$ then the current clustering partition is not unique, set: $r_c = r_c + 1$;
else
if $r_{max} < r_c$ then the current clustering partition is the best one so far, set:
$r_{max} = r_c,\ \alpha_z = \alpha_c + \Delta\alpha,\ S^* = S$;
end if;
Set: $r_c = 1,\ k_l = k$;
end if;

Step 6. Set: $\alpha_c = \alpha_c - \Delta\alpha$;
if $\alpha_c > 0$ and $k > 1$ then, go to Step 4;
else Output the clustering partition S^* (and optionally α_z).

Stop.

Both of the algorithms involve calculations on nxn matrices implying a quadratic space complexity, which is a serious drawback in practice considering limited memory availability. The next section analyzes the space requirement of the methods in detail and discusses a data set division technique to overcome the issue

III. Improving Memory Efficiency of FJP

First of all, it is important to note that both Algorithm 1 and 2 run in $O(n^2)$ time making them fast enough for even large data sets. A theoretical and experimental analysis of those algorithms in terms of calculation time can be found in [6].

The largest data set used in computational experiments consisted of 10000 data points in the previous works, even though the longest CPU time observed was only a few seconds. The reason why no bigger data sets were used was that the available RAM (4 GB) was insufficient.

Let us make a rough estimation of memory requirement for an FJP application, assuming that no additional space is used besides an nxn matrix of real numbers. Note that the number of attributes of a data point is irrelevant, since the matrix consists of the neighborhood values. If each cell holds a single precision floating point number, i.e. each number requires 4 bytes of memory, then $4n^2 = b$ and $n = \frac{\sqrt{b}}{2}$ hold, where n is the number of data points and b is the number of bytes required. Fig 1. gives a comprehensive illustration of the fact that a quadratic space complexity is impractical for a clustering application to work on big data.

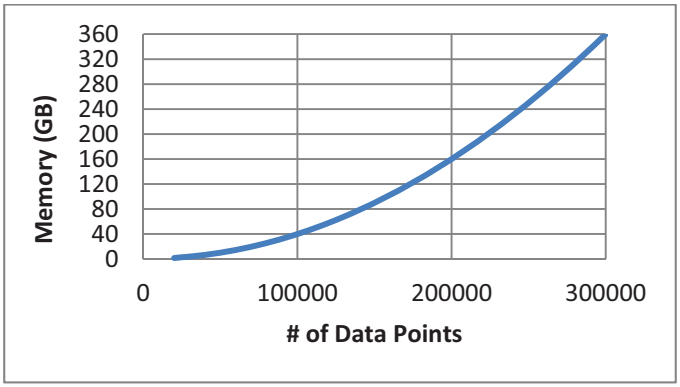


Fig. 1. Number of data points versus required memory in GigaBytes.

A baby step can be reducing the precision of floating point numbers, using the following observations: **1.** A fuzzy relationship matrix consists of numbers in [0, 1] interval; **2.** Empirically, a very rough precision is enough to operate. The first observation is obvious, whereas the second one is an implication of previous work [7,8], where a number of experiments with different α settings showed that sufficient scanning parameters changes between 0.382 and 0.059. Therefore, a representation of numbers in [0, 1] interval with a spacing of $\frac{1}{255} \cong 0.004$ would be enough to obtain the optimal clustering results. That is to say, 1 byte of memory could be sufficient instead of 4 bytes for each number in the matrix.

In order to improve the space efficiency, so that FJP is applicable regardless of size of data set, a more generic approach is needed. Dividing the data set into subsets to obtain local clustering results and aggregating the local results into a final clustering is a steady approach, which is applied to a number of clustering methods [9,10]. Naturally, it is also a way to adapt a method to the distributed computing. In a density based clustering method where neighborhood parameters are preset, this generic divide-and-conquer approach can be applied straightforward: Local clusters only need to exchange a small portion of their members that are

potentially within the neighborhood parameter of other local clusters (called border points in DBSCAN) and they are merged if any of 2 members from different local clusters are discovered to be neighbors.

In case of FJP, the neighborhood parameters are not present in advance. Thus, another way of exchanging data points between local clusters is needed. Assuming the clusters are convex sets, choosing the convex hull of local clusters as their candidate member to merge with other clusters can be considered. However, FJP may build clusters of irregular, non-convex shapes and there may be situations where convex hulls miss actual connecting points of 2 local clusters. Such a situation is illustrated in Fig. 2.

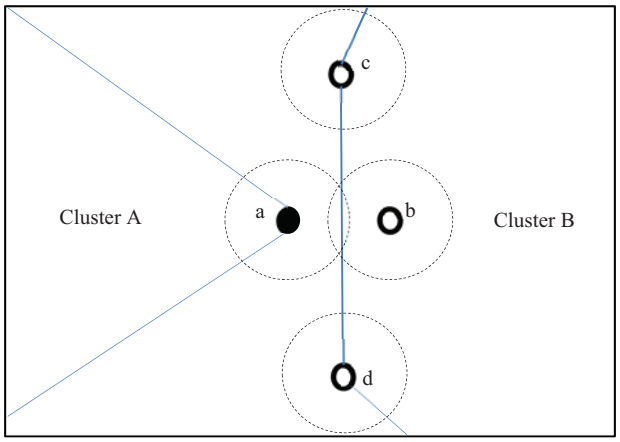


Fig. 2. Point a is in convex hull set of cluster A, and points c and d are in convex hull set of cluster B. Only a and b are α-neighbors.

A generalization of convex hulls are known as α-shapes [11,12]. In order to avoid confusion of α notations, they are referred as r-shapes from this point on. The concept of r-shapes is given in the following.

Definition 8. A generalized disc of radius $1/r$ is a disc of radius $1/r$ if $r > 0$, the complement of a disc of radius $-1/r$ if $r < 0$, and a half-plane if $r = 0$.

Definition 9. Let S be a finite set of points on plane. For $p, q \in S$, if there exists a generalized disc of radius $1/r$ with p and q on its boundary and which contains all other points in S, then p and q are r-neighbors.

r-shapes are constituted by connecting r-neighbors with edges to outline the shape of S. An Fig 3. shows an example of an r-shape. r value determines the detail of the shape by effecting number of r-neighbors and for $r = 0$, r-shape is a convex hull. Note that the above definitions are given in $\mathbb{R}^2$, but extensions to $\mathbb{R}^d$ for arbitrarily larger integers d can be easily made. On the other hand, efficient r-shape construction algorithms deal with data sets up to 3 dimensions. Extension to higher dimensions and algorithms to discover r-shapes, and more detail on the topic can be found in [12].

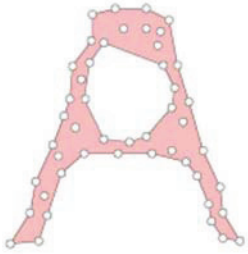


Fig. 3. An r-shape [12].

Set of r-neighbors of a local cluster of FJP on a divided data set is a natural list of a candidates to connect to another local cluster. This enables the aforementioned data set division technique to be applied to FJP, regardless of its parameter ambiguity disadvantage. A pseudo-code of such an algorithm is in the following.

Algorithm 3: Distributed FJP
Input: Dataset X, number of subsets t, disc radius r;

Step 1. Sort X according to an arbitrary attribute of data;

Step 2. Divide X into t consecutive distinct subsets of roughly equal size $X = \bigcup_{i=1}^{t} X^i$;

Step 3. Build local clusters of each subset X^i, $i = 1, \dots, t$ using FJP;

Step 4. For each local cluster $C^{i,j}$ in X^i, $i = 1, \dots, t$, $j = 1, \dots, k$ calculate sets of r-neighbors $r(C^{ij})$;

Step 5. Merge any 2 clusters from different subsets if $p \in r(C^{i,*})$ and $q \in r(C^{i+1,*})$, $i = 1, \dots, t-1$ are α-neighbors, where * denotes any integer in [1, k];
Output the global clustering after merging process;

Stop.

Before discussing how Algorithm 3 handles excessive space requirement of FJP, following issues about the algorithm itself emerge. How to determine r? Does it guarantee the same clustering results with FJP methods? How to maintain a common α-neighborhood among subsets as their critical-alpha values may differ?

Choosing r appears to be the most challenging issue of the algorithm. Intuitively, set of r-neighbors of each local cluster $r(C^{ij})$ is supposed to be small enough to ensure that merging step itself does not dominate local memory usage. This can be solved by another local procedure.

Whether Algorithm 3 can achieve same clustering performance with other FJP algorithms depends on maintaining an appropriate common α-neighborhood notion. The fuzzy relation function given in Definition 3 requires the maximum distance between data pairs, implying local clustering requires global information. If the data set is min-

max normalized, replacing d_{max} with $\sqrt{m}$, where m is number of attributes (dimensions) provides a consistent relationship function among subsets without need of global maximum distance. Having a common relationship function, it is easy for αScan to maintain a common α-neighborhood: At each iteration there is a constant α_c value, which is supposed to be the same for any subset. Conversely in FJP, α_z value that determines the final clustering is extracted from the data set at hand. Thus, α_z values might be different for each subset of the data set. Again, this issue can be solved by a procedure between consecutive subsets while merging.

Considering distributed version of the algorithm achieves optimal or sub-optimal clustering results, it is immediately seen that each local clustering requires $O((n/t)^2)$ space. Since t is determined by user, it can be adjusted to fit in the available memory. Likewise, if r parameter is appropriately set as mentioned above, memory limitation is surpassed.

IV. Conclusion

FJP is an unsupervised, density-based clustering method offering quality clustering. Previous studies presented a great improvement in terms of speed to handle big data using FJP. However, low memory efficiency disabled the method to extend its bounds further.

In this work, using r-shapes (or α-shapes originally) to determine border points of local clusters, we offered a distributed version of FJP that allows the users to adjust the memory requirements to their needs. It also serves as a scheme to apply the method in distributed environments.

In future, an experimental work could help investigating possible practical problems as well as measuring computational performances. Besides, devising algorithms to build r-shapes of high dimensional data sets might be a fructuous subject.

References

[1] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1 (Univ. of Calif. Press), pp. 281–297, 1967.

[2] M. Ester, H.P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining* (KDD-96), pp. 226–231, August 1996.

[3] E. N. Nasibov and G. Ulutagay, "A new unsupervised approach for fuzzy clustering," *Fuzzy Sets and Systems*, vol. 158(19), pp. 2118–2133, October 2007.

[4] E. Nasibov and G. Ulutagay, "Robustness of density-based clustering methods with various neighborhood relations", *Fuzzy Sets and Systems*, vol. 160(24), pp. 3601–3615, December 2009.

[5] G. Ulutagay and E. Nasibov, "On fuzzy neighborhood based clustering algorithm with low complexity", *Iranian Journal of Fuzzy Systems*, vol. 10(3), pp. 1–20, June 2013.

[6] E. Nasibov, C. Atilgan, M. E. Berberler and R. Nasiboglu, "Fuzzy joint points based clustering algorithms for large data sets," *Fuzzy Sets and Systems*, vol. 270, pp. 111–126, July 2015.

[7] E. Nasibov and C. Atilgan, "Parameter selection in fuzzy joint points clustering algorithms", *Proceedings of 9th International Conference on Application of Information and Communication Technologies,* pp. 8–12, 2015.

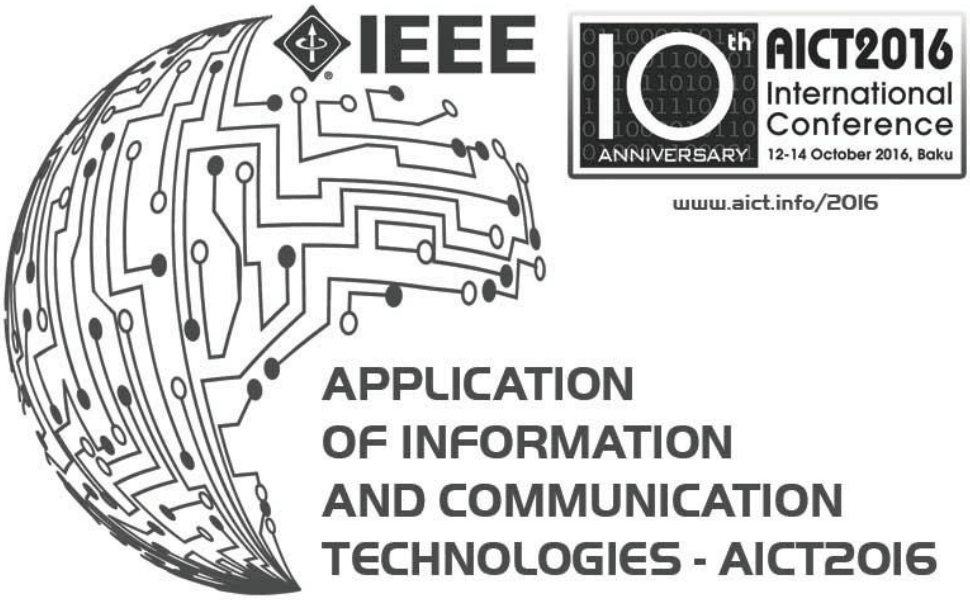
[8] Nasibov E., Atilgan C., "A Note on Fuzzy Joint Points Clustering Methods for Large Data Sets", *IEEE Transactions on Fuzzy Systems*, in press. (DOI 10.1109/TFUZZ.2016.2551280)

[9] X. Xu, J. Jager and H.-P. Kriegel, "A Fast Parallel Clustering Algorithm for Large Spatial Databases", *Data Mining and Knowledge Discovery*, vol. 3, pp. 263–290,1999.

[10] E. Januzaj.,H.-P. Kriegel,M. Pfeifle: "DBDC: Density-Based Distributed Clustering", *Proceedings of 9th International Conference on Extending Database Technology*, pp. 88–105, March 2004.

[11] H. Edelsbrunner, D. G. Kirkpatrick and R. Seidel, "On the shape of a set of points in the plane", *IEEE Transactions on Information Theory*, vol. 29(4), pp. 551–559, Jul. 1983.

[12] H. Edelsbrunner., "Alpha shapes - a survey", *in Tessellations in the Sciences; Virtues, Techniques and Applications of Geometric Tilings Visualization and Data Analysis*, Ed. Springer Verlag, 2010.



CONFERENCE PROCEEDINGS

SESSION 5.

COMMUNICATION, NETWORK AND HARDWARE

12-14 October 2016, Baku, Azerbaijan

2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)

Optimum Power Allocation for Non-orthogonal Multiple Access (NOMA)

Talgat Manglayev
School of Engineering,
Nazarbayev University,
Astana, Kazakhstan Z05H0P9
talgat.manglayev@nu.edu.kz

Refik Caglar Kizilirmak
School of Engineering,
Nazarbayev Universit,y
Astana, Kazakhstan Z05H0P9
refik.kizilirmak@nu.edu.kz

Yau Hee Kho
School of Engineering,
Nazarbayev University,
Astana, Kazakhstan Z05H0P9
yauhee.kho@nu.edu.kz

***Abstract*—An optimum power allocation scheme for non-orthogonal multiple access (NOMA) with successive interference cancellation (SIC) technique for downlink channels of next generation cellular networks is proposed in this paper. In NOMA downlinks, the signals for each user are superimposed at the transmitter side and SIC is implemented at the receiver side to separate them. Numerical results show that the overall system throughput performance can be improved significantly by the proposed power allocation algorithm with NOMA in comparison with orthogonal multiple access (OMA), which is implemented in previous wireless telephone technology generations.**

Keywords*—Future Radio Access, OMA, NOMA, SIC, Power allocation

I. Introduction

The increasing demands for high data rate and ubiquitous access to multimedia contents have resulted in an explosive proliferation of mobile equipment. This has put tremendous performance pressure on cellular networks in terms of providing more reliable, faster and energy efficient services in the upcoming wireless generations (G), especially 5G and beyond.

Conventionally, various orthogonal multiple access (OMA) techniques in time, frequency and/or code domains, have been implemented successfully in existing wireless systems where each user's signal can be differentiated easily. Time division multiple access (TDMA) is implemented in first generation analog telecommunication systems, while frequency division multiple access (FDMA) together with TDMA are being used in the existing global system for mobile communications (GSM). Code division multiple access (CDMA), on the other hand, is widely implemented in 3G and finally, the latest 4G network Long Term Evolution (LTE) and LTE-Advanced are employing orthogonal frequency division multiplexing (OFDM), orthogonal frequency division multiple access (OFDMA), multiple input multiple output (MIMO) and precoding etc. Suffice to say that OMA techniques have been successfully implemented in mobile telecommunication technologies since 1980. However, it is expected that data capacity and spectral efficiency for these wireless mobile communication networks will reach bottleneck soon.

Non-orthogonal multiple access (NOMA) technique, employing super positioning of signals emitted from the transmitter side (TX) in time, frequency and code simultaneously, is believed to be able to solve this bottleneck. It is stated as a solution for spectrum efficiency and improvement of throughput performance for upcoming radio access [1]. This technique, however, requires careful power allocation from the TX in relation to the distance of each receiver; those farther away from the TX shall be allocated more power to compensate for transmission loss. This will also ensure fairness in term of data capacity. Furthermore, due to super positioning of all receivers' signals during transmission, cancellation of unwanted interference and noise, i.e. other unrelated receivers' signals, will be required at the receiver side (RX). This will entail high computational complexity at both TX and RX; optimal power allocation is required at the TX while successive interference cancellation (SIC) is required at the RX.

In this paper we investigate an optimum power allocation scheme for NOMA with SIC for the downlink channels of next generation cellular networks. Some background information on NOMA with SIC model is explained in Section II, followed by our system model in Section III. We discuss our numerical results in Section IV and conclude our work in Section V.

II. Background

As mentioned earlier, the requirement for higher level of communication is being met by usage of new kinds of mobile devices, services and applications that enable high resolution video streaming (4K), media rich social network services, augmented reality etc. [2]. NOMA is being regarded as a potential scheme that can meet the above-mentioned needs and requirements. Many researchers have compared the performances of NOMA and SIC with other technologies in different scenarios. In [3], the capacity performances of OFDMA which is implemented in 4G, is compared against NOMA with SIC for two user equipments (UE). They implemented a highly-complex full search power algorithm (FSPA) at TX, and the results clearly prove NOMA's dominance over OFDMA with 32 percent and 48 percent increase in capacity for UE_1 and UE_2 respectively. Computer simulations in [4] also showed superior performance of NOMA cell throughput and cell-edge user throughput over those in OMA for wide band and sub band scheduling in high mobility and low mobility scenarios. Another results of computer simulations with transmission rate back-off algorithm applied in NOMA scheme [5] achieved clear average user throughput and cell-edge user throughput gains that are similar to perfect channel cases in OMA scenario.

NOMA implements UE multiplexing in power domain, however power allocation is not a novelty for NOMA. Ac-

cording to [6] the larger the difference in the allocated power ratio is between the UEs, the smaller the inter-user interference is. Such resource allocation is discussed in [7] where iterative sub-optimal power allocation algorithm based on difference of convex (DC) was offered for OFDMA-based NOMA systems. The best simulation result was approximately 16 x 10^6 bps for 25 UEs. Such results can be reached in the current 4G network and will be suitable for mobile device services. One constraint remains that power has to be allocated on the basis of providing fair capacity rates among UEs, besides sum capacity. Optimum power allocation algorithms with fairness rates were offered in [8] where simulations for OFDM systems with two UEs were done. The algorithm is implemented in two steps and complex for dynamic power allocation computations.

III. System Model

This section describes the underlying system model used in this paper where we considered downlink transmission only. In a network with K UEs (see Fig. 1), we assume the UE_1 is the closest to the base station (BS), and UE_K is the farthest one.

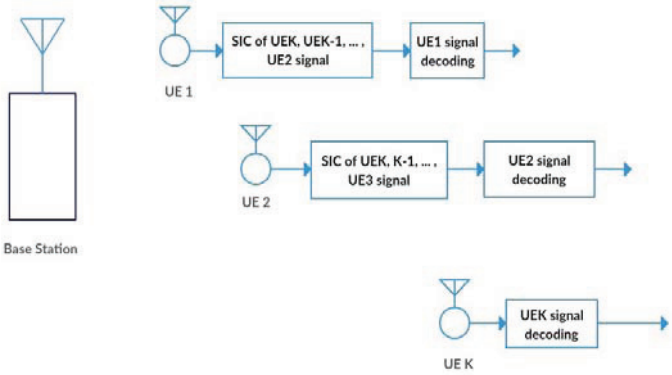


Fig. 1. Downlink NOMA for K UEs.

In NOMA, signals emitted from the BS for each UE are superimposed on top of each other. More power is allocated to UE located farther from the BS. In the network, all UEs receive the same signal which contains the information for all users. Then, each UE decodes the strongest signal first, then subtracts the decoded signal from the received signal. SIC receiver iterates the subtraction until it finds its own signal. UE located close to the BS can cancel the signals of the farther UEs. The farthest UE, since its signal weighs the most in the received signal, the first signal it decodes is its own signal.

The transmitted signal by the BS can be written as

$$x(t) = \sum_{k=1}^{K} \sqrt{\alpha_k P_T} x_k(t) \tag{1}$$

where $x_k(t)$ is the individual information signal and α_k is the power allocation coefficient for the UE_k and P_T is the total available power at the BS. The power allocated to each UE_k then becomes $P_k = \alpha_k P_T$. In NOMA, power is allocated according to the distance of UEs to the BS: UE_1 is the closest to the BS, so it is allocated the least power, whereas UE_k is the farthest one, therefore it has the highest amount of power.

The received signal at the UE_k is

$$y_k(t) = x(t) g_k + w_k(t) \tag{2}$$

where g_k is the channel attenuation gain for the link between the BS and the UE_k and $w_k(t)$ is the additive white Gaussian noise at the UE_k with mean zero and density N_0 (W/Hz).

When NOMA is used, the throughput (bps) for each UE can be written as [9]

$$R_k = W \log_2 \left(1 + \frac{P_k g_k^2}{N + \sum_{i=1}^{k-1} P_i g_k^2}\right) \tag{3}$$

where W is the available transmission bandwidth and N is the total noise power $N = N_0 W$.

In OMA, on the other hand, UEs are allocated orthogonal frequency or time slots in order to receive their information. When the total bandwidth and power are shared among the UEs equally, the throughput for each UE for OMA becomes

$$R_k = W_k \log_2 \left(1 + \frac{P_k g_k^2}{N_k}\right) \tag{4}$$

where $W_k = W/K$ and $N_k = N_0 W_k$.

The sum capacity for both OMA and NOMA can be written as

$$R_T = \sum_{k=1}^{K} R_k. \tag{5}$$

We further define fairness index as [10]

$$\text{F} = \frac{(\sum R_k)^2}{K \sum R_k^2} \tag{6}$$

which indicates how fair the system capacity is shared among the UEs, i.e, when F gets close to 1, the capacity for each UE gets close to each other.

The objective of our power allocation mechanism is to maximize the sum capacity under a fairness constraint for NOMA systems. The optimization problem considered in this paper is formulated as

$$\begin{aligned} \underset{\alpha_k}{\text{maximize}} \quad & \sum_{k=1}^{K} W \log_2 \left(1 + \frac{P_k g_k^2}{N + \sum_{i=1}^{k-1} P_i g_k^2}\right) \\ \text{subject to:} \quad & \sum_{k=1}^{K} P_k \le P_T \\ & P_k \ge 0, \ \forall k \\ & \text{F} = F' \end{aligned}$$

where F' is the target fairness index in the network. We obtain the power allocation coefficients α_k for each UE_k through exhaustive search. The algorithm 1 describes optimum powered allocation is given below.

Algorithm 1 Optimum power allocation (OPA)

```
initialization
initialize powerMatrix include all possible PAs
set fairnessConstraint to F'
for i in powerMatrix do
   calculate capacity
   calculate fairnessIndex
   if fairnessIndex ≤ fairnessConstraint then
      set capacity(i) to zero
   end if
end for
set maximumCapacity to zero
for i in capacity do
   calculate capacity(i)
   if capacity(i) ≥ maximumCapacity then
      set maximumCapacity to capacity
   end if
end for
```

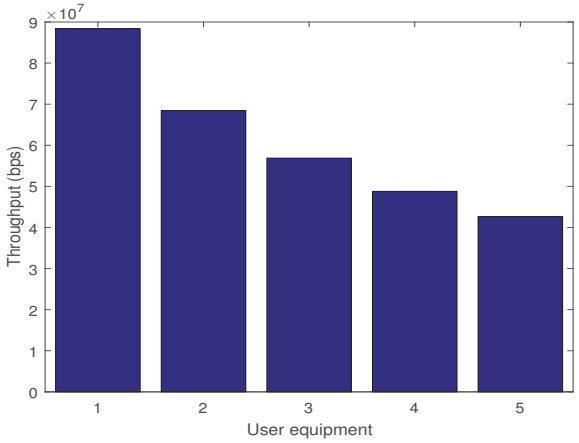


Fig. 2. OMA maximum capacity data throughput rate performance for five UEs with fairness index = 0.9.

IV. Numerical results

In this section, we present our numerical results that compare the achievable throughput for OMA and NOMA. We consider K = 5 UEs in the network. The available bandwidth is taken as W = 50 MHz. Total available power at the BS is P_T = 1 Watt. The distance between each terminal in the network is set at 50 meters. Carrier frequency is 1 GHz. The channel gains are obtained with Okumura-Hata propagation model and are calculated for 5 UEs as $[g_1 \ldots g_5]$ = [-33.21 -36.23 -37.99 -39.24 -40.20] in dB. The noise density N_0 is taken as 10^{-17} W/Hz.

Fig. 2 gives the achievable data rates for the UEs in the network with OMA when both power and bandwidth is equally shared between the UEs. The maximum rate of UE_1 is $9\text{x}10^7$ bps, UE_2 barely reaches $7\text{x}10^7$ bps, then the rate gradually decreases and reaches $4\text{x}10^7$ bps for UE_5. The rates are calculated using (4) and the sum capacity for this scenario becomes approximately $3.05\text{x}10^8$ bps. Moreover, in OMA configuration, the fairness is calculated as 0.9 using (6).

In evaluation of the performance of NOMA with optimal power allocation (OPA), we considered three different scenarios with fairness indexes 0.5, 0.7 and 0.9. Numerical results show clear dominance of NOMA over OMA for each fairness index case of NOMA.

In Fig. 3, the data rate distribution is given when OPA is applied when the fairness index is 0.5. UE_1 achieves $2.5\text{x}10^8$ bps, data rate of the UE_2 is slightly lower than $1\text{x}10^8$ bps. In general, all five UEs achieve higher rates than that with OMA. The low fairness of 0.5 can be observed. The sum capacity for this case is calculated as approximately $4.37\text{x}10^8$ bps. The optimum power allocation coefficients for this fairness index equal to $[\alpha_1 \ldots \alpha_5]$= [0.07 0.2 0.23 0.24 0.26].

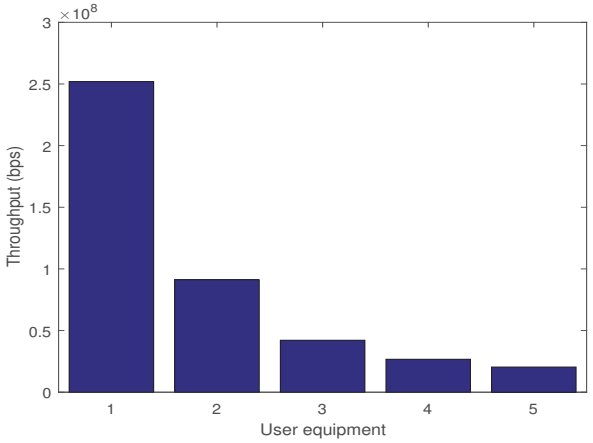


Fig. 3. NOMA maximum capacity data throughput rate performance for five UEs with fairness index = 0.5.

When the fairness requirement is increased to 0.7 (see Fig. 4), the data rates obtained for UE_4 and UE_5 are slightly less as compared to OMA, however UE_1, UE_2 and UE_3 perform better. UE_1 obtains slightly more than 16 x 10^7 bps, UE_2 achieves approximately 12 x 10^7 bps, however UE_3, UE_4 and UE_5 obtain 5 x 10^7, 3.8 x 10^7 and 3 x 10^7 bps respectively. The optimum power allocation coefficients for this fairness index equal to $[\alpha_1 \ldots \alpha_5]$ = [0.02 0.14 0.23 0.30 0.31].

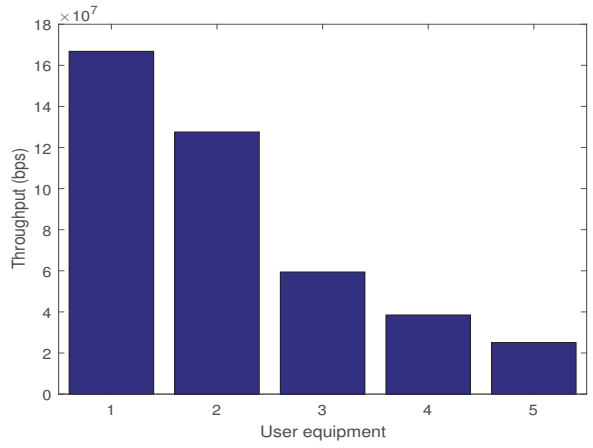


Fig. 4. NOMA capacity data throughput rate performance for five UEs with fairness index = 0.7.

In Fig. 5, NOMA performance with fairness index equal to 0.9 is implemented. Numerical results prove dominance of

NOMA performance over OMA for all five UEs. Cell edge UE_5 achieves 5 x 10^7 bps against 4.4 x 10^7 bps in OMA. The highest throughput 12.1 x 10^7 bps with only 0.01 power allocation coefficient is obtained for UE_1 in NOMA against 9 x 10^7 for OMA. The sum capacity with fairness index 0.9 by NOMA is approximately 3.95 x 10^8 bps, while achievable sum capacity for OMA is approximately 3.05 x 10^8 bps. The optimum power allocation coefficients for this fairness index equal to $[\alpha_1 \ldots \alpha_5]$ = [0.01 0.05 0.12 0.28 0.54]. Among

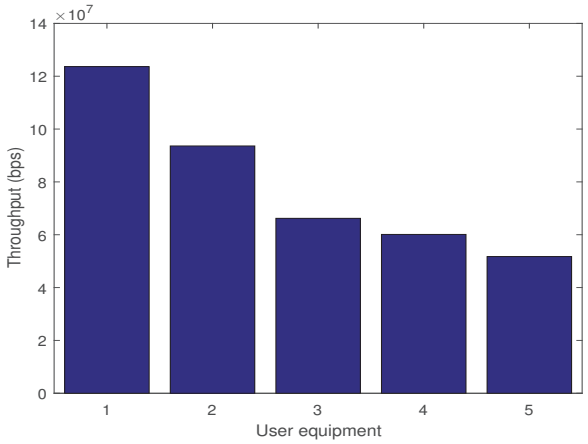


Fig. 5. NOMA capacity data throughput rate performance for five UEs with fairness index = 0.9.

different fairness index requirements, the highest sum capacity is obtained with fairness index of 0.5. Fig. 6 shows achievable data throughput obtained for different fairness indexes. It can be clearly seen that as the fairness requirement increases, the less is the difference between each UE throughput. For all fairness indexes, the UE_1 has the highest throughput and its value reduces as the fairness requirement increases. For UE_3, UE_4 and UE_5, on the other hand, the throughput values increase as the fairness index increases. The throughput for UE_2 gradually increases till it reaches its peak at the fairness index value of 0.7, and then it moderately reduces. The throughput for all the UEs is in the range of 0.5 x 10^8 and 1.4 x 10^8 bps at the 0.9 fairness index. Sum capacity curve confirms our previous results: the sum capacity for all UEs is reduced, when the fairness index requirement is increased.

V. Conclusion

In this paper, we implemented an algorithm to determine the optimum power allocation ratio for NOMA with SIC. Optimum power ratios for each UE are obtained to maximize the sum capacity while maintaining a target fairness index. Calculations were performed for three different fairness indexes: 0.5, 0.7 and 0.9. The sum capacity results for NOMA with SIC outperform OMA for all three fairness indexes. Furthermore, it was shown that the sum capacity with NOMA reduces as the fairness index increases. In conclusion, NOMA with SIC is expected to be implemented in the future radio access technologies and promises to improve the spectral efficiency.

References

[1] M. C. Bahadir and H. Arslan. "Theoretical Analysis of the Co-Existence of LTE-A Signals and Design of an ML-SIC Receiver", *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4626-4639, Aug. 2015.

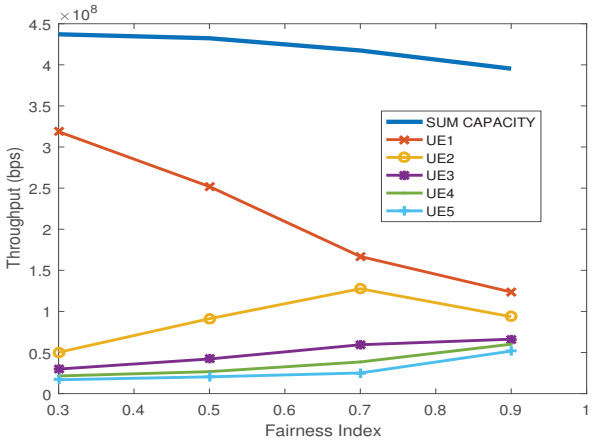


Fig. 6. Capacity obtained by five UEs with different fairness indexes.

[2] NTT DOCOMO INC. "5G Radio Access: Requirements, Concept and Technologies", in *DOCOMO 5G White Paper*, 2014, pp. 2-3.

[3] A. Benjebbour *et al.*, "Concept and Practical Considerations of Non-orthogonal Multiple Access (NOMA) for Future Radio Access", in *Int. Symp. Intelligent Signal Processing Communications Systems (ISPACS)*, Naha, 2013, pp. 770-774.

[4] A. Benjebbour *et al.*, "System-Level Performance of Downlink NOMA for Future LTE Enhancements", in *In Globecom Workshops (GC Wkshps)*, Atlanta, 2013, pp. 66–70.

[5] N. Nonaka, A. Benjebbour and K. Higuchi. "System-Level Throughput of NOMA Using Intra-beam Superposition Coding and SIC in MIMO Downlink When Channel Estimation Error Exists". In *IEEE Int. Conf. on Communication Systems (ICCS)*, Macau, 2014, pp. 202 – 206.

[6] Y. Saito *et al.*,"Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access", in *Vehicular Technology Conf. (VTC Spring)*, Dresden, 2013, pp. 1-5.

[7] P. Parida and S. S. Das. "Power allocation in OFDM based NOMA systems: A DC programming approach". In *IEEE Globecom Workshops (GC Wkshps)*, Austin, 2014, pp. 1026-1031.

[8] Zukang Shen, J. G. Andrews and B. L. Evans. "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints", *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, p. 2726-2737.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge: Cambridge University Press, 2005, pp. 240 - 241

[10] R.K. Jain, D.W. Chiu, and W.R. Hawe. "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System", Hudson, 1984, pp. 20-21

A General Extensible Framework for Mobile Location-based Information Systems

Mona A. E. Rizvi, Assan Toleuov, Dzhamshed Khaitov, Sagynysh Baitursinov
Computer Science Department, Nazarbayev University
Astana, Kazakhstan
mona.rizvi@nu.edu.kz

***Abstract*—Many of the world's great museums, cultural heritage sites, and other tourist locations offer their visitors some sort of companion mobile app. These apps have features that range from providing basic information to detailed exhibit data, guided tours, interactive displays and augmented reality. With very few exceptions, these apps are purpose-built for each specific venue. We describe a general extensible framework for mobile location-based information systems that provides a complete infrastructure for building apps of this type, with which only content creation is required to make a full-featured, venue-specific mobile app, but further customization is also possible and simple.**

***Index Terms*—Mobile computing, frameworks, location-based services.**

I. Introduction

Many of the world's great museums, cultural heritage sites, and other tourist locations offer their visitors some sort of companion mobile app. These apps have features that range from providing simple helpful information to detailed exhibit data, guided tours, interactive displays and augmented reality. With very few exceptions, these apps are purpose-built for each specific venue. We describe a general extensible framework for building apps of this type, with which only content creation is required to make a full-featured, venue-specific app, but further customization is also possible and simple.

The framework was modeled with museums and other cultural sites in mind − we collected requirements by reviewing tens of existing museum and cultural site apps, as well as by working with the National Museum of the Republic of Kazakhstan [1], for whom we are developing the first system based on the framework. Although we designed it using museums as a model, the framework is generic enough to support location-based information systems for a variety of places, whether indoor or outdoor, large or small, including college campuses, shopping malls, parks and gardens, or even entire cities. We have implemented a fully-functional prototype version of all the components in the framework, and are currently beginning an evaluation of the framework itself, as well as developing a number of extensions.

The framework consists of four main components, as shown in Figure 1:

- A content database to store all the information associated with the areas and objects in a given venue, along with a set of web APIs that allow access to the content from mobile clients
- A usage database to store information such as user data, access times and counts, and *likes*, along with the web APIs to update this data from mobile clients
- A web-based client application that allows content managers to maintain the content data and view the usage data
- A mobile app that allows visitors to access the content using a variety of different methods, and which updates the usage database based on user activity

The databases structure their data as generally as possible. In the content database, all the areas and objects in a venue are arranged in a simple hierarchy. All information can be entered and accessed in multiple languages. Content can be organized by *level*, to allow for customization for different user populations (e.g. children and adults). Content can be created and stored in an extensible set of media formats, including various text, audio and image formats.

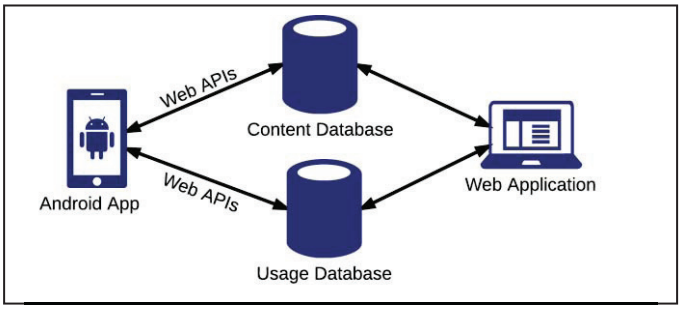


Fig. 1. The high-level model of the framework.

The mobile app can access content about objects, exhibits or general areas in a venue. It uses an extensible set of location determination and object identification mechanisms that currently includes: QR codes, NFC tags, and Bluetooth beacons. A prototype module for image recognition as an object identification method has been developed and is currently being integrated into the framework.

Section 2 of this paper covers some background and related work; section 3 discusses the database model and components, section 4 discusses the mobile application, section 5 discusses the web application and web APIs, and, finally, section 6 discusses the plans for evaluation and future development.

II. Background and Related Work

Mobile location-based information systems range from very general navigation-focused applications such as Google maps (maps.google.com) and in-car navigation systems to systems that can work indoors and give users specific information about each of the possibly many objects surrounding them. Our framework is designed to produce systems that fall at the more fine-grained end of this spectrum. A museum is a good representative example of a space that contains numerous individual areas and objects for which visitors might like to have access to additional related information. The additional information may not be displayable with the objects physically due to space, format or aesthetic considerations. Human, audio, and paper guides filled this need in the past, and now mobile apps are being created to provide the same services.

Our review of a large number of museum, cultural site and botanical garden apps showed that these types of apps are often built from scratch for specific venues or even for exhibits within a single venue. The results of that review are outside the scope of this paper; however, for example, the Smithsonian Museum main mobile app [2] lists over 30 other Smithsonian apps, for different buildings, different exhibits, using different input mechanisms, with different user interfaces, for different mobile platforms, etc. The result is wasted development effort and less than ideal user experiences. Although results for venue-specific applications do appear in the literature, many researchers have proposed more general solutions [3, 4, 5, 6, 7, 8].

In [7], the researchers have developed a general framework, but it is more suited to outdoor navigation than to complex indoor environments. Other researchers have proposed frameworks suitable to object-rich environments but limited them to a specific technology: in [8], RFID is the primary technology used for object identification, and in [6], image recognition for augmented reality is the focus of the work. Bruns et.al [4] presented interesting work that combines Bluetooth location determination with image recognition for object determination. A very flexible architecture was proposed in 2007 by Zafeiropoulos et. al [3], and, more recently, in 2013, researchers in Italy [5] presented their work on a mobile guide that uses NFC and QR codes for object determination, which is part of a more general framework for location-based information and social interaction.

Our framework is similar in many aspects to the work in [3] and [5], but ours offers more generality with respect to technologies than [5] and, unlike [3], is designed for extensibility.

III. The Database Components

The database components are central to our framework. Regardless of whether it is a museum, park, historical building, or shopping mall, basic information about a *place* or *venue* is stored in the content database; for example, its name, location, contact data, news and events. The set of items in a venue that can be individually identified by a user is stored hierarchically. For example, a museum may contain a number of buildings, and in each building there are various halls, and in each hall, there are multiple exhibits, and in each exhibit, there are individual works of art. In our database, each of these things – buildings, halls, exhibits, artworks – is called a *ScannableItem*. A ScannableItem is a generic *thing* that is uniquely identified, may *contain* other ScannableItems and has a number of attributes associated with it. The attributes of a ScannableItem include a unique ID, a type, a location, a short and a long textual description, as well as associated images, audio and video.

The content database also contains tables for configuration data: one table contains a record for each available language. For example, at venues in Kazakhstan, the primary languages are Kazakh, Russian, and English; whereas, in the United States, the primary languages would be English and Spanish. Another configuration table contains a record for each *level*. Levels allow for each piece of information to be customized for a certain audience. For example, the National Museum creates content customized for adults, students and small children. Each text attribute for a ScannableItem − name, short description, image caption − can be stored in any of the configured languages, and each longer textual attribute or audio can be stored in any combination of language and level. In an environment like the National Museum, with 3 languages and 3 levels, an object on display may have 9 available text descriptions.

Figure 2 below shows a partial entity-relationship (ER) diagram for the content database, specifically the ScannableItem table and the primary tables related to it.

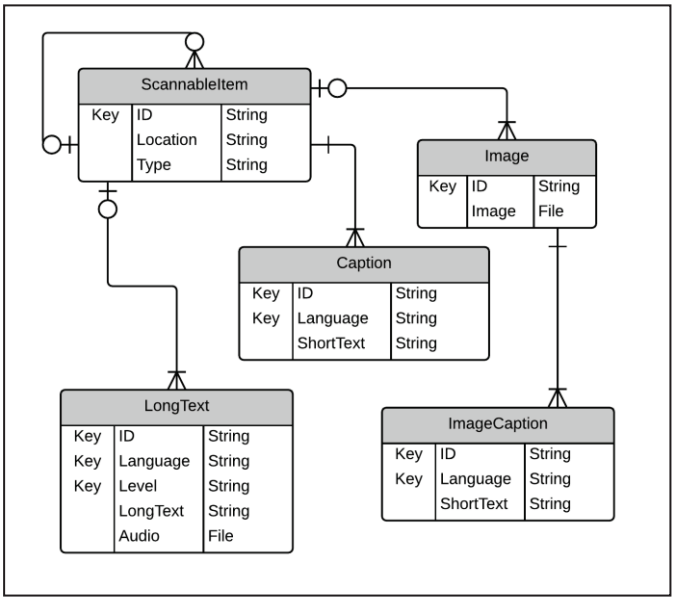


Fig. 2. A partial ER diagram for the content database.

In the current version of the framework, the content and usage database tables are stored together in one relational database. However, if a large amount of usage data were to be collected, it would be appropriate to keep that data in a companion non-relational database, such as MongoDB [9], designed for fast updates. In the current version, only the

ability to *like/favorite* an item is supported, so there are only two usage data tables: a user table and a favorites table that links users to items. Tables to store access history are planned. Usage data will be useful for users and venue managers to understand the relative popularity of the areas and items in the place.

IV. The Mobile Application

A. Location Determination and Item Identification

In any location-based information system, the method of determining the user's location is key. In many systems, there is just a single method: outdoor systems might use GPS, indoor systems might use RFID, Bluetooth or WiFi triangulation. For specific object identification (which of course also can determine location precisely), systems might use NFC, QR codes, or image recognition. Our framework allows for any one or a combination of these methods. Our current version includes the ability to identify areas or objects using Bluetooth beacons, NFC tags or QR codes, and we are adding the ability to use image recognition. For example, a specific item may be marked with QR code and an NFC tag, both of which resolve to the same item in the database. Similarly, the entrance to an exhibit may have an NFC tag, and a Bluetooth beacon may be located just inside. So a user of the system might scan the tag to get general information about the exhibit, or exhibit information might be pushed to the user automatically after she enters.

B. Android Mobile App

The mobile app is based on the Android platform. Its user interface is simple and uses the principles of Google's *Material design* [10]. The app's main screen consists of four panels:

- Home − basic information and latest news
- Favorites – list of the user's favorite items and areas
- Navigate – displays floorplans/maps
- More – includes settings, credits, links, etc.

Figure 3 shows the main app screen, with a current event displayed on the Home panel. The floating button launches the QR code reader, and when a code is recognized, the app switches to the item details screen. If an NFC tag is scanned, or a Bluetooth beacon is sensed, the app automatically starts the item details screen. The item details screen displays all the information available about a specific object or area. This information may include textual information, images, audio or video.

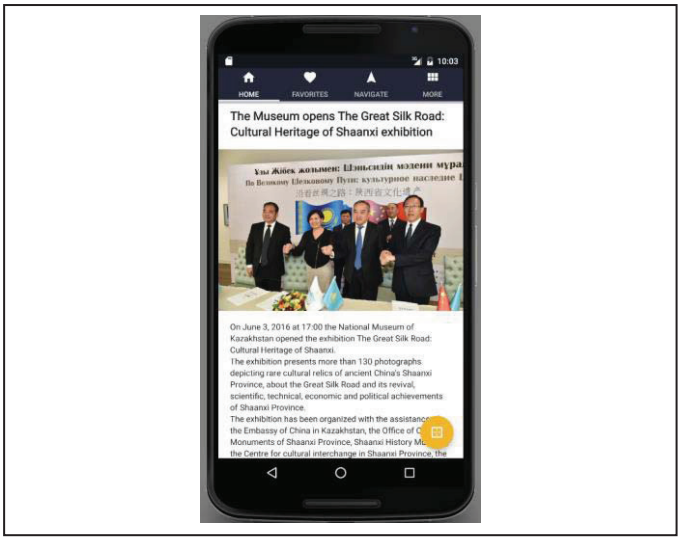


Fig. 3. The main screen of the Android mobile app (*content from* [1]).

C. Audio

An audio file can be stored for each area/object in the database together with its long text description, allowing for different audio by language and level. In this way the mobile app can use the same content as existing audio tours, and function in the same manner, but allow patrons to use their own mobile devices and headphones instead of the venue needing to own and maintain specialized devices and headsets.

Whenever a user enters an area or scans an object, the audio will be available for playback on the mobile app's item details screen. Furthermore, if the venue has not created special audio, the item details screen will allow the user to listen to the long description of an object using the mobile device's built-in text-to-speech capabilities. Although text-to-speech does not provide an ideal user experience, and the quality is not consistent across languages, it does at least provide an option for vision-impaired visitors, or those for whom reading on a small screen is uncomfortable.

D. Navigation

In the prototype version of the app, the navigation screen only displays maps/floorplans of the venue. However, the content database does contain the location of each area/object. A web interface has already been developed that allows a content manager to specify the location of an area/object by placing it on a map/floorplan, and to specify (draw) and store the navigable paths between areas/objects and between floors. A new navigation panel for the mobile app is being developed that makes use of the location and path data in the content database in order to allow the user to locate herself on the map, to search for items of interest and see their locations on the map, and to see the best path to an item of interest.

V. Web Client and APIs

The server-side components of the framework include a web client and a set of APIs. The web client gives content producers and managers a convenient way to enter, organize

and maintain their content. The APIs provide a way for the mobile app to query and update the content and usage data. Some of the current APIs are listed in Table I. The APIs return their results in JSON for easy handling by the mobile app.

TABLE I. APIS FOR CONTENT AND USAGE

Type	Address
GET	/api/scannables/{id}/{language}
GET	/api/scannables/{id}/{language}/{level}
GET	/api/languages
GET	/api/levels
POST	/api/like/{userId}/{scannableId}
DELETE	/api/unlike/{userId}/{scannableId}
GET	/api/favorites/{userId}
GET	/api/likenumber/{scannableId}
GET	/api/getnews/{language}

A section of the client interface for maintaining a ScannableItem is shown in Figure 4, illustrating the ability to add short descriptions of an area/object in multiple languages, as well as to enter the basic data associated with an item. Current functionality in the web application includes maintaining items and all their related data, manipulating maps and floorplans and specifying item locations and navigable paths, creating and editing the hierarchical relationships between areas and objects, and maintaining events and their related images, text and audio.

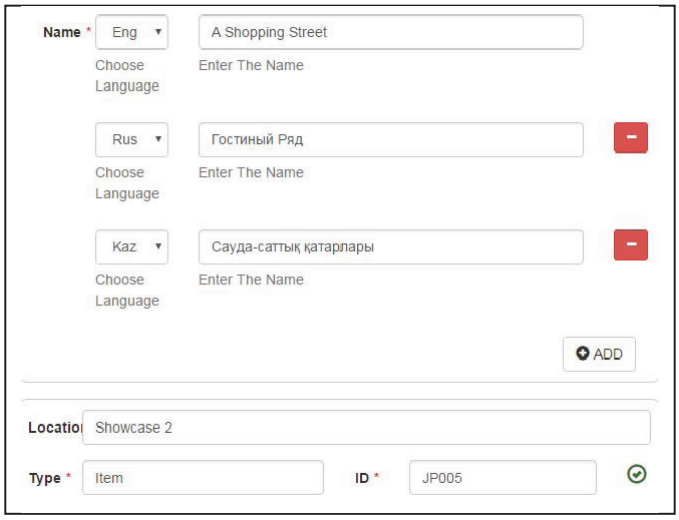


Fig. 4. A section of the client application interface.

VI. CONCLUSIONS AND FUTURE WORK

We have developed a general framework for mobile location-based information systems and implemented a fully-functional prototype version of all the components. The framework provides a simple, general capability for developing systems that are commonly used in museums, tourist sites, parks and shopping areas. It provides a comprehensive method for creating and displaying content customized to different audiences based on language and level. It allows for content to be presented in textual, audio, and image formats, and is easily extensible to allow for other media. It currently allows for location determination and object identification to be done using a combination of QR codes, NFC tags, and Bluetooth beacons; it is easily extensible to allow for GPS, image recognition and other methods. Future work on this project falls into two main categories: evaluation and further development.

A. Evaluation

We plan to use our framework to create a system for the National Museum of the Republic of Kazakhstan and install a prototype of the system in the museum for evaluation during the summer of 2016. We also hope to create and install another prototype at a museum in the U.S. during the same timeframe. These evaluation sites will help us improve the user interfaces and prioritize new features. We also are seeking a partner to develop a system using our framework for a very different type of venue, such as a shopping mall or botanical garden, in order to evaluate the generality of the framework.

B. Further Development

Some of the features that are currently under development include full-featured navigation, object identification using image recognition, and more comprehensive collection of usage data. We expect to begin development of an iOS version of the mobile app in autumn 2016. Future development will include personalization features based on analysis of usage data collected; for example, offering users suggestions for paths to take or areas to visit or producing post-visit content for the user. We should provide support for new types of content, such as augmented reality or location-based advertisements. We may add more social features beyond *liking*, such commenting on objects. Finally, many different types of mobile apps can be built to make use of the same content, such as games and quizzes.

A significant issue for the framework itself is that it assumes that the mobile device has a constant internet connection at all times, since content is retrieved on an as-needed basis. To be more general, an option must be available to intelligently pre-download content, and for the mobile app to be configured to work from that locally-stored data in an offline mode.

ACKNOWLEDGMENT

We would like to thank Kamilya Altynbekova, Saida Mussakhojeya, Madi Zhanbyrtayev, Shyngys Zhiyenbek and Korlan Zhumabekova for their contributions to current and future versions of the framework. We would also like to thank the many personnel from the National Museum of the Republic of Kazakhstan who have helped us with our requirements gathering.

REFERENCES

[1] National Museum of the Republic of Kazakhstan, http://nationalmuseum.kz, accessed June 16, 2016.

[2] Smithsonian Mobile, http://www.si.edu/apps/smithsonianmobile, accessed June 16, 2016.

[3] A. Zafeiropoulos, E. Solidakis, S. Zoi, N. Konstantinou, P. Stathopoulos, N. Mitrou, " Location based guidance services in a museum environment: deployment issues and a proposed architectural approach," in WINSYS, 7, pp. 217-24, 2007.

[4] E. Bruns, B. Brombach, T. Zeidler, O. Bimber, "Enabling mobile phones to support large-scale museum guidance," in *IEEE multimedia*, (2), pp. 16-25, 2007.

[5] U. Biader Ceipidor, C. M. Medaglia, V. Volpi, A. Moroni, S. Sposato, M. Carboni, A. Caridi, "NFC technology applied to touristic-cultural field: A case study on an Italian museum," in Near Field Communication (NFC), 2013 5th International Workshop on, pp. 1-6. IEEE, 2013.

[6] M. Schickler, R. Pryss, J. Schobel, M. Reichert, "An engine enabling location-based mobile augmented reality applications," in Web Information Systems and Technologies in Lecture Notes in Business Information Processing, Vol. 226, pp. 363-378, December 2015.

[7] M. Singhal, A. Shukla, "Implementation of location-based services in Android using GPS and web services," in IJCSI International Journal of Computer Science Issues, 9(1), pp. 237-242, 2012.

[8] J. Sieck, "Location based services and museum information systems," in Intelligent Systems, Modelling and Simulation (ISMS), 2012 Third International Conference on, pp. 663-666. IEEE, 2012.

[9] MongoDB, https://www.mongodb.com/, accessed June 16, 2016.

[10] Material Design, https://material.google.com/, accessed June 16, 2016.

Anomaly detection in network traffic using extreme learning machine

Yadigar Imamverdiyev[1], Lyudmila Sukhostat[2]
[1,2]Institute of Information Technology of Azerbaijan National Academy of Sciences
Baku, Azerbaijan
[1]yadigar@lan.ab.az, [2]lsuhostat@hotmail.com

***Abstract*—Intrusion detection systems are one of the most relevant security features against network attacks. Machine learning methods are used to analyze network traffic parameters for the presence of an attack signs. In this paper, extreme learning machine method is considered for intrusion detection in network traffic. The experimental results lead to the conclusion of practical significance of the proposed approach for attacks detection in network traffic.**

***Index Terms*—Anomaly detection, network attack, network traffic, extreme learning machine, NSL-KDD.**

I. Introduction

Network attacks are one of the causes of the abnormal phenomena observed in the work of the network equipment, as well as traffic transmission over the network. Anomalies of network traffic may result in incorrect operation of single channel or entire network segments, lead to a denial of service in this network equipment. Network attacks are constantly changing because attackers use individual approaches. It is also affected by changes in software and hardware.

Solution of anomaly detection problem is not trivial, since anomalies nature itself is changeable. Providing a comprehensive definition of abnormal or normal behavior in the context of a computer network is quite subtle [1, 2]. Another reason is that some anomaly detection methods require labels of normal and abnormal behaviors that are difficult to obtain [3, 4]. In addition, choose the right tool for anomalies detection is not easy. The intended tool may well be suitable for only one type of abnormality, but not in all [5]. Thus, when anomaly types are not known a priori, it is a very realistic assumption, that the choice of anomaly detection method is not trivial. In addition, the scale of the network is a problem. When it detects anomalies it is need to consider load balancing (distribution of the tasks implementation process between multiple network servers in order to increase the overall performance) and fault tolerance (the ability of the system to work in case of failure of its individual elements), especially with the growth of the existing network size [6, 7].

More information about anomalies detection can be obtained by machine learning methods. Theoretically, machine learning algorithms can achieve high productivity, i.e. can minimize the number of false alarms and maximize detection accuracy. However, it usually requires an infinite number of training samples (theoretically). In practice, this condition is impossible because of the limited computational capability and requiring real-time response.

In this paper, a method for the anomalies detection in network traffic based on the extreme learning machine (ELM) is considered.

II. Application of Machine Learning Methods to Intrusion Detection

The cyber security anomaly detection includes the detection of malicious activities. This approach consists of two phases: training and detection [8]. At the training phase, machine learning methods are used to create a normal traffic profile in the absence of an attack. At the detection phase, input events are labeled as an attack when the event records are different from the normal profile. Subsequently, anomaly detection can detect previously unknown attacks. However, the anomaly detection is complicated by high false alarm rate. In these cases, an attacker can use these anomalies to train an anomaly detection system, so that it will recognize the malicious patterns as normal.

All four possible anomaly detection metrics (true positive, false positive, false negative and true negative) are shown in Fig. 1 as a confusion matrix [9]. The problem is to minimize the false negative and false positive results.

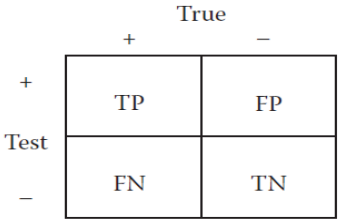


Fig.1. The confusion matrix for evaluation of the performance of machine learning methods.

The most commonly used performance evaluation metric for anomaly detection methods is the frequency rate of true positive results (TP) (also known as detection probability), which is the number of correctly classified intrusions. Another important indicator is the rate of false positive results (FP) (or a false alarm), which denotes the number of abnormal samples

classified as normal. The frequency of false negative results (FN) is the number of normal samples classified as an intrusion. The frequency of true negative results (TN) is the number of correctly classified intrusions. In addition, other parameters, such as receiver operating characteristic (ROC), are used to evaluate an intrusion detection system (IDS). ROC is the detection metric for performance evaluation, and is generated by plotting TP against FP at various thresholds. F-score (also known as F-measure) evaluates recognition system performance as a whole, combining the precision and recall.

III. Extreme Learning Machine

Extreme learning machine algorithm was proposed by Huang et al. in 2006 [10]. This algorithm is based on a single-layer feedforward neural network (SLFN). ELM also solves the classification and regression problems [11-14]. Consequently, ELM can be used to detect anomalies in network traffic. It is widely used in many researches [11-12, 15]. ELM has a high computation rate and does not require iterative parameter tuning for training.

In ELM model input data training is carried out within seconds or minutes, even for large data sets and complex applications, which cannot be achieved using conventional methods.

SLFN with L neurons in hidden layer and activation function $G(a_i, b_i, x)$ can be described as [10, 16]:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x), \quad (1)$$

where $\beta = [\beta_1, \beta_2, ..., \beta_L]^T$ is the matrix of output weights from the hidden layer to the output layer and $h(x) = [h_1(x), h_2(x), ..., h_L(x)]$ is a non-linear mapping space, which contains the output vector with respect to the input x.

The activation function $G(a, b, x)$ is a non-linear piecewise continuous function [10, 17]. In ELM model, hidden node parameters (a, b) are randomly generated, and it is therefore more effective than support vector machines or artificial neural networks [10, 13].

The output vector Y_j in matrix form is calculated as follows:

$$Y = H\beta, \quad (2)$$

where H is the matrix of output neurons' values of the hidden layer and is given by

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}, \quad (3)$$

where N is the number of input samples.

Y is the target output matrix:

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{1N} & \cdots & y_{Nm} \end{bmatrix}. \quad (4)$$

Network training consists in calculation of output weights matrix of the hidden layer by

$$\beta = H^+ Y, \quad (5)$$

where $H^+ = (H^T H)^{-1} H^T$ is the pseudo-inverse matrix (Moore-Penrose matrix) for the case of non-singular matrix H.

From matrix theory [18] it is known that the Eq. 5 is the best approximate solution of Eq. 2 by the least squares method, which avoids local minima during network training.

When the neural network is trained, i.e., according to the Eq. 5 found the output weights matrix β of the hidden layer, the output vector is calculated based on Eq. 2. And the input features vector (which is not contained in the training set) is applied to the network. Then the number of class to which the input vector belongs is calculated.

IV. Database description

The attack samples of transport and application levels can use the advanced database of signatures [19] built on the basis of KDD-99 database on the initiative of the American Association for the Defense Advanced Research Projects Agency (DARPA) [20]. A dataset of connections was collected to conduct the research in the field of intrusion detection, which covers a wide range of intrusions simulated in a medium that mimics the US Air Force network. NSL-KDD includes 42 attributes and has labels for each specific type of attack. Each vector is designated either as a normal or as an attack. Deviation from "normal behavior" is considered as an attack [3].

All attacks are divided into four groups [21]: DoS (Denial of Service Attack), R2L (Remote to Local Attack), U2R (Users to Root Attack), and Probe (Probing Attack).

V. Experimental results

Operation of the network, trained by ELM using Matlab package, is compared with various activation functions (radial basis activation function (RBF), triangular basis activation function (TriBas), and Gaussian activation function) and is tested using NSL-KDD database (Table I). The hidden layer has 30 neutrons with different activation functions. This number of neurons provides a high classification accuracy on the test sample and acceptable accuracy on the training set for the network structure and training method.

In this paper Gaussian activation function for different parameter values of σ and μ was proposed [22]. The best accuracy in detecting the DoS, U2R and R2L attacks was achieved using Gaussian activation function with $\sigma = 0.1$ and $\mu = 4$ (DoS – 86.89%, U2R – 99.99%, and R2L – 99.94%). In

the case of Probe attacks high accuracy (99.06%) was obtained for the Gaussian activation function with $\sigma = 0.2$ and $\mu = 0$.

It should be noted that the neural network training based on the ELM approach is superior to other methods, both in velocity and in classification accuracy for training and testing samples.

TABLE I. DETECTION ACCURACY COMPARISON FOR DIFFERENT ACTIVATION FUNCTIONS

	DoS	Probe	U2R	R2L
RBF	80.01%	91.79%	99.88%	**99.94%**
TriBas	81.60%	95.62%	99.65%	99.57%
Gaussian (σ =0.2, μ =0)	84.26%	**99.06%**	99.95%	99.70%
Gaussian (σ =0.1, μ =4)	**86.89%**	90.01%	**99,99%**	**99,94%**

The considered classifier based on ELM method provides an acceptable quality of attacks classification and a high performance, making it an attractive solution for the real-time intrusion detection systems. Thus, the proposed approach can be considered in the construction of IDS system.

CONCLUSION

Network attacks are one of the causes of the abnormal phenomena observed in the network equipment operation, as well as traffic transmission over the network. Anomaly detection is possible by means of machine learning methods to improve the reliability of the existing systems. In this paper, a method for the anomalies detection in network traffic based on ELM was proposed. An experimental database NSL-KDD was described. A description of the main types of attacks was given.

In the view of the above mentioned, it is possible to select the following research directions in the field of anomaly detection:

- increasing the reliability and accuracy of intrusion detection methods, and anomalies (reduction of first- and second-order error levels, particularly with regard to previously unobserved impacts of information technology);
- increasing the share of corrective processes that do not require human expert intervention, which reduces the level of heuristic decision-making and allows qualitatively improve the response time to attack (for example, by automatically generating signatures for new malware in a few minutes after the confirmation of its abnormally rapid spread on the network);
- resistance to the new technologies used by an attacker in order to conceal the fact of the intrusion, active influence on the very IDS by creating denial-of-service conditions, or the generation of excessive flow of false positives, which makes it impossible to use.

REFERENCES

[1] M. Xie, S. Han, B. Tian, and S. Parvin, "Anomaly detection in wireless sensor networks: A survey," Journal of Network and Computer Applications, vol. 34, no. 4, 2011, pp. 1302–1325.

[2] H. Nallaivarothayan, D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of different features and learning models for anomalous event detection," in Proc. of the 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA'13), Hobart, Australia, November 2013, pp. 1–8.

[3] J.J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: A review," Computers & Security, vol. 30, no. 6–7, 2011, pp. 353–375.

[4] U. Fiorea, F. Palmierib, A. Castiglionec, and A. D. Santis, "Network anomaly detection with the restricted Boltzmann machine," Neurocomputing, vol. 122, 2013, pp. 13–23.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, vol. 41, no. 3, pp. 15:1–15:58, July 2009. [Online]. Available: http://doi.acm.org/10.1145/1541880.1541882

[6] D. Difallah, P. Cudre-Mauroux, and S. McKenna, "Scalable anomaly detection for smart city infrastructure networks," IEEE Internet Computing, vol. 17, no. 6, 2013, pp. 39–47.

[7] E. Anceaume, Y. Busnel, E. Le Merrer, R. Ludinard, J. Marchand, and B. Sericola, "Anomaly characterization in large scale networks," in Proc. of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'14), Atlanta, USA. IEEE, 2014, pp. 68–79.

[8] D. Bhattacharyya, J. Kalita, "Network anomaly detection: a machine learning perspective," 2013, 336 p.

[9] Sumeet Dua and Xian Du, "Data Mining And Machine Learning In Cybersecurity," Auerbach Publications, 2011, 256 p.

[10] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme Learning Machine: Theory and Applications," Neurocomputing, vol. 70, 2006, pp. 489–501.

[11] N. Acharya, N.A. Shrivastava, B. Panigrahi, and U. Mohanty, "Development of an artificial neural network based multimodel ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine," Climate Dynamics, vol. 43, no. 5, 2014, pp. 1303–1310.

[12] R.C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern," Australia Atmospheric Research, vol. 153, 2015, pp. 512–525.

[13] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines," A review Neural Networks, vol. 61, 2015, pp. 32–48.

[14] J. Sánchez-Monedero, S. Salcedo-Sanz, P. Gutiérrez, C. Casanova-Mateo, and C. Hervás-Martínez, "Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier," Engineering Applications of Artificial Intelligence, vol. 34, 2014, pp. 199–207.

[15] A. Belayneh and J. Adamowski, "Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression," Applied Computational Intelligence and Soft Computing, 2012, pp. 1–13.

[16] M. Şahin, Y. Kaya, M. Uyar, and S. Yıldırım, "Application of extreme learning machine for estimating solar radiation from

satellite data," International Journal of Energy Research, vol. 38, 2014, pp. 205–212.

[17] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," Neurocomputing, vol. 70, 2007, pp. 3056–3062.

[18] F. Zhang, "Matrix Theory: Basic Results and Techniques," 2011, 399 p.

[19] P. Aggarwal, S.K. Sharma, "Analysis of KDD dataset attributes-class wise for intrusion detection," Procedia Computer Science, vol. 57, 2015, pp. 842–851.

[20] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, 2000, pp. 262–294.

[21] "NSL-KDD data set for network-based intrusion detection systems". November 2014. [Online]. Available: http://nsl.cs.unb.ca/NSL-KDD/

[22] R. Bracewell, "The Fourier Transform and Its Applications," New York: McGraw-Hill, 1999, pp. 98–101.

The Technology for the Automatic Formation of the Personal Digital Voice Pattern

R.R. Mussabayev, Ye. N. Amirgaliyev,
Laboratory of analysis and modeling of informational processes
Institute of Information and Computational Technologies
Almaty, Kazakhstan
rustam@ipic.kz

A.T.Tairova, T.R.Mussabayev, K.Ch. Koibagarov
Uniline Group LLP
Almaty, Kazakhstan
mail@uniline.kz

***Abstract*— The main purpose of the given paper is to describe the concept and the development process of the new open technology for the automatic formation of the personal digital voice pattern for the solution of the following problems on its basis: computer speech synthesis, continuous speech recognition and identification of a person by his voice (three major tasks in the field of speech technologies).**

***Index Terms*— Personal digital voice pattern (PDVP), speech cloning, speech synthesis, speaker identification.**

I. Introduction

The precondition for the development of the given technology was the increasing necessity of users of modern mobile devices and stationary computers in better quality and productivity of modern speech technologies [1]. Thus, one of few possible ways of speech technologies evolution in the direction of quality improvement is the creation of the new open standardized technology for the exact account of user individual voice features that will allow providing more flexible and adaptive personification of speech systems. The more speech information about a user is processed by the speech system in the process of work, the more exact and complete is the voice portrait of the person, and the more possibilities it has for adapting, in order to provide more qualitative results of synthesis, recognition and identification.

The previous scientific investigations and technical developments were basically concentrated on separate aspects in the process of different tasks solution in the field of speech technologies. New models, algorithms and methods in this field were offered, and the existing ones were investigated. The essential difference of the given technology approach from the previous ones is its complex and generalizing character directed at the creation of the unified technology for the solution of the primary tasks in the field of speech technologies due to more flexible adaptation to the specific features of a separate person's voice.

In the process of the PDVP technology development the following theories, methods and algorithms are integrated to the one monolithic technology:

- The speech formation theory [2-3];
- Parameterization of speech signals [4-5];
- Computer speech synthesis [6];
- Automatic speech recognition [7-8];
- Identification of a person by his voice [9];
- Speech enhancement [10];
- Speech cloning [11];
- Automatic segmentation of speech signals [12];
- Formation of speech databases [13];
- Application of additional biometric sensors jointly with a microphone (multisensory speech signal processing) [7-9];
- Solution of speech inverse tasks [14-16].

The investigations of Belarus scientists from the Laboratory of Speech Recognition and Synthesis of the Incorporated Institute of Informatics (Minsk), which are directed at the development of speech cloning system [11], are the closest to the given technology on its subject. The research of the Belarus scientists differs from the given technology, as their cloning of speech characteristics is only used in the course of speech synthesis, and other speech technologies are not involved.

Also for the given technology development, the works which are carried out under the supervision of V. N Sorokin at the Institute of Problems of Information Transfer of the Russian Academy of Sciences (Russia, Moscow) are of great value. For all types of the speech sounds they managed to obtain the solutions of speech inverse tasks with the inaccuracy comparable with the measurement inaccuracy [14-16].

II. The PDVP technology conception

The personal digital voice pattern (PDVP) represents a set of parameter values of various mathematical models and algorithms which describe the individual speech and the voice characteristics of a certain person with the given degree of accuracy in the compact and unified form. The identification of optimal values of model parameters is carried out by the algorithmic way with the use of a digital record of a natural speech signal pronounced by this person. The results of algorithmic estimations of various measures of similarity between the initial speech signal and the signal synthesized with the use of the PDVP serve as optimality criteria. The PDVP provide the possibilities for more precisely considering of individual voice characteristics of users, and its using ensured considerable improving the qualitative characteristics

of computer synthesis, speech recognition, and speaker identification. Also the important PDVP property is standardization and unification of the data on individual speech and voice characteristics of a certain person that allow to use the common and the unified data set while solving all the above-mentioned problems.

Potentially, the PDVP can obtain the social function similar to the function of a digital photo: users can keep in electronic form the voice portraits of people dear to them, whom they hold in remembrance, so as to have a possibility to sound any textual information by their voice. Potentially, the realization of the technology for the creation of the personal digital voice portrait can generate the new independent market of goods and services based on the given technology on international scale.

III. THE PDVP TECHNOLOGY DEVELOPMENT PROCESS

For PDVP technology realization the wide range of tasks are solved. At first the formation of the set of specialized mathematical models and algorithms for parameterization of various aspects of speech-formation process is conducted. These aspects are included: vocal tract configuration, vocal excitation signal, resulting speech signal, dynamics of intonation processes, etc. Also the research of applicability of different types of biometric sensors in addition to a microphone was conducted. These additional biometric sensors are needed for direct high-precision registration of vocal excitation signals and also for monitoring of dynamics of vocal tract configuration change in the speech-formation process (multisensory speech processing) [7-9].

The additional research is carried out for the purpose of choosing the optimum composition and the number of parameters of models and algorithms included into the formed set. It was important to identify the demanded accuracy of these parameters definition for providing the qualitative solution of the speech synthesis, speech recognition and speaker identification tasks. As the result of this stage of work was the PDVP standard which defines the composition of the used models and their parameters, the admissible ranges of the parameter values changing, and also the format of representation, storage and data exchange.

For the automation of the PDVP formation process the set of tools in the form of the information system were developed. In the process of PDVP formation the long digital recording of a natural speech signal pronounced by the person, whose personal voice digital portrait is used.

At the next stage the construction of the algorithms for the solution of the computer speech synthesis, recognition and identification tasks on the basis of the PDVP technology was performed with software realization of the corresponding information systems.

At this time, the documentation is prepared for detail describing the of the PDVP technology. All these information will be available on the web-portal, where all the technologies and the information resources developed in the framework of the given project will be placed in the source codes and open access.

IV. THE FUNDAMENTAL PDVP SUBTECHNOLOGIES, ALGORITHMS AND METHODS

A. Mathematical modeling of speech formation processes

The significance of the given project consists in the research of the existing mathematical models and algorithms and also in the development of new ones for the solution of the given task for the development of the new open technology for the PDVP automatic formation.

For the PDVP formation the results of the speech inverse task solution are used. These are the tasks of definition of a speech tract form, anatomic and articulatory parameters, commands of articulation control and also the form of vocal excitation impulses by the parameters measured on a speech signal. These problems are incorrect, as they essentially suppose ambiguous and frequently unstable solutions. However, there are preliminary results which show that it is possible to obtain the solution of speech inverse tasks with the inaccuracy comparable with the measurement inaccuracy. The bases of these solutions are mathematical models of speech formation which are used in Tikhonov's variation method along with physiologically and anatomically adequate restrictions and specific regularizers [14-16]. The constructive algorithms based on the solution of inverse tasks provide the essentially new approaches to the development of speech technologies. The preliminary experiments show that the quality of speech articulatory synthesis is comparable with the quality of natural speech; the compression of a speech signal to speed less than 2 Kbit/sec is possible with the full preservation of speech legibility and individuality, and the total probability of an error of the speaker verification by his voice can be below 0.01%. Also new possibilities in the development of speech automatic recognition systems are created.

In the framework of the PDVP project, the new methods for the solution of inverse speech tasks are developed, and the results of these solutions are used to increase the efficiency of the PDVP formation process.

In the composition of PDVP in addition to inverse task solution the parameters of the following models, algorithms and methods are included:

- Fujisaki intonational model which allows to consider all speaker dependant aspects of speech prosody [17].
- Sinusoidal model used for performing of harmonic analysis of separate speech segments and their subsequent compression [4].
- Speech segments clustering algorithms [18].

B. Using the throat microphone (laryngophone) as the additional modality for direct voice excitation signal registration and further analysis

In the process of the PDVP technology development, the research of applicability of different type biometric sensors in addition to a microphone is conducted. These sensors are used for the direct high-precision registration of vocal excitation signals and also for monitoring the dynamics of the speech tract configuration change during the speech formation process.

The signal of human vocal fold oscillations can be used as independent modality during the speech segmentation and recognition processes. It is well known that the vocal fold is the source of a voiced quasi-periodic excitation. The role of the vocal fold in the process of speech signal generation can be demonstrated on the basis of an aggregate model of speech production (Fig. 1). In accordance with this model, vocal fold vibrations are modeled by the sinusoidal impulse generator whose spectrum is given *P(f)*. Noise oscillations are modeled by the white-noise generator with spectrum *N(f)*. While mixing vocal and noise oscillations, their amplitude levels can be matched by the following parameters: *v* – for the voice generator, *u* – for the noise generator. Then output signals from both generators are summed and delivered to the input of block that models the vocal tract where their spectral shape is transformed in accordance with the transfer function *H(f)*. The emission characteristics of the lips are modeled by *L(f)*.

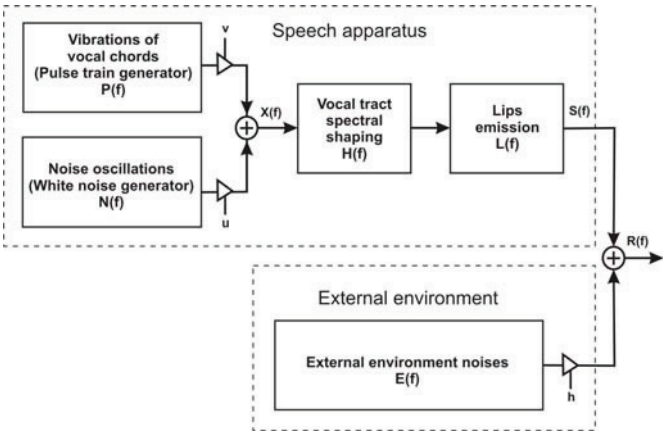


Fig. 1. Aggregate model of speech production

Hence the spectrum of the speech signal *S(f)* is defined as follows:

$$S(f) = (v \cdot P(f) + u \cdot N(f)) \cdot H(f) \cdot L(f) = \\ = X(f) \cdot H(f) \cdot L(f) \quad (1)$$

If external acoustic noise is taken into consideration in the model, then spectrum of the resulting signal will be:

$$R(f) = S(f) + h \cdot E(f) \quad (2)$$

For the impact on the speech signal we have the following parameters of the speech production model: a mixture of the internal voice and noise oscillations (determined by *v* and *u*), the fundamental frequency (determined by *P(f)*), the spectral form (specified by *H(f)*), the amplitude of signal (depending on *v, u, h*). While solving speech recognition tasks by the original speech signal *R(f),* there was an attempt to restore the parameters described above, and then to transfer them to the input of the recognizer. Due to objective circumstances (imperfect algorithms, the presence of background noise (*E(f)*, etc.), the solution of the problem of restoring the model parameters may contain some inaccuracies which can usually become one of the reasons of recognition quality deterioration.

The analysis of the used model (1), which is considered to be the most adequately reflecting the objectively occurring physical speech formation processes [2], allows to make the single conclusion that the signal of the voice excitation which is presented by the vocal fold oscillations plays the primary role in the speech formation process. Since the speech signal itself is the result of the complex transformations of the voice excitation signal with the further adding of the noise component, and we don't often know the parameters of the vocal tract where the given transformations (amplification and suppression of resonant frequencies) are realized, then it is quite reasonable to register directly the excitation signal with the use of a throat microphone. The given model substantiates the necessity of using a throat microphone for direct registration of the voice excitation signal, which in its turn allows getting the additional information for the realization of more accurate segmentation in comparison with the segmentation using only one speech signal. It is rather difficult to restore the source excitation signal directly from the speech signal. For this it is necessary to know the exact physical parameters of the speaker's vocal tract.

On the subsequent stages the voice excitation signal is passing through the different types of analysis. Particularly using the voice excitation signal the succeeding segmentation of source speech signal was conducted. The segmentation process was performed synchronously with the fundamental frequency of the speech signal. This level of segmentation is subphonemic. Also the laringophone signal was used for calculation of the new types of informative features which are allowed to improve the efficiency of the segmentation and recognition processes.

C. Automatic speech segmentation algorithm

While solving different speech synthesis and recognition tasks, there is a necessity of using speech segmentation algorithms [12]. The process of speech signal segmentation can be performed by applying the different algorithms at different levels: sub-phonemic, phonemic, syllabic, the word level, syntagmatic, etc. The accuracy and stability of the segmentation results have significant influence on efficiency of the speech recognition process, and also on naturalness of synthesized speech.

The effectiveness of segmentation algorithms can be evaluated by the following criteria: precision of the boundary determination, robustness, stability, computation speed. The main advantage of automatic segmentation algorithms in comparison with manual segmentation is the ability to achieve a high degree of segmentation results reproducibility. It means that the all boundaries between the two typical segments in different places will be detected in the same way throughout the whole signal.

In the given work the several new approaches for solving the problem of speech signal segmentation by applying the additional modality - a signal of human vocal fold oscillations is used. In this case, the term "modality" means an existence

mode of an object or behavior of an occurrence or as a way of understanding of an object or an event.

The automatic segmentation algorithms are used in the PDVP building process where, at the first stage, a continuous speech signal coming to the input is segmented, and then the process of multilevel classification and recognition of the obtained speech segment sequence is executed. These algorithms are also used for the formation of databases of speech segments that are directly used as patterns in the parameterized form during the recognition and synthesis processes.

During the process of PDVP formation the special algorithm is also used for the automatic speech signal segmentation based on the use of changing dynamics analysis of the harmonic features of speech signals [4]. They are sequentially calculated for each period of the segmented speech signal in accordance with its fundamental frequency (pitch synchronous). The analysis of the used harmonic features properties shows the possibility of their use as additional informative features for the phonemes differentiation in the speech recognition algorithms (Fig. 2).

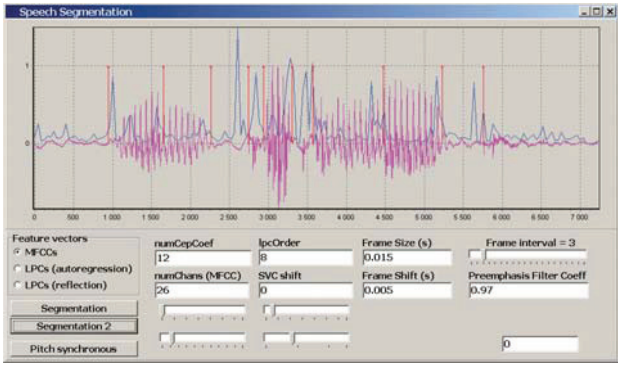


Fig. 2. The software unit with implemented algorithms of automatic segmentation of speech signal on the sub-word units. On this figure using the arrows the segment borders are marked.

D. Automatic tuning of the recognition system to individual speaker characteristics using PDVP

The important property of the automatic speech recognition system is the recognition accuracy. The recognition accuracy depends on the set of different factors. One of these factors is the ability of the system to recognize speech of different speakers. The majority of algorithms of recognition are constructed by the principle of comparison of a distinguished speech segment with reference segments of a database. Therefore those speakers will have the best results of recognition, whose speech characteristics correspond to speech characteristics of reference speech segments. The two ways of improvement of recognition quality are possible.

The first way consists in increasing of quantity of accessible reference signals in a database. Thus reference signals should correspond to various speakers with various speech characteristics. At the realization of the given requirement there are certain complexities: it is inconvenient to carry out speech recording of dozens of speakers with various voice characteristics. It is caused by the fact that the voice of any person has quite unique characteristics. After increasing of the database size the laboriousness of recognition also increases.

The second possible variant for the solution of the given problem is the development of the specialized algorithms for automatic adoption of the recognition system for the specific features of the speaker using PDVP. The given direction has been chosen in the process of the realization speech recognition system on the basis of this technology.

The investigation and realization of the following algorithms were made for the solution of the given task:

1. The algorithms for the estimation of the voice specific features (average values and temporal characteristics of formants, durations of phonemes, intonation contours, model parameters, etc.)

2. The algorithm of spectral transformation of the one speaker signals to the signals of another speaker on the basis PDVP;

3. The algorithm of fast reconfiguration of the recognition system from the voice of one speaker to the voice of another one.

The software realization of the given algorithms in the structure of the speech recognition system has allowed to increase the accuracy of the recognition with simultaneously decreasing the volume of speech databases.

E. Speech synthesis personification

The use of the PDVP technology in systems of speech computer synthesis which allow carrying out the personification of synthesis results, i.e. speech synthesis will be carried out by the voice of the person whose digital voice portrait was chosen for synthesis. The given aspect is promote the substantial increase of the synthesis result quality, because in the speech synthesis process the preference will be given basically to the high-quality PDVP, which provide the most natural sounding of the synthesized voice, while this voice itself is pleasantly perceived by ear. Thus, having the specialized software, every user of speech computer synthesis systems can create his own PDVP and use them for artificial speech synthesis. Users can carry out the PDVP exchange, choose for themselves the most pleasantly sounding voices, keep the PDVP of people dear to them. Most likely, of the greatest popularity will be the PDVP of celebrities (actors, singers, politicians), and also of people possessing unique voices (for example, professional announcers). Potentially, the PDVP can obtain the social function similar to the function of a digital photo: users can keep in electronic form the voice portraits of people dear to them, whom they hold in remembrance, so that to have a possibility to sound any textual information by their voice. On the basis of the given technology, the PDVP market presented in the form of specialized web-sites can be generated, where PDVP placing, exchange and sale are carried out.

F. Multilinguality, multiplatformness and open source

The given PDVP technology is developed on the principles of open systems. All the project source codes will be in open access. In the process of the given technology development, the principle of multiplatformness for providing the maximum independence of software and technical platforms is used. On the basis of the given technology, the applications of computer speech cloning and synthesis, automatic speech recognition and identification of a person by his voice for various platforms: Windows, Linux, Android, iOS are developed. Also the special emphasis will be made on the developed technology adaptation for mobile platforms.

The next important fundamental principle of the given project is multilingualism. The technology is developed, so that it would be possible to provide the support of personalized speech synthesis in most of the existing natural languages with minimum expenses, only at the setting level. In particular, in the course of the project realization, the synthesis support in the Kazakh, Russian and English languages with the possibility of further system expanding to other languages will be provided. For the system expansion to other languages, the resources of the international community of developers who will join the given project as its popularity increases will be used. All multilingual resources for the given system, and also its source codes will be consolidated in open access at the project official website.

V. Conclusion

In the framework of the PDVP technology, the question of unification of user individual speech characteristics within the limits of the solution of three primary tasks in the field of speech technologies is investigated for the first time: computer speech synthesis, continuous speech recognition and speaker identification. The developed technology is promoting the increase in the quality of speech human-machine interfaces.

In the course of the PDVP technology realization of the following main results are obtained:

1. The technology for the automatic formation of the digital voice pattern of a person including the following elements: mathematical models and algorithms, the PDVP standard, the information system for the PDVP automatic formation.

2. The PDVP based software for: computer speech synthesis, continuous speech recognition and automatic identification of a person by his voice

3. The databases of different PDVPs balanced by the properties of the speakers.

On the current development stage we finalizing the standardization of the PDVP technology and preparing it for open source availability. Also now we are work on improving the quality and effectiveness of PDVP based speech technologies.

Acknowledgment

This work is supported by the Science Committee of RK, under grant 0115PK00552 and grant 0115PK02741.

References

[1] S. Furui, "Fundamental technologies in modern speech recognition", IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 16-17, 2012.

[2] Fant G. "The Acoustic Theory of Speech Production", The Hague: Mouton, 1960.

[3] Sorokin V. N. The theory of speech-formation. Moscow, "Radio and communication", 1985. - 312 p.

[4] McAulay R. J., Quatieri T. F. Speech Processing Based on a Sinusoidal Model, The Lincoln Laboratory Journal, N.1 (2), 1988, pp.153-167.

[5] Vintsuk T.K. Analysis, recognition and interpretation of speech signals. - Kiev: Naukova dumka, 1987. - 264 p.

[6] Taylor P. Text-to-Speech Synthesis, Cambridge University Press, Cambridge, 2007, 597 p.

[7] Erzin E. "Improving throat microphone speech recognition by joint analysis of throat and acoustic microphone recordings", Audio, Speech, and Language Processing, IEEE Transactions on, V.17, N.7, 2009, pp.1316-1324.

[8] Graciarena M., Franco H., Sonmez K., and Bratt H. Combining standard and throat microphones for robust speech recognition, IEEE Signal Processing Letters, V.10, N.3, 2003, pp.72-74.

[9] Mubeen N., Shahina A., Nayeemulla A., Vinoth G. "Combining spectral features of standard and throat microphones for speaker identification", Int. Conf. on Recent Trends In Information Technology (ICRTIT), 2012, pp.119-122, 19-21 April 2012.

[10] Yegnanarayana B., Mahadeva Prasanna S. and Sreenivasa Rao K. "Speech enhancement using excitation source information", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP ' 02), V.1, 2002, pp.541-544.

[11] Lobanov B. M, Tsirulnik L.I. Computer speech synthesis and cloning. Minsk, «Belarus science», 2008. - 344 p.

[12] Sharma M. and Mammone R.J. "" Blind "speech segmentation: automatic segmentation of speech without linguistic knowledge", in Proc. ICSLP, 1996.

[13] Demenko G., Grocholewski S., Klessa K., Ogorkiewicz J., Wagner A., Lange M., Sledzinski D., and Cylwik N. "LVCSR Speech Database - JURISDIC," in Signal Processing Algorithms, Architectures, Arrangements, and Applications (SPA), 2008, pp. 67 - 72.

[14] Sorokin V., Leonov A., Trushkin A. Estimation of stability and accuracy of inverse problem solution for the vocal tract, Speech Communication V. 30, N1, 2000, pp. 55-74.

[15] Sorokin V., A.S. Leonov. "Inverse problem for the vocal tract: Identification of control forces from articulatory movements", Pattern Recognition and Image Analysis V. 10, 2000, pp.110-126.

[16] Sorokin V., Tananykin A. Initial conditions in the problem of voice recognition. Journal of Communications Technology and Electronics V. 55, N12, 2010, pp. 1542-1549.

[17] H. Mixdorf, "A novel approach to the fully automatic extraction of Fujisaki model parameters", Proceedings of ICASSP, pp. 1281-1284, 2000.

[18] H. Kamper, A. Jansen, S. King and S. J. Goldwater, "Unsupervised lexical clustering of speech segments using fixed-dimensional acoustic embeddings", Proc. SLT, 2014.

Comparison of QoS Architectures for VoIP Traffics in IPv4 and IPv6

Hayder AL-Fayyadh
Rifai Education Department
Ministry of Education
Dhi-Qar, Iraq
haderfayad@gmail.com

Murat Koyuncu
Department of Information Systems Engineering
Atilim University
Ankara, Turkey
mkoyuncu@atilim.edu.tr

***Abstract*—The Internet has become an integral part of human`s life with an ever increasing number of applications and services that it provides to its customers. Both the numbers of Internet users and services provided by the Internet continue to rise tremendously. Quality of Service (QoS) is the ability of a network to provide better services for the users to divide the bandwidth according to the needs of different applications. With an increase in demand for multimedia applications like video streaming, VoIP and video conferencing, larger bandwidth requirement increases drastically since such applications are very sensitive to delay, packet loss, and jitter. IPv6 has new features which are designed to improve QoS supported by IPv4, as well as other improvements. This paper gives a comparative analysis of QoS architectures, namely Integrated Services (IntServ) and Differentiated Services (DiffServ), which are examined to see their effects for Voice over IP (VoIP) in IPv4 and IPv6 environments. For comparisons, OPNET Modeler is used for network modeling and testing. The obtained test results show the importance of IPv6 to support multimedia applications on the Internet.**

***Index Terms* — QoS, IPv4, IPv6, IntSer, DiffServ, OPNET**

I. INTRODUCTION

There is an increasing demand for multimedia applications such as videoconferencing, VoIP and video streaming which cause serious data congestion on the IP networks. In recent years, the Internet has become the communication environment not only for social and human interaction but also for business relations as well as to support education in universities and schools. In addition, technological developments have led to creation of enormous new multimedia contents such as 3D videos, interactive environments, network gaming, virtual worlds which require larger bandwidths to support such applications [1].

With the developments of multimedia applications over IP networks, the bandwidth consumption has become a serious problem between the Internet service providers and Internet users. Since the bandwidth is a precious resource in the networks, the IPv4 and IPv6 protocols, which are the main two network layer protocols, provide Quality of Service (QoS) for the applications to a certain extent. Here, QoS represents providing the requirements of a network application by managing the delay, delay variation (jitter), bandwidth, and packet loss parameters [2].

In IPv4, QoS is provided by Type of Service (ToS) Field (also known as Differentiated Service field in the latest update) in IP header. The IPv4 ToS field is an 8 bit field with the first 6 bits known as the DiffServ Code Points (DSCP). The datagram's priority is defined with these precedence bits. It is a field in an IP packet that enables different levels of service to be assigned to network traffic [3].

IPv6, which is the successor of IPv4, implements QoS with the assistance of two fields: flow label and traffic class [4]. The 20 bit flow label field is designed to provide specific handling for a data flow. The flow is a unique combination of non-zero flow label and source address. The IPv6 routers provide service to the packets that belong to a particular flow in a similar way. The traffic class field is represented in 8 bits and is used for prioritizing the data packets similar to the TOS field of IPv4. The first 6 bits are used for classification of various packets [3].

The concept of Quality of Service (QoS) refers to the ability to provide different services to network traffic of different classes. The ultimate goal of QoS is to provide better network services with dedicated bandwidth, controlled jitter and latency, and improved loss characteristics [2]. QoS is the ability to ensure the delivery of critical data flows; in other words, to ensure various performance criteria that determine the level of satisfaction of a network application. In addition, it offers an ability to define the attributes of the services provided, both qualitatively and quantitatively. Assurance of QoS is important if the network capacity is not enough for all applications, especially for multimedia applications such as real-time VoIP, online games and IP-TV, since these applications are sensitive to delay and often require fixed bit rate in contrast to other applications such as HTTP, FTP and E-Mail which are not sensitive to delay.

QoS architectures which are designed by the Internet Engineering Task Force (IETF), namely integrated services (IntServ) and differentiated services (DiffServ), provide end-to-end QoS in an IP network. These architectures have some QoS parameters such as bandwidth, latency and jitter which are under the control of service providers, and can be

predicted and matched with the needs of the applications. The IntServ architecture provides a guaranteed end-to-end service through negotiations [5]. Applications ask for the level of service required to operate and rely on QoS mechanism provided by the network before beginning of the transmission. Applications do not send traffic before receiving a sign that the network is able to receive the burden to be loaded by the application and to provide the requested QoS. The other QoS model is the DiffServ architecture which provides a set of classification and queuing mechanisms to protocols or applications [3, 6]. DiffServ relies on the ability to classify packets of different types at edge routers or hosts. Network traffic can be classified based on network addresses, protocols and ports, the ingress interface, or other parameters as long as it is supported by a standard access lists or extended access lists. Then, the routers are expected to give priority to specific classes depending on the application requirement.

The main contribution of this paper is to provide a comprehensive analysis for QoS architecture's performances in IPv4 and IPv6 networks for VoIP applications. IPv6 is an emerging technology and many studies evaluate its performance when it is used on traditional IP networks. Various studies compare IPv4 and IPv6 performances for different traffic types such as FTP and HTTP. However, there are only a few studies evaluating the performance of VoIP, especially real-time, applications with other traffics on both IPv4 and IPv6 networks.

The paper is organized as follows: In Section II, we provide an overview of QoS architectures. Section III explains the test environment. Obtained test results are given in Section IV with discussions. Finally, Section V concludes the paper.

II. QoS Architectures

The IETF has designed two architectures for delivering end-to-end QoS in an IP network. These architectures are IntServ and DiffServ [7] which are explained in the following sub-sections.

A. Integrated Services (IntServ)

IntServ is a communication model for data networks designed and developed by IETF to meet a wide range of QoS requirements of various devices and applications in a computer network [5]. In this architecture, network applications make a request in advance about QoS requirements that they need. The request is usually sent by signals in the communication process to form a flow as a type of virtual circuit between the transmitter and receiver including also all routers along the path. Resource Reservation Protocol (RSVP) is used for this signaling process [8]. This model realizes and maintains all individual flows in the routers and assures timely delivery of packets. A user or application sends information to the network about the QoS requirements for the traffic profile. The traffic profile determines the requirements, e.g., how much bandwidth and delay should be provided. After getting confirmation about the QoS request from the network, the application is allowed to start sending and receiving data. The transmitted data must be within the limits given by the QoS device without any exceptions.

In the IntServ model, applications are required to reserve resources for each flow. IntServ creates a flow as a virtual-circuit that requires a signaling system to build up virtual circuit before the data transmission. RSVP is a signaling protocol to reserve certain amount of bandwidth for the requesting applications. RSVP is an independent protocol from the Integrated Service model. Therefore, RSVP can also be used in other models to create virtual-circuits [8].

In RSVP protocol, the sender sends a PATH message towards the receiver to inform it about the attributes of the desired traffic. Along the path each intermediate router moves this message to the next hop specified with the routing protocol. When the receiver approves the flow, it sends a RESV message that is moved hop by hop in the reverse direction towards the sender. The intermediate routers along the path may accept or reject this request; if the same request is accepted, resources are allocated, and then RESV message is moved towards the sender. On the other hand, if the request is refused, RESV-ERR message (error message) is sent by the intermediate routers back to the sender to inform that the request is rejected. If the sender receives a RESV message, then that means the resources have been allocated successfully and data can be send to the receiver. At the end of communication, the receiver sends a RESV-TEAR message towards the sender to delete the resource allocation. In turn, the sender sends PATH-TEAR message to remove the path stated in all routers [8].

Although IntServ provides guaranteed services for network applications, there are two issues with this model when it is applied to the Internet:

1. In this model, each router stores information for each stream. In fact, if this model is used widely in the Internet, then the routers will be overloaded and become slow, since too much information is stored for each flow by the routers on the path (scalability problem).

2. The IntServ model only provides two kinds of services: guaranteed and control-load. If an application requires another service different than these two types, then it is not possible (restriction on service type).

B. Differentiated Services (DiffServ)

The DiffServ model is a QoS support technology which is based on the class of service quality designed for IP. The purpose is to provide differentiated services for different types of applications on the Internet. It provides a traffic flow with predictable parameters like latency, bandwidth and packet loss. The IETF has developed this model to address the weaknesses of the IntServ model [6]. There are two main changes in it as follows:

1. The basic process shifts from the centre of the network (core routers) to the edge of the network (edge routers or hosts). As the routers do not need to store information about the flow, scalability issue in the IntServ model is solved. Applications or hosts define the type of services they need before sending packets.

2. The service per-flow is converted into per-class service. A router has to route packets formed on the class of services, e.i., the service class is not specified in the flow, but specified in the packet. This feature can solve the restriction issue on the type of service accordingly and it can define different classes depending on the application demands.

The IETF has decided to change the TOS (type of service) field in IPv4 and the traffic class field in IPv6 with Differentiated Services (DS) field [3]. Therefore, both the IPv4 and IPv6 packets have a field named the DS field. The value of this field is set at the boundary of the network by the sender host or the first router designated as the boundary router. The DS field has two sub-fields. The first one is named DiffServ Code Point (DSCP) consisting of 6-bits which determines the activity of router per-hop behaviour (PHB). The second field is defined as Currently Unused (CU) in RFC 2474 [3] and modified as Explicit Congestion Notification (ECN) in RFC 3168 [9], which is used by end-systems for notification about network congestion.

A router with DiffServ capability uses the 6-bit DSCP to refer an index table in which the packet handling mechanisms of the classes are defined when they are being processed. Some group of packets share a DSCP value so it refers to one Behavior Aggregate (BA), therefore core routers need to take only BAs into account instead of individual flows [3].

The per-hop behavior (PHB) specifies the packet forwarding properties relating to a particular class of traffic. This makes it possible for the network devices to prioritize traffic accordingly. Intermediate routers in the network can use the DSCP marking to make decisions based on how they are configured, to choose a particular PHB to ensure the flow of packets. This enables applications to function as designed. When DiffServ bits are set to zero, it indicates that there is no special handling requirement for this package in the network. Other marks should be in accordance with a specific forwarding behavior [10]. Differentiated services are split into two categories termed as assured forwarding and expedited forwarding.

Assured Forwarding (AF): AF provides a number of DS code points assigned to four AF classes (queues) known as AF1 to AF4. In each class, there are 3 bits of AF code which is described as low, medium and high, indicating the possibility of packets being dropped from that class. For each class, various levels of service are guaranteed, and drop precedence is used inside a class to specify the relative importance of the packet in case of congestion. The major objective of this PHB is to guarantee delivery as long as a given rate is not overridden. AF traffic is not subject to quantifiable time requirements such as jitter or delay [11]. Table 1 shows the DSCP values for the generally used four AF classes.

TABLE 1: THE DSCP VALUES FOR THE FOUR AF CLASSES [12]

Drop-Precedence	Class 1	Class 2	Class 3	Class 4
Low drop	AF11=001010	AF21=010010	AF31=011010	AF41=100010
Medium drop	AF12=001100	AF22=010100	AF32=011100	AF42=100100
High drop	AF13=001110	AF23=010110	AF33=011110	AF43=100110

Expedited Forwarding (EF): This service is described as Premium Service which is specially designed for traffic flow that requires low latency, low jitter such as multimedia real-time applications (video, voice, etc.). EF minimizes delaying of the packets. So, packets face very small or no queues in some cases and their processing are independent from the network load [13]. At edge routers of a network, the total EF traffics maximum rate must be less than the bandwidth of the slowest link in the network. As a result of that, this approach ensures bandwidth availability even in the congestion, all EF packets are not overloaded and can present desired behavior.

III. TEST ENVIRONMENT

OPNET Modeler is used for modeling and testing a network appropriate for the QoS architectures. The basic characteristic of OPNET is to provide different real-life network configuration capabilities which can make the simulation environment closer to reality [14]. The advantages of OPNET compared to other simulators include Graphical User Interface (GUI) for modeling, comprehensive library of network protocols and models, graphical interface to view the results, availability of documentation for the user to develop the network models.

Figure 1 shows the network topology used in this study. As seen in the figure, the router R1 is connected to the router R2 using a bi-directional DS1 link (1.5 Mb/s), whereas other workstations and servers are connected to each other via 10 BaseT (10Mb/s) links. The Profile, Application and QoS Attribute definitions are modified to support the network design.

The following six scenarios are modeled and simulated in this study.

- Scenario 1: DiffServ in IPv4 Network
- Scenario 2: DiffServ in IPv6 Network
- Scenario 3: IntServ in IPv4 Network
- Scenario 4: IntServ in IPv6 Network
- Scenario 5: Best effort in IPv4 Network
- Scenario 6: Best effort in IPv6 Network

The attributes of the routers, workstations and servers are defined to support the related scenarios. The QoS configuration is modified to present each simulated service model. The modeled networks are simulated for one hour duration using three applications which are web (HTTP), VoIP, and video conference. Among these different applications, VoIP has the highest priority and we aim at seeing the QoS provided by different architectures to this application. The statistics, which are important in the

evaluation of the networks, are collected. These statistics can be global representing all network components or local representing only one selected network component.

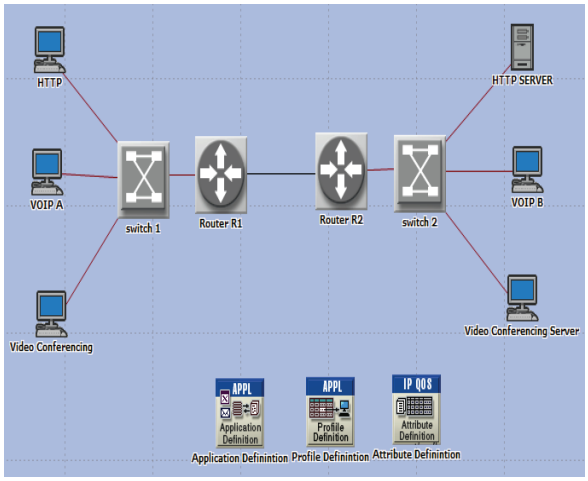


Fig. 1. Network topology for QoS architectures

IV. Test Results and Discussion

Figure 2 shows end-to-end delays for IntServ, DiffServ and Best Effort over the IPv4 and IPv6 networks. The values of IntServ and DiffServ over both IP versions are closer to zero in contrast to the Best Effort services. The large affect of QoS model on minimizing the delay compared to Best Effort technique, which uses FIFO as queuing algorithm, is noticeable with the obtained results.

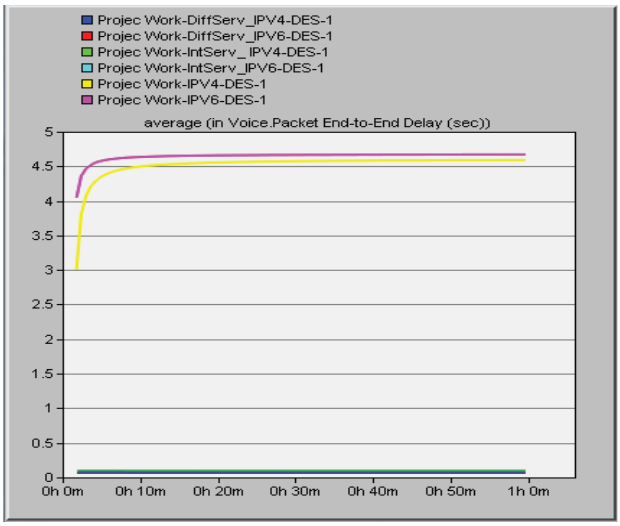


Fig. 2. Average end-to-end delays for 6 scenarios

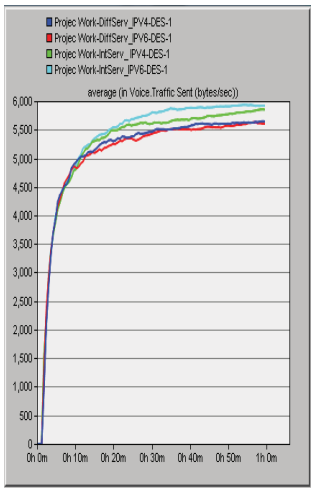


Fig. 3. Traffic sent

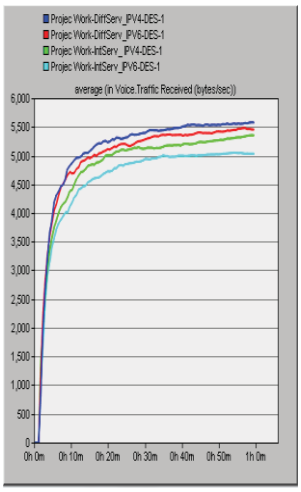


Fig. 4. Traffic received

Figure 3 and Figure 4 show the average traffic sent and the average traffic received by the voice application, respectively. As seen in the graphic, generally traffic sent in IntServ is higher for both IPv6 and IPv4. IntServ assigns a dedicated bandwidth for voice flows, so these flows of packets have more facilities to reach to destination than DiffServ which handles the classification of packets at edge routers. There is a slight difference between IPv4 and IPv6.

Figure 4 shows the average traffic received for the voice application. DiffServ has more traffic received than IntServ and both of them have less traffic received in IPv6 with small differences. DiffServ has more traffic received because it classifies the packets according to DCSP values assigned to packets at the network's edge, so, DiffServ processes packets much faster than IntServ in the core network. For DiffServ, the Expedited Forwarding (EF) is used in this work. EF minimizes delay that the packets face with and it has very small or has no queues in some cases and their operation is independent of the other loads in the network. IPv4 has more traffic received than IPv6, because the header size of IPv6 is bigger than IPv4. As a result, IPv6 has more packet drop and less traffic received compared to IPv4 [15].

Table 2 shows the percentage of loss packets in voice application over IPv4 and IPv6 when IntServ and DiffServ mechanisms are used. The heavy traffics load in the network is the main reason to increase loss packet ratio because the buffer becomes full which causes packet drops. The obtained results show that DiffServ has better performance in terms of packet loss.

TABLE 2: Values of Loss Packet Ratio in Voice Application.

Voice Application	IntServ		DiffServ	
	IPv4	IPv6	IPv4	IPv6
Traffic Sent (byte/sec)	5,855	5,936	5,646	5,631
Traffic Received (byte/sec)	5,352	5,031	5,577	5,453
Loss Packet%	8.59	15.11	1.05	3.16

Figure 5 shows the average end-to-end delay in voice application. DiffServ presents less packet delay in both IPv4

and IPv6 in comparison to IntServ, and the delay over IPv6 is less than the one over IPv4 in both QoS architectures [16].

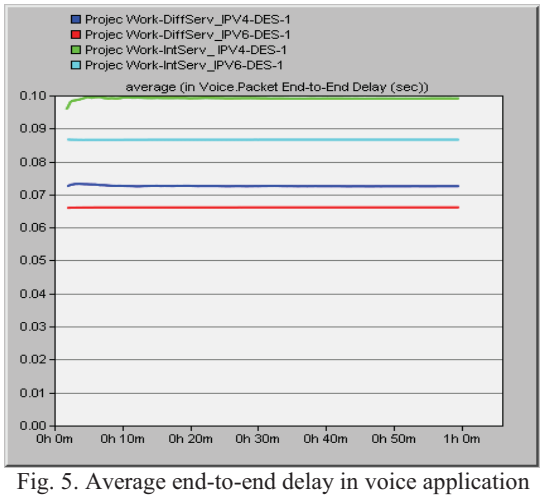


Fig. 5. Average end-to-end delay in voice application

In DiffServ based networks, all flows categorized into a given QoS class share the common network resources, therefore, the end-to-end queuing delay of a specific flow must be based on the amount of traffic from other flows which are part of the same class. The queuing delay in a DiffServ-based network may be less than the one in an IntServ-based network. In IntServ-based networks, each flow reserves network resources and the reserved resources are utilized by the flow with the highest priority. Hence, the maximum end-to-end queuing delay of a flow is possibly only predicted precisely from the amount of the reserved network resources. As a result, we find that DiffServ architecture causes smaller queuing delay than IntServ architecture. The obtained results are in line with the results given in [16], which investigates the performance of DiffServ and IntServ.

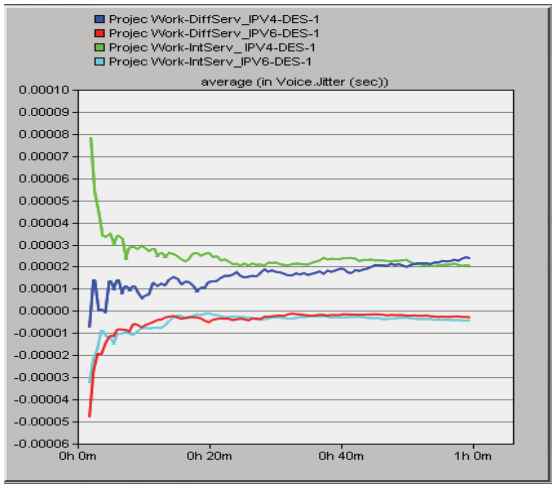


Fig. 6. Average jitter in voice application

Figure 6 illustrates the average jitter in voice application. The value of Jitter over IPv4 is higher than IPv6 for both QoS architectures. Both DiffServ and IntServ have closer jitter values to each other over IPv6. IPv6 supports classification of packets by using the flow label in header field, so, this field is used for reducing the total time processing load at the routers in a network and this situation results in minimizing jitter affect in the network. For the sake of voice application to be understandable, the voice packets need to reach at regular intervals.

V. Conclusions

In this paper, we have tested the performance of QoS architectures for VoIP applications over IPv4 and IPv6 networks. The obtained results show that QoS service architectures provide clear advantage compared to best effort service. When we compare QoS architectures, DiffServ gives better performance than IntServ over both IPv6 and IPv4 networks when multimedia traffic exists simultenously with other traffics on the network. IntServ produces more delay and jitter. On the other hand, IPv6 provides better delay and jitter than IPv4. Considering four scenarios with QoS, DiffServ over IPv6 network is the best solution for VoIP applications.

References

[1] Internet Live Stats, online: http://www.Internetlivestats.com/Internet-users/, accessed on March 2015.
[2] J. Gozdecki, A. Jajszczyk, R. Stankiewicz, “Quality of service terminology in IP networks”, Communications Magazine, IEEE, vol. 41, pp. 153-159, 2003.
[3] K. Nichols, S. Blake, F. Baker, D. Black, “Definition of the Differentiated Services Field (DS Field)in the IPv4 and IPv6 Headers”, RFC 2474, IETF, 1998.
[4] R. Hinden, S. Deering, “Internet Protocol, Version 6 (IPv6) Specification”, RFC 2460, IETF, 1998.
[5] R. Braden, D. Clark and S. Shenker, “Integrated Services in the Internet Architecture: An Overview”, RFC 1633, IETF, 1994.
[6] M. Carlson, W. Weiss, S. Blake, Z. Wang, D. Black, E. Davies, “An Architecture for Differentiated Services”, RFC 2475, IETF, 1998.
[7] G. Huston, “Next Steps for the IP QoS Architecture”, RFC 2990, IETF, 2000.
[8] J. Wroclawski, “The Use of RSVP with IETF Integrated Services”, RFC 2210, IETF, 1997.
[9] K. Ramakrishnan, S. Floyd, D. Black, “The Addition of Explicit Congestion Notification (ECN) to IP”, RFC 3168, IETF, 2001.
[10] A. B. Forouzan, “Data Communication and Networking”, Book, 4th Edition, Chapter 20, pp. 579-583, 2007.
[11] J. Wroclawski, W. Walter, H. Juha, B. Fred, “Assured Forwarding PHB Group”, RFC 2597, IETF, 1999.
[12] M. T. Aziz, and M. S. Islam, “Performance Evaluation of Real–Time Applications over DiffServ/MPLS in IPv4/IPv6 Networks”, (M.Sc.Thesis), School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden, 2010.
[13] K. Nichols, J. Van Jacobson, P. Kedarnath, “An Expedited Forwarding PHB”, RFC 2598, 1999.
[14] K. Salah and A. Alkhoraidly “An OPNET-Based Simulation Approach for Deploying VoIP”, International Journal of Network Management, vol. 16, no. 3, pp. 159–183, 2006.
[15] A. Ghaida, B. Amin, A. Mustafa, “Comparison between Ipv4 and Ipv6 Using OPNET Simulator”, IOSR Journal of Engineering, vol. 4, pp. 44-50, 2014.
[16] S. Shioda, K. Mase, “Performance Comparison between IntServ-Based and DiffServ-Based Networks”, In Global Telecommunications Conference, IEEE, vol.1, pp 1-6, 2005.

Visual Quality Estimation Technique for Services in Cloud Environments

Aleksandr Skatkov, Elena Mashchenko, Victoria Shevchenko, Dmitry Voronin
Sevastopol State University,
Sevastopol, Russian Federation
{AVSkatkov, maschenko, VIShevchenko, voronin}@sevsu.ru

Abstract—Scientific problems, related to the interaction between IT-services suppliers and consumers in the cloud environments, has been described. The analysis of quality metrics, used for IT-services estimation, has been carried out taking into account the specificity of the cloud infrastructure that focuses on large-scale scientific problems solving. The technique, used for the visual express quality estimation of IT-services, has been proposed. The method of parallel coordinates has been used for multidimensional data visualization. The proposed technique can be used as an effective instrument for visual analysis of cloud services functioning (the dynamics of their quality indicators is considered).

***Index Terms*—Cloud Computing, Quality of Service, Quality Metrics, Data Visualization, Parallel Coordinates Method.**

I. Introduction

Cloud services mean an approach to the IT-infrastructures creation in which the consumer gets the required IT-resources from external providers in the form of web services. The key features of such services are the following: self-service on-demand, broadband network access, resource pooling, instant flexibility, measurability of the provided services [1]. The cloud solutions implementation can be a rather profitable alternative decision for the enterprise (in comparison with private IT-infrastructure maintenance). Clouds can save the costs that are usually spent on the purchase and support of the private computer equipment. Moreover, it can be helpful to adjust the amount of simultaneously used resources paying only for the consumed volumes. In addition, this approach allows business processes to be simplified within the company as the most of issues (relating to IT) are outsourced to cloud provider. Under these conditions, the actual task is to ensure the required level of cloud services quality.

There are many techniques helping to organize the management activities for IT-service provision. For example, there are the following: ITIL (IT Infrastructure Library), MOF concept (Microsoft Operations Framework), COBIT standard (Control Objectives for Information and Related Technology) [2–4]. However, none of these techniques can provide exhaustive and scientifically grounded recommendations on the formation or improvement of IT-service management structure in order to increase its efficiency in cloud environments.

II. Problem Statement

In this study, the authors discuss the visual quality estimation technique for cloud services that solve scientific problems. Cloud services usage can significantly increase the availability and simplicity of the large-scale computational experiments implementation needed for interdisciplinary scientific research. At the same time, there is a problem of "trust" to the cloud environment from the researchers' side. It will be appreciated that some of the users, involved in large-scale scientific problems solving, do not have the necessary level of competence for the installation, configuration and sharing scientific applications in the cloud. The collusional situations that arise during the scholars' work just exacerbate the problem of "trust" to cloud technology. Thus, cloud providers urgently need new approaches and technologies that are suitable to meet the following functional requirements: 1) to estimate the degree of data processing criticality for large-scale scientific problems solving; 2) to organize the guaranteed management of IT-services quality to resolve the considered contradiction.

Cloud service providers are interested in using simple visual tools that give users an opportunity to evaluate the quality of IT-services in the cloud and to calculate the financial benefits from cloud computing introduction.

The object of the study is associated with methods for the IT-services quality evaluation in the high-performance cloud environments, focused on the large-scale scientific problems solving.

The subject of the study is connected with the process of the interaction between the actors in cloud environment. Properties of the actors and the main scenarios of their interaction are described in the extended reference model [5–7] that takes into account the specificity of large-scale scientific tasks. In this model, the following cloud actors are considered: Consumer, Provider, Auditor, Broker, Crisis manager and Composite architect.

The aim of the study is to offer an efficient adaptation of the existing imaging techniques to the problems of visual quality estimation of services in cloud environments that focuses on large-scale scientific problems solving.

III. Quality Indicators Analysis for IT-services in Cloud Environment

Many of the standards that are currently applied to cloud computing have been developed for "before-cloud" technologies (such as web and telecommunication services). Currently there is an active development of standards and guidelines that are designed specifically for cloud services. The "cloud" standards are developed in two directions: technical solutions and business solutions. Basic provisions of these standards that take into account the business purposes were formulated in ETSI (European Telecommunications Standards Institute) [8]. Among the technical solutions, it is possible to distinguish four categories of cloud standards that are used for IT-services quality estimation. These categories have the following notations: security, interoperability, data portability and reversibility.

Based on data that was taken from the works [9-11], the authors proposed a hierarchy of indicators for cloud services' quality estimation (fig.1). The level L_0 corresponds to the global level estimation of the quality of cloud services. The level L_1 corresponds to a set of requirements that are imposed on the participants that provide or use cloud services. At the level L_2 the decomposition of requirements for cloud services was carried out. As the result, the three groups were obtained. They have the following notations:

- Requirements to financial indicators of business processes.
- Requirements to indicators of supporting processes of IT-services.
- Requirements to technical indicators.

The levels $L_3 - L_4$ detail the elements of the second level by presenting different types of measured metrics and calculated indicators, for example, such as the following: financial indicators, qualitative indicators, probabilistic indicators; indicators that take into account the service level management and the availability of service.

The technical performance requirements are based on the NIST standards [12,13] that include five most important characteristics of cloud services (on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service).

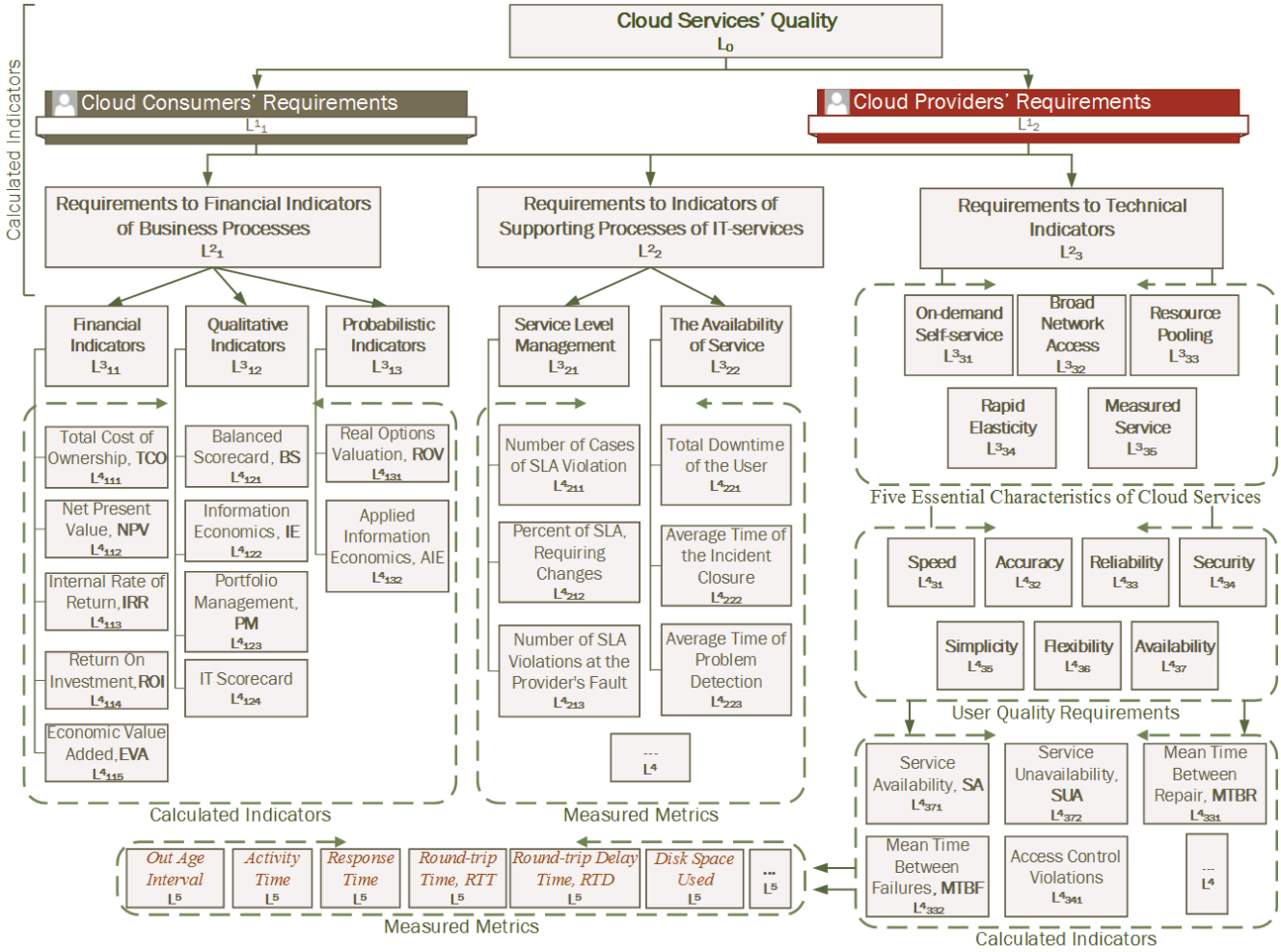


Fig. 1. Hierarchy of quality indicators of services in cloud environments

IV. Visualization Technique

The proposed technique of visualization is based on the modern concept of Visual Analytics (VA) that is also called Visual Data Discovery (VDD) [14]. Such systems are designed to help the low-complemented user to get brief, easy to understand information about the values of cloud services' quality indicators. It can improve the efficiency of distributed calculation used for the solving of large-scale scientific problems. Such systems have a thick layer of automatic data analysis. In the context of the study, this layer implements the calculation of complex quality indicators based on measurable metrics and then transmits the information to the visualization layer.

Dashboards are one of the most important elements of the user's interface in visual analytic systems. They provide multiple visual indicators on a single page. The authors propose to use the method of parallel coordinates [15] and petal charts for multidimensional data visualization. The method of parallel coordinates provides an opportunity of effective multidimensional data visualization even if the information is measured in different scales. In addition, a researcher can easily determine if there is a correspondence between the requirements (SLA, etc.) and the values of IT-services' quality indicators.

The proposed technique has the following steps:

1) Getting the data of cloud services monitoring. The information, containing measured values of metrics, is received from the monitoring systems and then it is transmitted to the analytical processing system. The metrics are grouped in accordance with the requirements of the financial, technical and other indicators (fig.1).

2) Complex quality indicators calculation. For each group of indicators and for a certain period of time the complex indicators are calculated based on the values of measurement metrics.

3) Visualization. Visual researcher's interface is represented as a Dashboard. For each group of quality indicators it is possible to choose the form of presentation: radar chart or parallel coordinates. Data normalization is performed only for the method of parallel coordinates. It allows to visualize the degree of closeness between the quality indicators' values and the critical values that are defined in the SLA.

Example. Let's represent the values of technical indicators shaped like parallel coordinates (fig.2). The considered indicators are the following: Service Availability, Service Unavailability, Mean Time Between Repairs (MTBR), Mean Time Between Failures (MTBF), Access Control Violations.

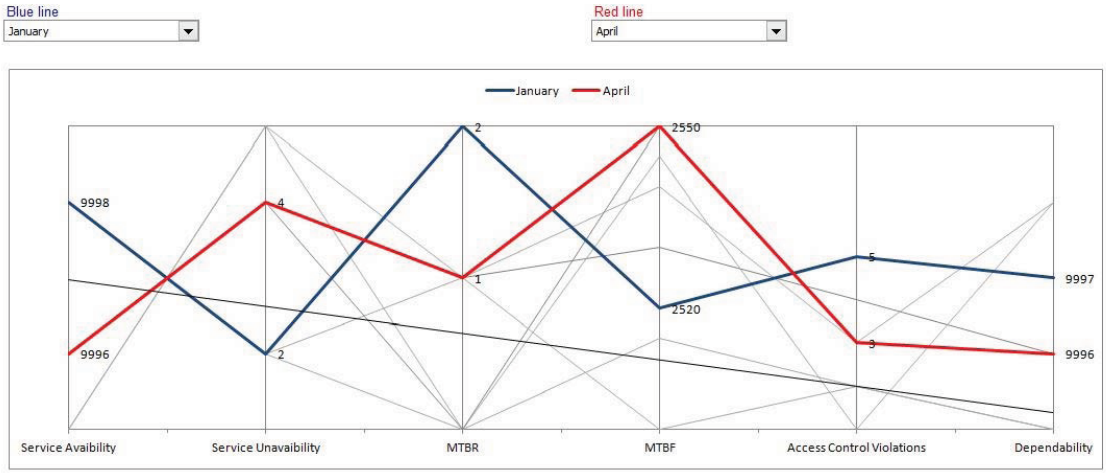


Fig. 2. Visualization with parallel coordinates

The presented information has been collected for a period of six months. Each line represents a set of indicators for a particular month. The researcher has the opportunity to enhance the color of the necessary visual elements and to analyze the relevant indicators. For example, in January 5 events of Access Control Violations took place. The current value of the Service Availability indicator was 0.9998, etc.

The radar charts are a more compact form of visual representation of the monitoring results that are concerned with the quality of cloud services. To construct the radar chart a researcher selects a set of IT-services' quality indicators that is denoted as {Pi}. The selected indicators are plotted on each of the radar chart axes. The relative indicator value is determined for each axis. It forms the outline of balanced solutions that are denoted as K (fig.3). This outline limits the balanced solutions area that is denoted as S_K.

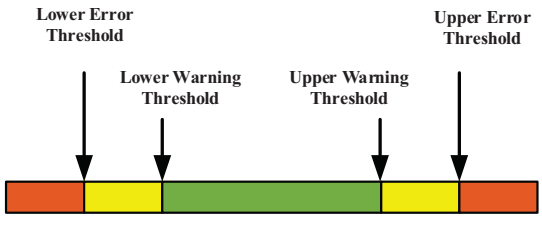


Fig. 3. P_i – Parameter Thresholds

The outline of acceptable QoS level (green zone for each indicator) forms the area that is denoted as S_Y. This area should be located: 1) within the outline K, if the $\{P_i\}$ is

bounded above; 2) between the zones S_{K1} and S_{K2}, if the $\{P_i\}$ is bounded above and below (fig 3); 3) over the area S_K, if the $\{P_i\}$ is bounded from below.

The difference between areas that are limited by the K and Y outlines (Eq. 1) forms the "imbalance" zone that is denoted as D. This zone illustrates the reserve of quality services' indicators on the radar chart. This reserve is formed according to the amount of computing resources that were allocated in conformity with the service level agreement (SLA) that is signed between the Cloud Consumer and the Cloud Provider. In efficient SLA this reserve is minimized.

$$S_D = S_K - S_Y. \quad (1)$$

An example of quality indicators visualization by using a radar chart is shown in Figure 4.

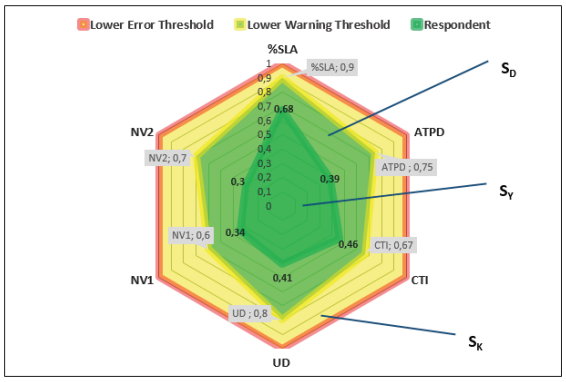


Fig. 4. Visualization with radar chart

Measurable metrics from the L^2_2 group were selected for visualization. They are the following: Percent of SLA Requiring Changes (%SLA); Average Time of Problem Detection (ATPD); Average Time of the Incident Closure (CTI); User Downtime (UD); The number of SLA violation (NV1); Number of SLA Violations at the Provider's Fault (NV2). All of the selected metrics are limited to the top. The averaged results of a survey of 150 respondents working on the problem of distributed simulation and limit values of the metrics are shown in Table 1.

TABLE I. QUALITY METRICS VALUES

Quality levels	Measurable metrics					
	%SLA	*ATPD (min)*	*CTI (min)*	*UP (min)*	*NV1*	*NV2*
Lower Error Threshold	40	120	180	600	25	10
Lower Warning Threshold	36	90	120	480	15	7
Respondent	27	47.2	82	248	8.5	3
Normalization	0.675	0.39	0.46	0.41	0.34	0.3

A researcher can analyze the dynamics of the cloud services' quality indicators by months based on the complex of diagrams similar to those that are shown in fig. 4.

V. CONCLUSION

The proposed estimation technique for visualization of cloud services' quality, dispute from the known decisions, is oriented on the visualization of calculated complex quality indicators. This technique can be implemented in the decision-support systems for management of cloud services' quality.

It can be used as a part of Visual Analytics system that can be rather efficient in services' quality management for cloud environments that focuses on large-scale scientific problems solving.

ACKNOWLEDGMENT

This work was supported by the Russian Foundation for Basic Research (project № 15-29-07936).

REFERENCES

[1] M. Litoiu, M. Woodside, J. Wong, J. Ng, and G. Iszlai, A business driven cloud optimization architecture, Symposium on Applied Computing, 2010, pp. 380–385.

[2] Ya.V Bon, G. Kemmerling and D. Pondaman. IT Servis-menedzhment, vvedenie, IT Expert, 2003.

[3] "Cross Reference ITIL V3 and MOF 4.0", Web: http://microsoft.com/mof, May 2009.

[4] "Vendor Management: Using COBIT 5", ISACA, Web: http://www.isaca.org/knowledge-center, Feb. 2014.

[5] A. Skatkov and V. Shevchenko, "Expansion of reference model for the cloud computing environment in the concept of large-scale scientific researches" (in Russian), Proceedings of the Institute for System Programming, Vol. 27 (Issue 6), 2015, pp. 285–306, DOI: 10.15514/ISPRAS-2015-27(6)-18

[6] A. Skatkov, E. Maschenko, V. Shevchenko and D. Voronin. Actors Interactions Research in Cloud Computing Environments Using System Dynamics Methodology, Proceedings of the 18th FRUCT & ISPIT Conference, April 2016, pp.612–619.

[7] A. Skatkov, V. Shevchenko and Y. Nikolaeva. Multi-agent model of cloud computing environment for multidimensional objects monitoring (in Russian), Environmental control systems, Issue 1(21), 2015, pp. 45–51.

[8] ETSI Technical Report Cloud; SLAs for Cloud services, 103 125 v1.1.1, Nov. 2012, Web: http://www.etsi.org

[9] Cloud Standards Customer Council (2014). Practical Guide to Cloud Computing, Version 2.0, Web: http://cloud-council.org/resource-hub.htm#practical-guide-cloud-computing

[10] Cloud Standards Customer Council (2015).Practical Guide to Cloud Service Agreements Version 2.0, Web: http://www.cloud-council.org/deliverables/CSCC-Practical-Guide-to-Cloud-Service-Agreements.pdf

[11] Tele Management Forum (TMF), Ed., SLA Management Handbook: Volume 4 - Enterprise Perspective, Published by The Open Group, Oct. 2004.

[12] P. Mell and T. Grance. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology, 2010.

[13] NIST Special Publication 500-307 Cloud Computing Service Metrics Description, NIST Cloud Computing Reference Architecture and Taxonomy Working Group, 2015.

[14] Cindi Howson. Visual Data Discovery: 4 Storytelling Approaches Compared, 2014.

[15] Alfred Inselberg. Parallel Coordinates: Visual Multidimensional Geometry and Its Applications. Springer Science & Business Media, 2009.

Analysis of the status of marine automation systems problems on the example of modern technological networks

Sergei Sokolov, Sergei Chernyi,
Department of information security
Admiral Makarov State University of Maritime and Inland Shipping
Saint-Petersburg, Russia
sergiiblack@gmail.com

Aleksandr Bordug, Ivan Titov
Department of ships electrical equipment and automation
Kerch State Maritime Technological University
Kerch, Russia
laboratory45@ya.ru

***Abstract*—At present, the interest in application of synchronous machines in the various systems of the electric drive and energy sources is still growing. Synchronous motors and their modifications (ac converter-fed motor, etc.) enable to develop low-noise, reliable and economically efficient electric drive systems. The article is devoted to research and development of a common methodology for the creation of mathematical models and information technologies for solution the problem of decision-making under conditions of uncertainty of input data. The peculiarity lies in the fact that all types of interval uncertainties have a specific semantics and, as a consequence, the "dimension".**

***Index Terms*—Ethernet, static error, control systems, oil and gas industry**

I. STATUS PROBLEMS IN THE MARITIME INDUSTRY

With the view of the allocation of resources in the Russian waters, it is clear the future direction for development of Russian offshore hydrocarbon production. It will be aimed at the waters of the Arctic shelf in the long run, where most of oil and gas reserves are concentrated. The Far Eastern shelf will be paid attention to as well, where at the moment one of the most important strategic oil fields (shelf of Sakhalin) are located, and the Caspian, the Black and Azov Seas, including the recently annexed territory of the Crimean Peninsula with a number of nationalized marine and terrestrial deposits (mostly gas ones).

If we take into account the assessment by O.Prischepa, General Director of the Federal State Unitary Enterprise "All-Russia Petroleum Research Exploration Institute" (VNIGRI), it is still too early to speak about the guaranteed prospects for the Russian shelf, because of the complexity of production and the lack of the required technology and expertise to develop the Arctic shelf, as well as all the environmental hazards which can be associated with its development. Severeoperational conditions (storm, movement and the formation of ice, the ice cover of the sea, the critical temperatures, etc.) and the borders of areas of the Arctic waters under dispute, cause problems, due to which the permanent suspension of the plans of works, carrying out for the development of the Arctic shelf and legal restrictions of the development activities will last for more than one year [1-10]. According to the General Director of VNIGRI and results of survey to 108 respondents [2, 11], among which oil and gas giant Russian market research companies and organizations engaged in services and equipment, it becomes clear that at the moment the most promising hydrocarbon deposits are located on the Caspian shelf (Korchagin field V. Filanovsky (12-15 million tonnes / year), as well as Sarmatian, and cross-border Khvalynskoe Rakushechnoye). Rights for the development of these fields belong to the JSC "Lukoil". The second promising area is the northern Sakhalin shelf, namely the projects "Sakhalin-1" and "Sakhalin-2" (oil extraction is estimated as 20 million tons / year and gas as 30 billion cubic m / year). Operators are Royal Dutch Shell and JSC "Gazprom. Further, the Arctic shelf and the shelf of the Black and Azov seas are next in the ranking.

Analyzing the assumption by O.M.Prishchepa and the survey results, we can understand that the main focus is aimed at the development of Russian oil and gas fieldsalready available by their natural (available depth, the impact of weather conditions), technical (available to the technical equipment) and economic (profitability of production) factors. Now, the technical equipment, required for the development of the shallow shelvesin the Caspian, Black, Azov Seas and the Sea of Okhotsk, as well as natural and weather conditions, are conducive to the normal development of fields located in these waters almost without any problems. These factors are the primary cause of slow development of the Arctic shelf, though the bulk of offshore hydrocarbons are found there [9-13].

Implementation of innovative techniques and equipment capability for the development of the Arctic shelf is still poor. The reason for this is the insufficient knowledge about areas and lack of world practical experience gained over the years of the offshore production in the Arctic with its harsh conditions. However, this does not in any way will be the cause of refusal in favor of Arctic hydrocarbon production in the long term. For example, Prirazlomnoye field in the Pechora Sea is still quite promising. It is located 55 km offshore. This field was put into service in 2013 and currently is the only Russian Arctic offshore field. For the first time production of hydrocarbons is

carried out in the Arctic shelf the in a similar way, with the use of a fixed platform "Prirazlomnaya".

The Ministry of Natural Resources and Environment also provides the data (Table 1) on the resource potential of the RussianArctic waters in terms of billion tons of oil equivalent, according to which you can see the actual distribution of offshore resources in 2013 [9-14].

TABLE I. THE RESOURCE LEVEL OF CERTAIN ARTIC WATER BODIES IN THE RUSSIAN FEDERATION IN THE BILLION TONS OF THE OIL EQUIVALENT

Arctic waters	Resource level	Booked reserves
The Barents and Pechora Seas	30.3	4.95
The Bering Sea	1.08	0
The East Siberian Sea	5.58	0
The Kara Sea	41.2	4
The Laptev Sea	3.26	0
The Sea of Okhotsk	8.7	1.7
The Chukchee Sea	3.3	0
Totally by water bodies	93.42	10.65

Thus, with the further development of marine energyproduction, the priority will be sea hydrocarbon production in the Russian continental shelf, which share in the Russian total production will grow from year to year. Over the period from 2010 to 2014, there were drilled 18 prospecting and 17 exploration wells on the Russian continental shelf with the 26 discovered fields, mainly in the Far East region and almost 122 licenses were issued for development in the period from 2015 to 2020. Most (60%) of the licenses shall apply to the waters of the Arctic seas. Main companies, that will develop these fields, were Rosneft, Gazprom and Lukoil. At the beginning of 2015, twenty-two deposits were discoveredin the Russian Western Arctic seas and eighty-eight exploratory wellswere drilled.

Taking into account the geographical position of Russia, where a large part of the shelf is located in the Arctic and northern regions, one should expect the implementation and development of breakthrough technologies there. This is one of the important aspects in solving the problem how to extract the challenged Arctic hydrocarbons. At the same time if you really look at the situation in the near future, Russia will not be able to switch to the Arctic hydrocarbon reserves as an alternative energy supply. It is necessary to arrange the deep-water exploration and production of oil and gas reserves at this stage, as most of the shelf and deep waters remain unexplored (Black Sea). According to [4-11], 60% of experts believe that to develop Russian offshore projects only on their own, without the use of foreign technologies is not possible. This fact should be taken into account as well. These technologies include:

- Technologies and equipment providing subsea extraction, drilling equipment.
- Support fleet with available heavy lift capacity.
- High-frequency and wideband 3D seismic survey.
- Information Systems.
- Technology of controlled deep-sea drilling.
- The equipment capable of being operated at extreme temperatures and icing conditions, for year-round drilling in the Arctic and Far Eastern regions.
- Drilling platforms.

II. METHODS OF NETWORK ARRANGEMENT SELECTION

As the physical level and network protocol information channel for most of the personal computers (PC) and the equipment units controllers Ethernet network protocols used at a data rate of 100 Mbit / s.

Communication through Ethernet is performed between the workstation and the remote input-output V-DAQ ™ (placed on workstations), and the corresponding controller hardware devices.

Copper cable Category 5 (CAT-5) is typically used for devices, placed near each other (in the same cabinet or adjacent cabinets). At large distances, typically use a multimode optical fiber cable range of 62.5 / 125 micron.

The Amphion network is commonly used communication via the bus PROFIBUSDP. Communication protocol PROFIBUS is an international standard channels of communication between hardware controllers, remote input and output devices, actuators, valves, operator interfaces and computers. The PROFIBUS network supports data rates over copper twisted pair or fiber optic cable up to 12 MBd. And application-level interface communication link between the remote controller and the equipment input-output devices are typically implemented as PROFIBUS networks.

The methods involve performing the following steps:

To generate a range Ψ of possible network interfaces;

To generate a range Ω of analyzed parameters in the network interfaces;

To generate a range of weights V, where $v_i \in V$, is weight assessment of the characteristics $\omega_i \in \Omega$;

Reset to 0 s_i, for all $s_i \in S$;

To build up a set of total weights S, where si∈S, is the total weight of the network interface $\psi_i \in \Psi$;

To choose the parameter $\omega_i \in \Omega$, $s_i = s_i + F\left(O\left(\psi_i \in \Psi, \omega_i \in \Omega\right)\right)v_i$, where $O\left(\psi_i \in \Psi, \omega_i \in \Omega\right)$ is the function equal to deviation of the network analyzed parameter $\psi_i \in \Psi$, from the parameter reference value ωi∈Ω; F(x) is a serial number in the ascending-sorted set of values for the function $O\left(\psi_i \in \Psi, \omega_i \in \Omega\right)$;

To eliminate $\omega_i \; from \; \Omega$, *if* $\Omega \neq \varnothing$, go to para 6;

To choose $s_{\min} = \min_{s_i \in S} s_i$, that means that the network interface ψi is best suited for this problem of automation.

Table 2 is a comparative table of parameters of the most popular local microcontroller networks in respect of the busbar characteristics.

TABLE II. SPECIFICATION FOR NETWORKS

Busbar type	RS-485	I2C	LonWork	CAN	MicroLAN
Max length of wire, m	2000	8	72	40	300
Max number of attachment units	32	14	127	127	256
Mains supply	-	+	-	-	+
Number of wires in a busbar	2	4	2	2	2
Availability of stand-alone devices	-	-	+	-	+

In analyzing networks as well as their performance and search for optimal data transfer routes, it is helpful to apply the graph theory [1-6]. The weight of the arc in the graph is as follows [7-15]:

$$u_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (1)$$

where x_i, y_i are Cartesian coordinates of the vertex v_i, xj, y_j are Cartesian coordinates of the vertex v_j.

In [8] the notion of the vertex separation number is introduced:

$$C(v_i) = \max_{v_j \in V}[d(v_i, v_j)], \quad (2)$$

where $d(v_i, v_j)$ is the shortest distance between vertexes i and j.

$$d(v_i, v_j) = \sum_{k,j \in T_{ij}} u_{kj}, \quad (3)$$

where T_{ij} is a graph simple path $G\ (V, U)$ with the beginning in the vertex vi and the end in the vertex v_j. [13-17].

Thus, the length of the maximum segment is as follows:

$$C(v_m) = \max_{v_j \in V}[d(v_m, v_j)],$$

where v_m is the vertex where the network manager is located.

The graph radius, being the minimum one of the eccentricities of the connected graph vertexes, is determined with the expression [9-18]:

$$R_0 = \min\left[d(v_m, v_j)\right].$$

Use of the graph theory makes possible to search the optimal route for transfer of the information control packets. In searching this route, such options as reliability of the data transfer, network delay, discard probability; channel loading, etc. should be taken into account. Setting of the network load simulation problem. The main parameter of the data network is the network traffic. In this system, traffic is a random fuzzy variable. While studying and assessing the performance of the proposed model, one may make load forecasts both for certain network channels and for the network as a whole. This simulation system may be a part of the computer-aided design engineering of data-processing networks.

Simulation time means step-by-step time, i.e. network traffic changes at all the network nodes with every step. Alternators of the network traffic are divided into the following types: expansion-operating generators (generators that go off just once per a finite number of steps of time); and continuously operating ones.

Dependence of transmission reliability on the traffic volume.

Assessment is produced as a fuzzy random value (FRV) of the following type:

$$T_r = \{A_1 / P_1, A_2 / P_2, A_3 / P_3\},$$

where A_1, A_2, A_3 are certain fuzzy sets and P_1, P_2, P_3 are a fuzzy membership function [14-17].

Remote input-output devices collect data for the hardware controller and can be configured to address a wide range of local and remote input-output tasks.

Sensors and controls equipment and workstation using a wiring connected to the remote input-output device. Hence, by a single cable PROFIBUS communicates Remote input-output devices to the controller hardware or other information processing devices.

WITSML, PRODML, RESQML allow greater detail description of objects - roughly an order of magnitude. And because the standards are designed so that there is no pre-set restrictions on how to name the parameters, it is possible to extend the range of data going beyond standard, respecting the full compliance with the standard. Theoretically, you can expand the standard to an unlimited number of parameters.

Motor winding is non-uniform heat transfer system from conductors through several layers of insulation to the boundary separating mechanism and the external environment. The high thermal conductivity allows the winding to prevent the absence of significant temperature gradients in some of its points. Thus, despite heterogeneity of heat transfer, it may be considered that the average winding temperature differs slightly from the temperature in each of its points. Parameters of the process of heat transfer may be treated as average ones throughout. The heat capacity of the environment is infinite; the ambient temperature is constant.

Availability Server field, supporting this standard, it became possible to combine all fields data from all sensors and rig in one place, which in turn allows you to: conduct a full and thorough analysis, as the maximum number of available data and nomenclature; monitor important and critical parameters in

a single control center; replace with minimal costs and add new operating equipment from different manufacturers, provided that it supports the standard; use any data processing software (provided it supports standard), since it is possible to receive any data from the field; remote monitoring and analysis of all available data in the field.

The well-known methods of the integral automation for electric power complexes at the offshore drilling rigs are characterized by solving the problems of the individual automation of certain processes and subsystems, which are integrated into the unique system of control at the highest level of hierarchy. This results in great excess of hardware facilities in the electric power system and its increased cost.

The boundary conditions at the left extreme point are represented by the equation

$$\frac{dT}{dt} = B_1 V - B_2 (T_1 - T_2), \qquad (1)$$

where V is control function (electric power supplied to the motor winding), W;

T_1 and T_2 are temperatures of the winding and the water surface border, respectively, ^oC; t is current time, c.

We assume the thermal process in the motor winding with relatively slow integration pitch ΔT. Then, applying to (4) Euler's method, we obtain:

$$T_{1,k+1} = T_{1,k} + (B_1 V_k - B_2 (T_{1,k} - T_{2,k}))\Delta t .$$

Equation in the finite differences

$$T_{2,k+1} = R_2 T_{2,k} + q_2 (T_{1,k} + T_{3,k}) ;$$

$$T_{3,k+1} = R_3 T_{3,k} + q_3 (T_{2,k} + T_{4,k}) ;$$

……..

$$T_{n,k+1} = R_n T_{n,k} + q_n (T_{n,k} + T_{n,k}) .$$

In a more compact form

$$T_{1,k+1} = T_{1,k} + (B_1 V_k - B_2 (T_{1,k} - T_{2,k}))\Delta t ;$$

$$T_{j,k+1} = R_j T_{j,k} + q_j (T_{j,k} + T_{j,k}) \, j = 2, n ,$$

Where $T_{1,k+1}$ is the right boundary condition;

$T_{1,k}$ is the left boundary condition for each interval n along the coordinate x;

$R_j = 1 - 2\Delta t \cdot a_j h^{-2}$, for the *k*-th at the time moment;

$q_j = \Delta t \cdot a_j \cdot h^{-2}$, for the *k*-th at the time moment.

Values R_j and q_j were obtained by applying the Euler's standard procedure of transition from differential equations to linear equations.

III. Conclusion

A new group of automation can help to overcome many problems typical of the oil/gas industries, to improve the efficiency of oilfield development. This group can be attributed to new developments in the field of downhole sensors that will provide measurements of high quality; new developments in the design of drill pipes with cable or fiber optic wiring, with the help of which it will be possible to provide of large amounts of receive dinformation to the surface. Such tubes will ensure an opportunity to monitor pressure and temperature in the annulus along the drillstring, which will allow monitoring of the parameters along the whole barrel.

References

[1] S. Chernyi, "Analysis of the energy reliability component for offshore drilling platforms within the Black Sea," Neftyanoe Khozyaystvo - Oil Industry, 2016, Issue 2, pp. 106-110.

[2] A. Zhilenkov, and S. Chernyi, "Investigation performance of marine equipment with specialized information technology," Procedia Engineering, vol. 100, 2015, pp. 1247–1252.

[3] S. Chernyi, and A. Zhilenkov, "Modeling of complex structures for the ship's power complex using XILINX system," Transport and Telecommunication, 2015, vol. 16 (1), pp. 73–82.

[4] Francy D. Rodríguez, Silvia T. Acuña, and Natalia Juristo, "Design and programming patterns for implementing usability functionalities in web applications," Journal of Systems and Software, Volume 105, pp. 107-124, 2015.

[5] Ward Passchyn, Dirk Briskorn, and Frits C.R. Spieksma, "Mathematical programming models for lock scheduling with an emission objective," European Journal of Operational Research, Vol. 248, Issue 3, pp. 802-8141, 2016.

[6] Nanyan Jiang, and Manish Parashar, "A programming system for sensor-based scientific applications," Journal of Computational Science," Volume 1, Issue 4, pp. 206-220, 2010.

[7] Guilherme Galante, and Luis Carlos Erpen De Bona, "A programming-level approach for elasticizing parallel scientific applications," Journal of Systems and Software, Volume 110, pp. 239-252, 2015.

[8] E. Pinheiro, and L. A. Weber Barroso, "Failure Trends in a Large Disk Drive Population," Proceedings of the 5th USENIX conference on File and Storage Technologies (FAST'07), pp. 17-28, 2007.

[9] S. Chernyi, "Use of Information Intelligent Components for the Analysis of Complex Processes of Marine Energy Systems," Transport and Telecommunication Journal, 17(3), 202–211, 2016. DOI: 10.1515/ttj-2016-0018

[10] Olifer, V.G., Olifer, N.A. Computer networks. Principles, technologies, protokols: manual for higher educational institutions, Piter-press, 2002.

[11] Alekseev, N.A., Baranov, A.P. Microrocessor systems for control of shipborne electric power plants "ASA–S", St. Petersburg, Admiral Makarov State Maritime Academy, 2000.

[12] Nadezhnost' avtomatizirovannykh IS: uchebnoye posobie (Reliability of automated Iformation Systems: manual), Fatuev, V.A., et al. Ed., Tula, 1998.

[13] Gorbatov, V.A., Smirnov, M.I., Khlytchev, I.S., Logicheskoye upravlenie raspredelennymi sistemami (Logic control of the distributed systems), Moscow, Energoizdat, 1991.

[14] Bolotin B.I., Weiner V.L., Engineering methods of calculation of stability of ship's automated systems. Shipbuilding, 1974.

[15] Tolshin, V.I, Bolotin, B.I., Konks, G.A. Influence of fuel equipment on the stability of diesel-generator with capacity of 150 kW. Proceedings of the Scientific Research Institute of Information of the Heavy Industry, 1969, 4, pp. 3-9.

[16] A. Nyrkov, S. Sokolov, A. Zhilenkov, S. Chernyi and D. Mamunts, "Identification and tracking problems in qualimetry inspections in distributed control systems of drilling platforms," 2016 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), Saint Petersburg, Russia, 2016, pp. 641-645. doi: 10.1109/EIConRusNW.2016.7448265

[17] A. Nyrkov, S. Sokolov, A. Zhilenkov and S. Chernyi, "Complex modeling of power fluctuations stabilization digital control system for parallel operation of gas-diesel generators," 2016 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), Saint Petersburg, Russia, 2016, pp. 636-640.

[18] S. Chernyi, "Techniques for selecting topology and implementing the distributed control system network", IOP Conf. Ser.: Mater. Sci. Eng., vol. 124, p. 012048, 2016.

Diagrammatic models processing in designing the complex automated systems

A.N. Afanasyev, N.N. Voit
«Computer facilities» department
Ulyanovsk state technical university
Ulyanovsk, Russia
{a.afanasev, n.voit}@ulstu.ru

R.F. Gainullin
Ltd «Ecwid»
Ulyanovsk, Russia
r.gainullin@gmail.com

***Abstract*—The paper proposes the syntax-oriented processing (analysis, checking and translation) methods of diagrammatic models in the design of complex automated systems. The methods are based on the author's finite-state graphical grammar.**

***Index Terms*—Diagram, processing, error checking**

I. Introduction

The use of diagrammatic models especially at the conceptual stages of complex automated systems (CAS) design contributes to the success of the CAS design as a result of the unification of the language of interaction among the CAS development staff, strict documentation of architectural design and functional solutions, formal checking of the diagram notations correctness and automatic (automated) diagram translation into code. The visual languages (UML, IDEF, eEPC, BPMN, etc.) have been developed and widely used for diagrammatic model development.

However, in the modern theory of graphical and diagrammatic languages there is no effective method of analysis and checking (especially with regard to complete checking) of syntactic and semantic errors of diagrammatic notations. The 'direct' checking methods have implemented in diagram development tools. These methods require multiple passes, allowing a learning system to analyze the syntactic correctness of diagrams and identify the main errors. However, they can't identify errors of so-called "the torn context" caused by the use of such connectors as AND, OR, EXC OR, etc. The context-sensitive semantic errors related to a text of any diagram are not checked. The above-mentioned facts reveal additional types of "expensive" errors in the CAS design that are difficult to diagnose. Analysis and checking of these errors are an important scientific and technical challenge.

II. Overview of the problem

The purpose of the study is to expand the class of diagnosed errors in the CAS design based on developing and implementing the methods and tools for diagrammatic notions analysis and checking. As a result, the number of errors and the time taken for the CAS design should be reduced. The method of diagram translation into code that allow partially automating the CAS software.

The study subject is the Unified Modeling Language (UML) used for the business-processes description at all stages of the CAS design. To present the diagrammatic models in the theory of visual languages two basic syntactic models called spatial and logical models are used. The spatial model contains relative or absolute coordinates of graphic objects. The use of such model to check and analyze structural, i.e. topological (syntactic) features of diagrams is very difficult. Therefore, to describe the syntax of the visual languages of diagrams the logical model is used. This model is processed by the graphical grammars.

John L. Pfaltz and Azriel Rosenfeld have proposed two-dimensional pattern generating web-grammars [1]. The proposals generated by web-grammars are directed graphs with symbols at their vertices ("webs"). Zhang has offered the positional graphical grammar [1], later it has been developed by Costagliola [2]. It is related to the context-free grammars. Wittenberg and Weitzman [3] have developed a relational graphical grammar. Zhang and Orgun [4, 5] have described the preserving graphical grammars in their studies.

The mentioned methods have the following shortcomings.

1. Positional grammars developing on the base of plex-structure don't use touchpoints and can't be used for languages the objects of which have dynamically changeable number of inputs\outputs.

2. The relational grammars have poorly implemented the neutralization tools for generating a complete list of errors.

3. There is no semantic checking of text diagrams.

4. The common shortcomings of the above-mentioned grammars are: when designing the grammar for unstructured visual languages the number of rules is increased (if there is no change in the number of the language primitives for describing the all variants without structure, the significant quantity of rules will increase), the complexity of developing the grammar. Furthermore, it takes much time to analyze diagrams (analyzers based on the considered grammars have polynomial or exponential time of analysis).

Table 1 shows the results of analysis of the functionality of modern tools for designing the diagrams. It shows that there is no semantic analysis of the text of the diagrams in all the tools.

TABLE I. COMPARATIVE CHARACTERISTICS OF TOOLS ANALYSIS

Indicator	MS Visio	Visual Paradigm	Rational Software Architect	ARIS Toolset
Checking of syntax model	Partially	Partially	Yes	Yes
Model visualization	Average	Average	Good	Good
Standards	UML, ER	UML, ERD, BPMN, DFD	UML, BPMN, EPC	UML, EPC, ERM, DFD-partially
Dynamic modeling	No	No	Yes	Yes
Optimization model	No	No	No	Yes
Cost-effectiveness analysis	Bad	Bad	Average	Good
Software code generation	No	No	Yes	No
Usability	High	High	Low	Low
Availability	High	High	Low	Low
Group working	Yes	Yes	Yes	Yes
Checking of syntax errors	No	No	Yes	Yes
Checking of semantic errors	No	No	No	No

III. ANALYSIS AND CHECKING OF DIAGRAMMATIC MODELS

To analyze and check models is proposed on the base of finite-state RVM-grammars:

$$\mathrm{RVM} = (\mathrm{n}, \Sigma^{\mathrm{n}}, \mathrm{RV}^{\mathrm{n}}, \mathrm{r}_0). \quad (1)$$

where n is a grammar index; Σ^{n} is an alphabet of the n-th grammar; RV^{n} is a set of the n-th grammar rules; r_0 is an axiom of upper-level grammar.

As one of the states RV^{i}grammar contains RV^{j}grammar. RV^{j}grammar can also be compound:

$$\mathrm{G} = \left(\mathrm{V}, \Sigma, \tilde{\Sigma}, \mathrm{R}, \mathrm{r}_0\right). \quad (2)$$

where RV^{n}– the L (R) language grammar is an ordered set of five nonempty sets, where $\mathrm{V} = \{\mathrm{v}_{\mathrm{l}}, \mathrm{l} = \overline{1, \mathrm{L}}\}$ is an alphabet of operations over the internal memory; $\Sigma = \{\mathrm{a}_{\mathrm{t}}, \mathrm{t} = \overline{1, \mathrm{T}}\}$ is a terminal alphabet of the graphical language (the set of the graphical language primitives); $\tilde{\Sigma} = \{\tilde{\mathrm{a}}_{\mathrm{t}}, \mathrm{t} = \overline{1, \tilde{\mathrm{T}}}\}$ is a quasi-terminal alphabet extending the terminal alphabet. The alphabet includes: quasi-terms of graphic objects that won't continue analysis; quasi-terms of graphic objects with more than one input; quasi-terms of links – marks with specific semantic differences defined for them; quasi-term for completing the analysis; $\mathrm{R} = \{\mathrm{r}_{\mathrm{i}}, \mathrm{i} = \overline{1, \mathrm{I}}\}$ is the scheme of the grammar G (a set of names of complexes of rules, where each complex r_{i} consists of a subset P_{ij} of rules $\mathrm{r}_{\mathrm{i}} = \{\mathrm{P}_{\mathrm{ij}}, \mathrm{j} = \overline{1, \mathrm{J}}\}$); $\mathrm{r}_0 \in \mathrm{R}$ is an axiom of RV-grammar (the name of the initial complex of rules), $\mathrm{r}_{\mathrm{k}} \in \mathrm{R}$ is a final complex of rules.

The rules of $\mathrm{P}_{\mathrm{ij}} \in \mathrm{r}_{\mathrm{i}}$ are given as:

$$\tilde{\mathrm{a}}_{\mathrm{t}} \xrightarrow{\mathrm{W}_{\vartheta}(\mathrm{V})} \mathrm{r}_{\mathrm{m}}. \quad (3)$$

where $\mathrm{W}_{\vartheta}(\mathrm{V})$ – n-ary relation defining the form of operation on the internal memory depending on $\vartheta \in \{0,1,2,3\}$; $\mathrm{r}_{\mathrm{m}} \in \mathrm{R}$ is the name of the receiver-rules.

The internal memory consists of stacks for processing the graphic objects that have more than one output, and elastic tapes for processing the graphic objects that have more than one input. The TRVM-system receives the symbols of terminal alphabet from input tape and transmits them to the appropriate level. The elements that transmit the grammar to another level are called "subterms". Then the RV^{n}-grammar description is given as:

$$\mathrm{G} = \left(\mathrm{V}, \Sigma, \tilde{\Sigma}, \bar{\Sigma}, \mathrm{R}, \mathrm{r}_0\right). \quad (4)$$

where $\bar{\Sigma}$ is a set of subterms, i.e. grammar elements that transmit the automat to the next lower level. The rules containing subterms are the following:

$$\bar{\mathrm{a}}_{\mathrm{t}} \xrightarrow[\mathrm{r}_0^{\mathrm{n}+1}]{\mathrm{W}_{\vartheta}(\mathrm{V})} \mathrm{r}_{\mathrm{m}}^{\mathrm{n}}. \quad (5)$$

where $\mathrm{r}_0^{\mathrm{n}+1}$ is a receiver-rules – the initial complex of next level grammar, $\mathrm{r}_{\mathrm{m}}^{\mathrm{n}}$is a receiver-rules to which the transition is done when $\mathrm{r}_{\mathrm{k}}^{\mathrm{n}+1}$.

The RVM-grammars of diagrams of UML-language activity is given in Table 2.

In collective design of CAS it is important to check the ontological consistency of the complex of designed diagrams. This type of errors is semantic. Therefore, to analyze the semantic correctness the multi-level grammar is proposed to be used [6, 7]. The upper level of RVM- grammar is a grammar of use case diagrams. During the analysis the semantic information about the domain field is stored as a graph. In a graph the semantic concept is textual information, presented by the blocks. The analysis shows that the semantic relation is formed among the blocks. For each new diagram the conditions of consistent

expansion (a new concept shouldn't conflict with existing concepts in a subject domain) are checked.

The proposed method allows a learning system to diagnose the following semantic errors: Large synonyms, Object's Antonyms, Conversion of links. Furthermore, the RVM-grammar identifies other four errors: Infinite cycle, Circular links, Multiple links, Synchronous call until get answer, Error of absence of context (Table 3).

IV. RESULTS

We review industrial areas that have an interest in application tools to check and diagnose business process and workflow as diagrams. So, we define BPML, UML, IDEF0, IDEF3, Petri net, and extract 20 types of errors that may occur into them.

TABLE II. THE TABULAR FORM OF THE GRAMMAR OF UML-DIAGRAM ACTIVITY [9]

№	Complex	Quaziterm	The complex-the successor	RV-relation
1	r_0	R_I	r_3	$W_1(i^{t(1)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
2		R_E	r_4	$W_1(i^{t(1)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
3		R_G	r_5	$W_1(i^{t(1)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
4		$\underline{R_I}$	r_6	$\emptyset$
5		$\underline{R_E}$	r_7	$\emptyset$
6		$\underline{R_G}$	r_8	$\emptyset$
7	r_1	R_A	r_2	$\emptyset$
8		R_G	r_5	$\emptyset$
9	r_2	C	r_0	$\emptyset$
10	r_3	C	r_0	$\emptyset$
11	r_4	C	r_0	$\emptyset$
12	r_5	C	r_0	$\emptyset$
13		A	r_1	$\emptyset$
14	r_6	C	r_0	$W_1(i^{t(2)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
15	r_7	C	r_0	$W_1(i^{t(2)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
16	r_8	C	r_0	$W_1(i^{t(2)})/\ W_3(i^{t(1)} \equiv \emptyset \cup i^{t(2)} \equiv \emptyset)$
17		A	r_1	$\emptyset$

TABLE III. THE ERROR TYPES OF UML DIAGRAMS [9]

№	Type of error	UD	AD	SD	CD	DD
1	Lack of link	+	+	+	+	+
2	Error of transfer of control					
3	Error of multiplicity of inputs		+			+
4	Error of multiplicity of outputs		+			
5	Invalid link	+	+	+	+	+
6	Error of link	+	+			+
7	Error of access level				+	
8	Error of message transfer		+	+		
9	Error of delegation of management				+	
10	Quantitative error of diagram elements		+			+
11	Excluding links of the wrong type				+	
12	Call to the lifeline			+		
13	Unbounded links	+	+	+	+	+
14	Violation of multiplicity of dependencies	+	+			
15	Mutually exclusive relations	+				
16	Multiple links	+				
17	Infinite cycle	+				
18	Circular links	+				
19	Synchronous call until get answer			+		
20	Error of absence of context		+			

The table uses the following abbreviations: UD the use of diagram; AD- activity diagram; SD – sequence diagram, CD-class diagram, DD – deployment diagrams.

Modern tools can only check 16 types of errors. Thus, we develop a new RVM-grammar to check and fix all type of errors, which has a linear require of time to analysis oppositely with other grammars with exponential and polynomial requires of time. Figure 1 shows the efficiency. Graphic objects are graphic figures as a circle, a rectangle, a rhombus, a square, a line and etc. We propose a formula [6-8] for calculation of the efficiency that can be written as:

$$\text{Required time} = c \cdot L_s, L_s = \sum_{i=1}^{m} \left(\sum_{j=1}^{V_i} v_{in_{ij}} + \sum_{j=1}^{V_i} v_{out_{ij}} \right) + \sum_{i=m+1}^{t} \left(V_i + \sum_{j=1}^{V_i} v\_out_{ij} \right) + \text{no\_label}. \quad (6)$$

where c – the constant of realization of algorithm, which determines a quantity of time (operators) that are spent to analysis one graphic object; L_s – the number of graphic objects; V_i – the number of graphic objects of i-type; v_in_{ij} – the number of inputs to j-graphic objects of i-type; v_out_{ij} – the number of outputs from j-graphic objects of i-type; t – a total of object types; m – a quantity of object types that have more than one output.

I. Conclusion

The proposed grammars allow a learning system to extend the class of diagnosed errors of diagrammatic models and organize semantic checking of text of diagrams. Authors developed RV- and RVM-grammars for the visual languages which are actively used in practice of design of CAS: IDEF, eEPC, BPMN, SharePoint. The further directions of the researches are: development and research of temporal RV-grammars for workflows in real time, development of methods of the semantic analysis for the specified languages, development of the translating RV-grammars.

The authors have developed the tools for processing the above-mentioned models used in the CAS design [8].

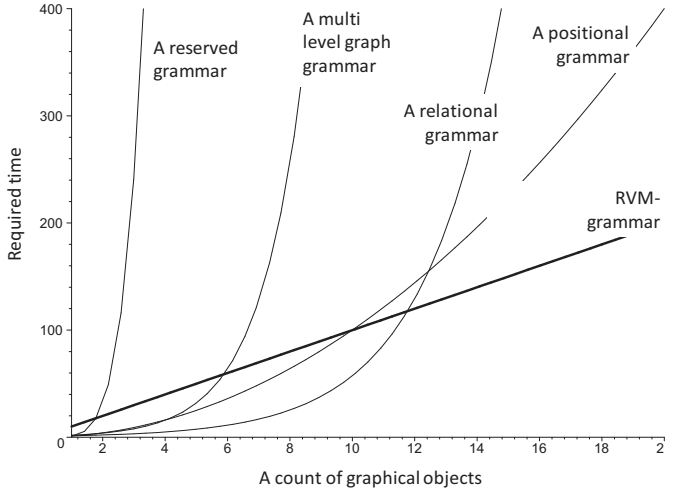


Fig. 1. Efficiency of checking and fixing errors with the help RVM-grammar into diagrams

ACKNOWLEDGEMENT

The authors acknowledge that this paper was supported by the Russian Fund of Basic Research RFFI-15-07-08268a.

References

[1] K. Fu, Structural methods of pattern recognition. Moscow: Mir, 1977.

[2] G. Costagliola, A.D. Lucia, S. Orece, G. Tortora, “A parsing methodology for the implementation of visual systems,”. *(http://www.dmi.unisa.it/people/costagliola/www/home/papers/method.ps.gz.)*

[3] K. Wittenburg, L. Weitzman, “Relational grammars: Theory and practice in a visual language interface for process modeling,”, 1996. *(http://citeseer.ist.psu.edu/wittenburg96relational.html)*

[4] D. Q. Zhang, K. Zhang, “Reserved graph grammar: A specification tool for diagrammatic VPLs,” IEEE Symposium on Visual Languages, pp. 284-291,1997.

[5] K. B. Zhang, K. Zhang, M. A. Orgun, “Using Graph Grammar to Implement Global Layout for A Visual Programming Language Generation System,”, 2002.

[6] O.G. Sharov, A.N. Afanas'ev, “Syntax-directed implementation of visual languages based on automaton graphical grammars,” Programming and Computer Software, vol. 6, pp. 56-66, 2005.

[7] A.N. Afanasyev, R.F. Gainullin, “Development of methods and tools for analyzing and monitoring the diagrammatic business-processes in the design of automated systems,” Congress on Intelligent Systems and Information Technology IS & IT'14, vol. 4, pp. 159–165, 2014.

[8] A.N. Afanasev, R.F. Gainullin, O.G. Sharov, “Program system for visual languages analysis,” Software & Systems, vol. 3, pp. 138-141, 2012.

[9] A.N. Afanasyev, N.N. Voit, E.Yu. Voevodin, R.F. Gainullin, “Control of UML diagrams in designing automated systems software,” Proceedings of The 9th IEEE International conference on Application of Information and Communication Technologies: AICT – 2015, pp. 285-288, 2015.

Improving Emergency Management DSS through the CAP Protocol

The case study of the Italian National Fire Service

Stefano Marsella
Ministry of the Interiors - National Fire Service
Rome, Italy
stefano.marsella@vigilfuoco.it

Marcello Marzoli
Ministry of the Interiors - National Fire Service
Rome, Italy
marcello.marzoli@vigilfuoco.it

***Abstract* — DSS is commonly intended as a computer-based information system that supports business or organisational decision-making activities. Such definition implies the capacity of a DSS of analysing and processing data generated or communicated by multiple sources. A DSS developed to help a civil protection or a fire service Authority should be fed by data and information provided not only by the citizens to emergency numbers, but also from any other organisation involved in the rescue process as well as by available sensors networks, from simulation tools using such data and from the wealth of information provided by GIS data services. Experiences gathered in the course of recent emergencies involving either large areas or very high numbers of people have shown that, even in recent years, the coordination of rescue activities rarely, if not never, was able to take advantage from ICT tools. The main obstacles to data exchange are political attitudes and lack of interoperability services. As a consequence, whenever an uncommon scenario demands such data exchange, the resulting political pressure brings to either exchange data anyway, through improper (and potentially risky) means, or to avoid such data exchange (and miss the related advantages). The sole possibility to overcome such situation is to reach an agreement between the different authorities, aimed at converting and exchanging data in a common protocol, which can be read by non homogeneous systems. Such solution has been tested in Italy in several real-life situations: L'Aquila earthquake (2009), summer forest fire season (2009), Venice local emergency management (2008 – today) and in several well aimed exercises in Europe and U.S.A.**

***Index Terms*—Interoperability, CAP, Emergency management**

I. Introduction

Even if does not exist a standard definition of DSS, it is commonly intended as a computer-based information system that supports business or organizational decision-making activities. When applied to daily or large scale emergencies, such definition implies the capacity of a DSS of analyzing and processing data generated or communicated by multiple sources. In more practical terms, a DSS developed to help a civil protection or a fire service Authority should be fed by data and information provided not only by the citizens to emergency numbers, but also from any other organization involved in the rescue process as well as by available sensors networks, from simulation tools using such data and from the wealth of information provided by GIS data services. The available technologies are adequate enough for developers to deliver even complex systems, however such systems are still rarely adopted due to a main obstacle: the data which could be timely fed to such systems are insufficient in quantity and quality and most often not up-to-date, mostly for both political and technical reasons. Experiences gathered in the course of recent emergencies involving either large areas or very high numbers of people have shown that, even in recent years, the coordination of rescue activities rarely, if not never, was able to take advantage from ICT tools. The main obstacles to data exchange are political attitudes and lack of interoperability services. Most often they are cross-related: on one hand, the extreme care with which emergency data is rightly treated brings most emergency managers at avoiding any exchange of data (e.g., not trusting readily available services able to erase part of the information), on the other hand, due to such attitude there is a lack of properly designed and developed interoperability services aimed at exchanging emergency data. As a consequence, whenever an uncommon scenario demands such data exchange, the resulting political pressure brings to either exchange data anyway, through improper (and potentially risky) means, or to avoid such data exchange (and miss the related advantages). In most contexts, this issue can pose severe problems, since even if the political pressure is aimed at improving coordination through automatic data exchange, the existing systems cannot be updated in time in order to ensure such functionalities. The sole possibility to overcome such situation is to reach an agreement between the different authorities, aimed at converting and exchanging data in a common protocol, which can be read by non homogeneous systems. Such solution has been tested in Italy in several real-life situations: L'Aquila earthquake (2009), summer forest fire season (2009), Venice local emergency management (2008 – today) and well as in several well aimed exercises in Europe and U.S.A., up to be adopted and implemented from several major European emergency management systems.

II. Interoperability functionalities: L'Aquila earthquake, provisional measures for cultural heritage

On April, 6th 2009 the Italian city of L'Aquila and the surrounding area have been striken by a 6,3 Mw earthquake, causing 309 victims, more than 1.600 injured and 10 billion euro of damages. The Italian National Fire Corps responded

swiftly, bringing in place some 1.000 professional rescuers within the first 24 hours, raised to more than 2.300 within the third day, together with some 1.100 vehicles and the needed resources and logistics. Of course the first and foremost target was to save lives, but soon after this task had been completed it was clear the urgency to deploy provisional measures for buildings to restore minimal safety conditions and avoid further damages.

L'Aquila was not a common town: besides the 73,000 civil buildings (half of which damaged), there were more than 600 registered monuments to save (172 of which damaged). More than 100 expert engineers of the Italian National Fire Corps have been working daily to assess civil buildings damages, but monuments required a more complex approach: firemen and their engineers had to work in team with cultural heritage experts provided by several Italian universities under the coordination of the Cultural Heritage Ministry. In fact, the design of provisional measures of each monument required several high-level expertises, as well as the practical approach of firemen, to adapt the design to an often compromised scenario. Such activity has been developed on a long term basis (it lasted more than one year). As a result, the involved professionals were periodically rotated: while firemen teams rotated with a week-long shift, the university teams could not always stay in place. A tool to work remotely was needed. Luckily, at that time the Italian National Fire Corps was testing the first release of the interoperability functionalities for the 100 provincial Control Centres.

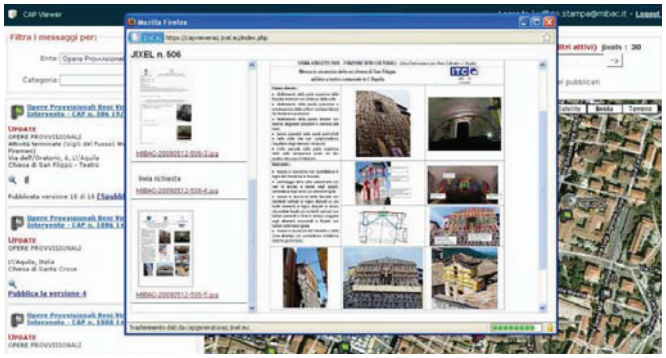


Fig. 1 Web page showing the interface used to exchange data between Cultural Heritage experts and Firefighters working on the damaged listed buildings.

This first release was based on a centric approach, with web interfaces accessing a common datastore. The access to the system, allowed both the firemen team leader and the cultural heritage experts to share information on the same monument, through the access to a common log as well as through upload and visualization of images and drawings (see Fig. 1).

III. Interoperability functionalities: Venice, daily emergency management.

Since then the Italian National Fire Corps developed and tested more advanced, flexible and adaptable interoperability functionalities, which allow the implementation of operational data exchange agreements between the 100 Control Centres and other emergency-related organizations.

In 2011 the Fire Corps HQs of Venice (Venezia) was the first at signing an agreement with the Municipality of Venice to exchange operational data (in CAP format) with the civil protection service, to share in real-time information regarding emergency calls status, as well as the response to incidents qualified as civil protection emergencies. The resulting common operational picture enhances the awareness of the actual citizens' needs, so as to allow an optimized allocation of resources, even if belonging to different organizations.

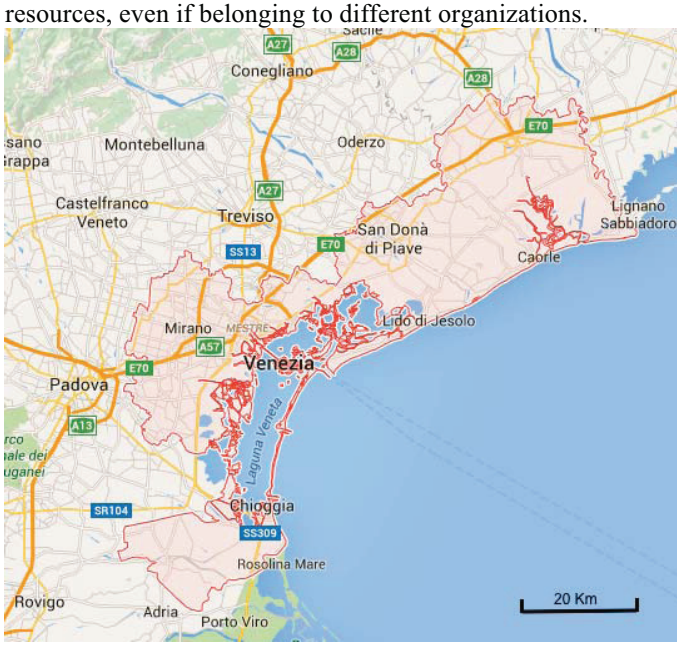


Fig. 2 The Venice Area involved in the agreement.

More recently the same Fire Corps HQs of Venice was closing some more agreements, in order to reach a better operational coordination through the use of the protocol with different Authorities of the area surrounding the city (see Fig. 2):

1) with the municipality of Portogruaro as leader of the "civil protection district of Portogruarese" (11 municipalities in the northern part of the province), with the same targets adopted in the agreement with the Municipality of Venice;

2) with the *Consorzio di bonifica del Veneto Orientale* (Land reclamation and drainage authority for the east Veneto region) to share in real time the status of call and incident concerning floods, so as to allow a faster optimization of water management (the area is reclaimed, so that the water flow is mainly managed through water pumps).

3) with *Autovie Venete* (managing part of the regional motorways network) to share in real time the status of call and incident concerning road accidents or anyway impacting on the network.

Since 2014 the Control Centres of the 10 HQs of Calabria and Puglia regions automatically send the data they have concerning forest fires to the Control Centre devoted to the management of the systems for remote detection and monitoring of forest fires.

In the course of 2016 the Fire Corps Regional Directorate for the Puglia region is finalizing a framework agreement with the Civil Protection Service of the Puglia Region, to allow real-time operational data exchange between the six interested control centres (in CAP format). In the same period, the Civil Protection Service of the Sicilia Region has started to work on an agreement with the Italian National Fire Corps for the same purposes. It should be noted that the emergency management system recently acquired by Sicilia Region allows the issue of public alerts in voice on radio frequencies reading CAP messages through text-to-speech services.

IV. Interoperability functionalities: European State of the Art and Future Perspectives

The Italian National Fire Corps interoperability initiatives took the move in 2008 from the FP7 European R&D project named REACT [1]. Since then, the participation of the Fire Corps to several other R&D projects contributed actively to the selection of realistic interoperability scenarios and to the analysis of technologies readily available to answer to the scenario requirements through the design of more advanced solutions and the consequent update of the functionalities implemented into the Fire Corps Control Centres emergency management system. Thanks to these implementations the system can now satisfy different technical and operational requirement maintaining its performance level. Of course, some of the scenarios, while most interesting, have not been implemented yet. This is the case of the scenario tested within the framework of the EU R&D project IDIRA [2] in 2012, into the Dresden premises of the German Red Cross, which, between others, was focused on data exchange between different systems designed to restore family links in the aftermaths of disasters (adopting the EDXL-TEC standards), e.g., Restoring Family Links [3] from the International Red Cross service and the Google Person Finder service [4]. Both these services were developed soon after the 2005 Hurricane Katrina and the 2004 Indian Ocean earthquake and tsunami, but are not able to exchange their data. Adding to the lot the numerous systems to register victims used by shelters, hospitals and rescuers, it become clear the potential impact of a widespread adoption by such systems of a common standard to reunify families while respecting applicable privacy rules.

An important side effect of the participation to European R&D projects was the effort spent into dissemination activities, to participate to thematic workshop, exchange ideas and inform other rescuers and developers on the opportunities offered by an actual implementation of interoperability functionalities. For sure, this effort contributed to the strategic decisions taken by other European stakeholders which have introduced (or have been introducing) similar functionalities into their renovated emergency management systems. E.g., this is the case the new management system for the European emergency number 112 developed in Portugal at National level [5], which adopts the CAP protocol with the aim of exchanging data with the Civil Protection, as well as of the management system for the European emergency number 112 to be developed in France for the Prefecture of Paris, which included too the CAP protocol into the related tender just published [6].

In the meanwhile, the adoption of CAP standards is spreading further on, e.g., Canada and U.S.A. are working hard on a multi-annual program named 2011 U.S. – Canada Beyond the Border Action Plan [7], aimed at raising the resilience level of their response systems near the common border, even through a series of exercises targeted to assess the interoperability level reached. As it states the final report [8] of the third edition of the Canada-U.S. Enhanced Resiliency Experiment (CAUSE) data interoperability between the emergency management systems of the many organizations involved adopts the CAP standard.

The introduction of services for real-time operational data exchange in the framework of humanitarian aid and civil protection international initiatives, as the United Nations OCHA the European Union Civil Protection Mechanism is another issue considered in managing the CAP protocol. A specific scenario has been proposed and tested in the framework of the EU R&D projects IDIRA (Interoperability of data and procedures in large-scale multinational disaster response actions – FP7 - 2011) as well will be tested in the course of specific exercises in Italy, Austria, Germany and Greece in the project NEXES (NEXt generation Emergency Services - H2020 - 2015) [9].

A comprehensive approach to the problem has been adopted by the recent Decision of the European Parliament and of the Council of 17 December 2013 on a Union Civil Protection Mechanism [10], which promotes the development of civil protection assistance intervention modules, consisting of the resources of one or more Member States. The aim of the Decision is reaching a fully interoperability, in particular with other modules, whatever their National provenance. When testing the compliance to the interoperability requirements, the European Civil Protection exercises mainly focusses on "hardware" interoperability, e.g., between base-camp power plugs and systems, or water networks and connectors, but could easily include "software" interoperability to allow data exchange services between control centres (in CAP format).

V. Conclusions

The targeted scenarios proposed and tested replicate difficulties and obstacles experienced by international (or simply non-local) rescue teams in coordinating their efforts with the emergency services in place. Notwithstanding the efforts spent to improve the civil protection systems' preparedness, the stress of actual emergencies prevents a thorough application of agreed procedures and is intrinsically limited by the number of local rescuers properly trained. On the other hand, properly designed interoperability services could allow the use of the same emergency management systems utilized for daily emergencies to exchange operational data in real-time to coordinate the activities of international rescue teams too. The widespread, worldwide adoption of Common Alerting Protocol – CAP standard could allow the success of a further initiative of the Italian National Fire Corps: the introduction of services for real-time operational data exchange in the framework of

humanitarian aid and civil protection international initiatives, as the United Nations OCHA the European Union Civil Protection Mechanism. In fact, such services could greatly support efficiency and usefulness of any international assistance initiative, allowing improved and smoother coordination of the civil protection module providing assistance with the emergency services of the host Nation, as well as with the parent services at home.

REFERENCES

[1] Cristaldi, M., Delprato, U., & Marzoli, M. (n.d.). CAP in the IST project REACT. Retrieved from http://www.wmo.int/pages/prog/www/ISS/Meetings/WIS-CAP_Geneva2008/DocPlan.html

[2] IDIRA project: Interoperability of data and procedures in large-scale multinational disaster response actions. (2011). Retrieved June 12, 2016, from http://www.idira.eu/

[3] Restoring Family Links. (n.d.). Retrieved June 12, 2016, from http://familylinks.icrc.org/en/Pages/home.aspx

[4] Google Person Finder. (n.d.). Retrieved June 12, 2016, from https://google.org/personfinder/global/home.html

[5] 112.pt. (n.d.). Retrieved June 12, 2016, from http://www.112.pt/Paginas/Home.aspx

[6] Ted - tenders electronics daily. (n.d.). from Retrieved http://ted.europa.eu/udl?uri=TED:NOTICE:175185-2016:TEXT:FR:HTML

[7] *UNITED STATES – CANADA BEYOND THE BORDER: A Shared Vision for Perimeters Security and Economic Competitiveness*. (2011). Retrieved from http://www.whitehouse.gov/sites/default/files/us-canada_btb_action_plan3.pdf),

[8] Cotter, D., Gusty, D., Marc, C., Booth, J., Thomas, J., Cohen, S. E., … Amoabeng, M. (n.d.). *Canada-U . S . Enhanced Resiliency Experiment Series "Cause III": Northeastern Experiment after Action Report*. Retrieved from http://www.crhnet.ca/sites/default/files/library/Joint CA-US CAUSE III EAST Report Final.pdf

[9] Nexes - NEXt generation Emergency Services. (n.d.). Retrieved June 12, 2016, from http://nexes.eu/

[10] Decision No 1313/2013/EU of the European Parliament ad of the Council (2013). Retrieved June 12, 2016, from http://doi.org/10.1017/CBO9781107415324.004

Algorithms in Portable Digital Device UHDTV Panoramic Image Formation

I.S. Konstantinov, S.A. Lazarev
Belgorod State University
Belgorod, Russia
lazarev_s@bsu.edu.ru

K.A. Rubtsov, Y.N. Maslakov
Educational-Scientific laboratory IIiUKS
Belgorod State University
Belgorod, Russia
rubtsov@bsu.edu.ru

***Abstract*—*In this article, the review of components and algorithms which were used by the authors in the development of portable digital device UHDTV panoramic image formation. The system is based on array of 12 light-sensitive sensors with lenses arranged in the center of the facets of a dodecahedron. Primary image processing is performed with the use of designed by the authors, on the algorithms PLD (programmable logic devices). Image processing includes the removal of distortions, masking, calibration, image formation on the sphere describing the dodecahedron and obtaining equidistant projection. The principles and features of their work are described.*

***Index Terms*—*panoramic images, regular dodecahedron, digital photography, digital video.*

I. Introduction

The issue of a single panoramic image formation from multiple overlapping images captured by different cameras is the combination of their common fragments and "bonding" line correction. One common method of combination is based on the determination of singular points found at both adjacent images, which is an alternative to the approximation of geometric object boundaries, located on two adjacent images [1]. However, in a portable device this method is difficult using. This is due to high demands on the processing unit. Modern systems of panoramic photography and video are being treated, as a rule, on the basis of an array of light-sensitive sensors, with different ways of positioning in space. The known spatial arrangement of cameras and their characteristics allows simplify the process of forming a panoramic image [2, 3].

II. Problem Determination

Development of algorithms for forming a panoramic image 12 of light-sensitive sensors, their optimization in order to reduce the computational load on the processing unit and the maximum utilization of PLD in the pretreatment step and the image forming equidistant projection.

III. Materials and Methods

The authors conducted a study task was to create a high-quality portable system panoramic photo and video shooting using a 12 light-sensitive sensors and the formation of the video standard Ultra HD 4K. According to the simulation results, it was found that the optimal arrangement of 12 the light-sensitive sensors in conjunction with lenses is their location in the center of the dodecahedron faces. In this case, part of the image formed on each photosensitive lens sensor is unique (not repeating other sensors) as part of the panorama of a regular pentagon. This design allows you to create a panorama of 360-degrees.

The problem of projecting high quality panoramic images from 12 cameras of the sphere, in real time, requires large computing resources, which makes difficulties the practical implementation in a compact device. This problem is solved by the authors by means of construction and optimization, by reducing the computational procedures, methods of forming the panoramic image on the basis of a mathematical model of a dodecahedron and obtaining transform coefficients from the information of each light-sensitive sensor. In the investigations it was found that the optimal ratio of the sides of light-sensitive sensor is 4:3. The use of different lenses in the system has shown the need to eliminate the distortion, and to compensate for possible deviations from the ideal placement of camera models in the mathematical model are calibration parameters for each light-sensitive sensor. Calibration parameters take into account the linear displacement cameras from the center of the dodecahedron and turns. Calibration of the system and eliminating distortion of 12 cameras (at this stage of software development) is performed automatically using a specially calibration table. As a result of the calibration coefficients of the matrix are formed to perform fast image sampling correction and formation of spherical panoramas.

To solve the problem of forming a panoramic image in a portable device, the authors following algorithms have been developed:

- correction algorithm for the distortion of the input frame;
- algorithm input frame conversion to the pentagon mean in local coordinates;
- algorithm for constructing a model of a dodecahedron and conversion of local input frame coordinates in the spatial coordinates;
- algorithm model building sector and the conversion of the spatial coordinates of the dodecahedron in spherical coordinates;

- algorithm for constructing an equidistant projection of the sphere onto a plane.

A. Correction algorithm for the distortion of the input frame

The algorithm implements the procedure for compensation of distortion caused by the camera lens. The main stages of the algorithm [4]:

- initially receives the image obtained by the light-sensitive sensor, and the numerical values of the coefficients of distortion;
- for each pixel of the input image is calculated taking into account the new situation of specific coefficients of distortion;
- in order to eliminate edge effects produced crop image distortion compensated.

B. Algorithm input frame conversion to the pentagon mean in local coordinates

The algorithm is designed to determine the affiliation of each pixel of the input frame is inscribed in his pentagon:

$$\mathbf{P} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \bigcup_{i=1}^{12} \mathbf{P}_i \,, \qquad (1)$$

where **X, Y, N –** the sets (arrays) pixel values of the coordinates 12 of light-sensitive sensors, in the form of regular pentagons, linked to the center of the dodecahedron and explicitly given values of calibration coefficients [5].

The main stages of the algorithm can be represented as follows:

- downloading the video frames received for each light-sensitive sensor;
- using the developed system of inequalities for each light-sensitive sensor based on the calibration coefficients determined by membership of a single pixel of the selected light-sensitive sensor inscribed pentagon;
- output of the algorithm is formed by an array of local coordinates.

C. Algorithm for constructing a model of a dodecahedron and conversion of local input frame coordinates in the spatial coordinates

The algorithm implements a procedure for constructing a model of a dodecahedron and conversion of local input image coordinates in the spatial coordinates on the selected faces of a dodecahedron. Algorithm using a mathematical model of a dodecahedron, allows you to perform conversion operation: $\mathbf{P}_n=\mathbf{R}_{z,\Theta}\cdot(\mathbf{R}_{y,\varphi}\cdot\mathbf{p}_n)+\mathbf{P}_{0n}$, where $\mathbf{p}_n$, $\mathbf{P}_{0n}$, $\mathbf{P}_n$ – matrix displacement of local coordinates the n-th light-sensitive sensor relative to the center of mass of the dodecahedron, $\mathbf{R}_{y,\varphi}$, $\mathbf{R}_{z,\Theta}$ – array of 12 rotation matrix.

Main stages of algorithm can be represented as follows:

- input data for the algorithm is the numerical value of the length of the rib fits into the pentagon frame, and an array of local coordinates of pixels belonging to the pentagon inscribed in a frame;
- transfer is made to the origin of system to all faces of a dodecahedron;
- computes spatial coordinates of image of each light-sensitive sensor.

D. Algorithm model building sector and the conversion of the spatial coordinates of the dodecahedron in spherical coordinates

The algorithm performs a procedure for constructing a model of a sphere and the conversion of existing array of spatial coordinates in spherical coordinates to obtain a display of the dodecahedron in a sphere:

$$\mathbf{S} = \begin{pmatrix} R \\ \Xi \\ \Psi \end{pmatrix} = \bigcup_{i=1}^{12} \begin{pmatrix} c \cdot \left(1+\sqrt{5}\right)\sqrt{3}/4 \\ \arctan\left(\sqrt{X_i^{\,2}+Y_i^{\,2}}\,/Z_i\right) \\ \arctan\left(Y_i/X_i\right) \end{pmatrix}, \qquad (2)$$

where X_i, Y_i, Z_i – associated with center dodecahedron coordinate arrays of elements (cells) of the image for *i*–side of dodecahedron, *c*–the face of dodecahedron.

The main stages of the algorithm can be represented as follows:

- initially enter a numeric value length ribs fit into the frame of a pentagon, an array of local coordinates of pixels belonging to the frame inscribed pentagon, and an array of spatial coordinates of pixels arranged on the faces of a dodecahedron;
- for each of the twelve sensors is performed conversion of spatial coordinates of pixels to a spherical coordinate system;
- forming an array index, which establishes the correspondence between identical pixels in the local and spherical coordinate systems.

E. Algorithm for constructing an equidistant projection of the sphere onto a plane

The algorithm implements the procedure for constructing an equidistant cylindrical projection of the sphere on the basis of predetermined angles. The main stages of the algorithm can be represented as follows:

- initially receives an array of pixel coordinates in a spherical system, the array indexes, establishing correspondence between identical pixels in the local and spherical coordinate systems, the numerical values of the scale factors;
- for the scope of each element is calculated its coordinate on the equidistant cylindrical projection;
- based on the coordinates formed equidistant cylindrical projection matrix and the image is thinned to the required dimension.

IV. Conclusion

The algorithms considered in this article used to creation of a spherical panoramic images with 12 cameras and have optimized for implementation on the PLD. This allows you to

process images in real time and generate video UHDTV standard.

ACKNOWLEDGMENT

The study has been conducted under subsidy #14.581.21.0003 (project ID – RFMEFI58114X0003) with the Ministry of Education and Science of the Russian Federation.

REFERENCES

[1] I.S. Konstantinov, B.J. Shuljak, S.A. Lazarev and A.N. Zalivin, "Singular Point Comparisons During Panoramic Image Formation," The International Journal of Soft Computing, vol. 10 (6), 2015, pp. 454-457.

[2] B. Summa, J. Tierny and V. Pascucci, "Panorama Weaving: Fast and Flexible Seam Processing", ACM Transactions on Graphics, vol. 31(4), 2012, pp. 83.

[3] R. Szeliski, Computer Vision: Algorithms and Applications, Springler, 2010, 979 p.

[4] E.G. Zhilyakov, A.A. Chernomorets and E.V. Bolgova, "About Image Decomposition on the Eigenvectors of Subband Matrix," Belgorod State University Scientific Bulletin. History Political science. Economics. Information technologies Series, vol. 15 (186), 31/1, 2014, pp.185-189.

[5] K.A. Rubtsov, G.F. Romerio, "Hyperoperations, for science and technology: New algorithmic tools for computer science," Lambert Academic Publishing, 2011, 185 p.

Implementing Cycling Training Application for Mobile Devices

Andrey Bogdanchikov, Rassim Suliyev, Bekzhan Bolatov
Computing Engineering Department
Suleyman Demirel University, SDU
Kaskelen, Kazakhstan
andrey.bogdanchikov@sdu.edu.kz
rassim.suliyev@sdu.edu.kz
bolatovbekzhan@gmail.com

***Abstract*— This work is focused on researching and developing Android application for cycling training and monitoring. It compares various training applications in similar area and defines some important points. Then thorough implementation concepts are provided using UML diagrams and development tactics. The usage of application is explained, and important device screens are included in the paper. In conclusion it is said about necessity of such applications and importance of staying healthy by attending sport activities.**

***Index Terms*— Android; cycling; training; system; application**

I. Introduction

Today, the hi-tech are more and more domestic applicable.

People tend to use mobile applications that form an integral part of life. With digital tech dawn breaking, healthcare is still of paramount importance. Going in for sports is critical in keeping fit. Bearing the facts in mind, we developed a mobile application for bicycle ride monitoring. The application is designed to keep records of bicycle rider's statistics and to figure a distance covered respective of speed.

The application employs acceleration meter sensors to display a rider's speed. It also makes use of a GPS technology to calculate a distance driven and a rider's location. The application shows a workout outcome to be kept, which will allow a user to schedule things in an orderly fashion.

The application is supported by any Android based mobile facilities. Which can use acceleration meter sensors, also we managed to lessen a power consumption of the mobile facility.

Similar idea is already being used in software engineering textbook by I. Marsic [1]. 6 students developed application using this idea. They develop software solutions for continuous monitoring of person's physical activities and vital signs, and motivating people to be more physically active. It uses GO RUN SYSTEM [2].

II. Aim

Aim is to create android application using gyroscope and accelerometer, which allows cyclists to see progress and update results of training.

Such application provides necessary information for cyclist in neat and user-friendly manner, so that he can use it in everyday training and keep track of all improvements. Seeing which can be extra motivation to stay healthy and fit.

III. Technologies currently in use

For now, customers always believe that the more money they spend the better equipment they would get which could help them to be much healthier.

Nike+ [3], one of the most popular technical equipment, is well developed and widely used. Nike+ is worth of Nike's reputation, accuracy and easy portability are its most unique characteristics.

There are many other application and devices that are oriented to provide environment for jogging and training control. In the project focus was bicycle training.

IV. Project Development

In UML diagrams you can see an architecture and logic of the project. According to the sequential UML diagram of the project, user can see three activities and select of them. First activity is Main page. Main page has two functions such as start and stop. If user presses start, accelerometer and GPS are started and synchronized with database. If user presses stop, accelerometer and GPS are closed, and results are saved in database.

Second activity is Map part. Map part shows current location of user, starting point, finishing point, travelled way of user. Third activity is Results. Results part saves results and shows previous results synchronize with database. View of UML diagram is shown in Figure 1.

UML interaction diagram architecture shows how system works and describes step by step each function. View of interaction diagram of how accelerometer works in my project is shown in Figure 2.

Interaction diagram shows how accelerometer works when user presses start. Most android phones have accelerometer sensors. Theses sensors measure acceleration forces. Acceleration is the measurement of the change in velocity, or speed divided by time.

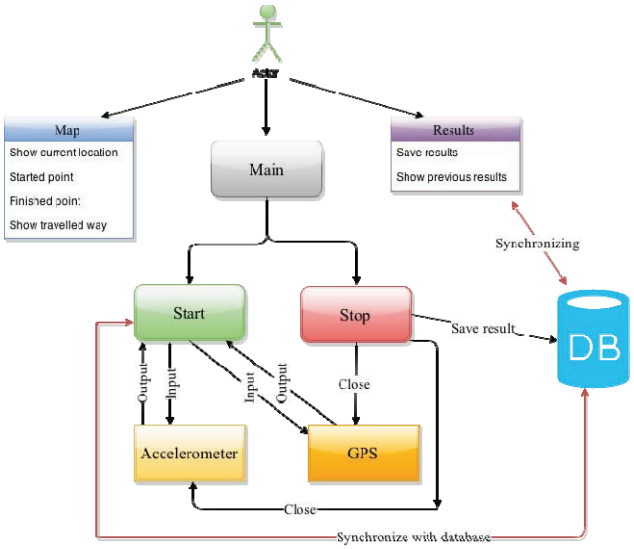


Fig. 1. Use-case UML diagram of project

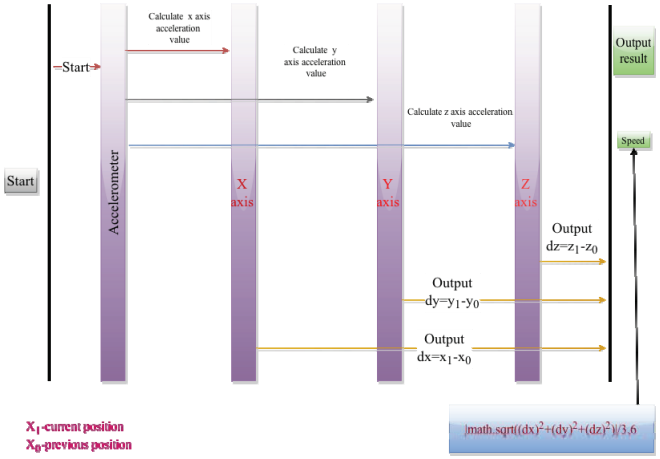


Fig. 2. Interaction diagram of how accelerometer works

Accelerometers that can measure acceleration in one, two, or three orthogonal axes are already available. When accelerometer is started, accelerometer calculates X, Y, Z axis acceleration values. Then it calculates average acceleration at X, Y, Z axis by this formula dx=dx1-dx0 and value x changes to another axis. This formula calculates average acceleration by subtracting previous from the current acceleration. Suffice it to say that it is necessary for some time to take readings, then install your XYZ Accelerometer appropriate dX, dY, dZ. This procedure cannot be neglected, because while we sleep constantly operates the acceleration of gravity, and a sensor to measure it. Accelerometer results are updated every 0,5 seconds. At the end acceleration is converted to speed.

Android devices use the same global positioning technology as Google Maps and most third-party GPS tools do. This allows users to locate themselves on a map, find and navigate to destinations via detailed directions, and search maps using a number of different methods.

GPS is needed to show the location of the user and the distance traveled on a map. If user presses button map, user will see a world map. If user presses on circle icon, then the user will know the approximate location. And to realize GPS I use android map and GPS libraries. GPS works with Internet.

Users can zoom in and out their location on the map. Users can see travelled distance of their trainings on the map. Their travelled distance marked like ‘start’ and ‘finish’ points. And user can see starting and finishing points on the map. (see Figure 3 below)

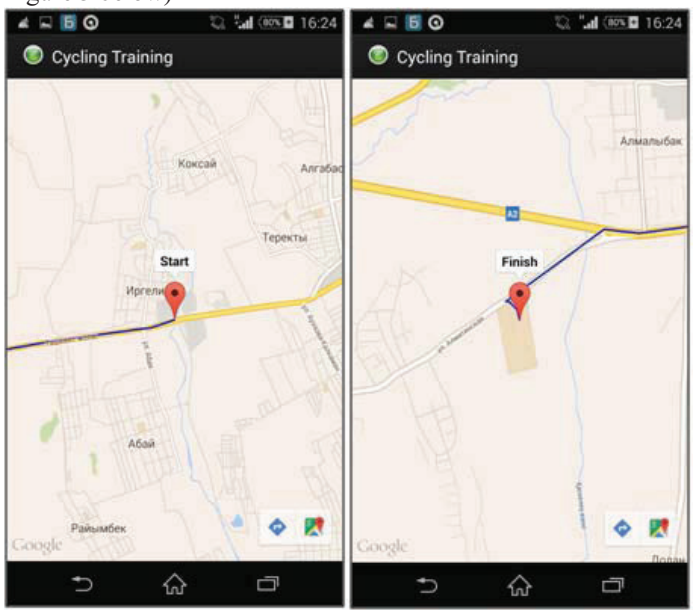


Fig. 3. Start and Finish points

Results part shows results of training in details, such as travelled time, distance and speed, also results of travelled distance on the map. Travelled time is shown in minute and seconds. Travelled distance is shown in meters, and it is calculated by GPS coordinates. Travelled speed is shown in kilometers per hour, and it is calculated by accelerometer sensors of Android Smartphone. View of Result part of application is shown in Figure 4.

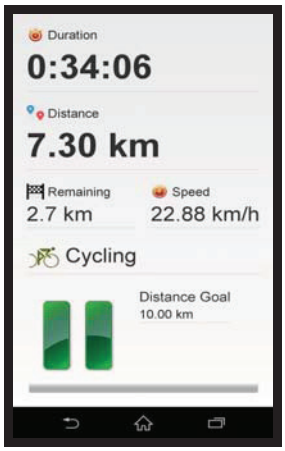


Fig. 4. Result part

Also results were compared with Adidas micoach application. Micoach application is calculating speed by GPS technology each 5-10 seconds. Results of training using Adidas micoach application are shown in Figure 5.

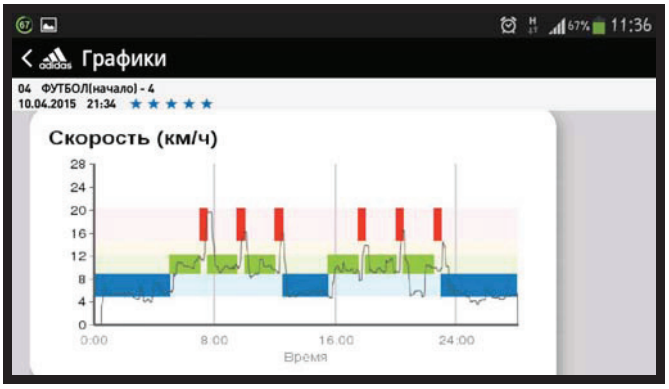


Fig. 5. Speed change on a chart

According to the chart, we see that the official application of the company Adidas, working on GPS for time lags. Nevertheless, on the chart, where user ran evenly and sometimes accelerated by changing the speed, you may notice fluctuations in the speed shown on the chart. Given this fact, we can safely assure you that the fault rate on a bicycle in a 3-5 km / h is not too large fault.

Calculating distance using GPS technologies works normal in this application. GPS starts calculating distance after 15 seconds in application. We tested it by comparing with Adidas official application.

V. Conclusion

The goal of this paper was to develop an application for cycling monitoring on android platform. The paper itself provides the project overview. It starts with some background information about application development and presents features of mobile applications from the point of view of android developer.

It then moves on the description of the design of the application, providing some short justification for application design decisions taken during development. Then the paper describes overall architecture of the application, covers some of the technically challenging or otherwise interesting features that have been implemented and, finally, it is concluded with the testing techniques used in the development.

We hope this application will be useful for people whose like bicycle riding.

References

[1] Ivan Marsic. Software Engineering. Rutgers University, New Jersey. 2012

[2] Angjie Li, Anqi Yan, Prakhar Srivastava, Sang Lin, Siyuan Li, Yadvaindera Sood. GO RUN. Rutgers University, New Jersey. 2013

[3] Saponas TS, Lester J, Hartung C, Kohno T (2006) Devices that tell on you: The nike+ ipod sport kit. Dept of Computer Science and Engineering, University of Washington, Tech Rep

[4] http://cdixon.org/2014/04/07/the-decline-of-the-mobile-web/ "The decline of the mobile web" by [cdixon], 2014.

The Usage of the Mobile Ad-Hoc Networks in the Construction Industry

Igor Konstantinov, Kostiantyn Polshchykov, Sergej Lazarev, Olha Polshchykova
Belgorod State University
Belgorod, Russia
polshchikov@bsu.edu.ru

Abstract— **This article defines the terms "remote under-construction facility" and "hazardous under-construction facility". The up-to-date sci-tech problem of provision for efficient communication at remote and hazardous construction facilities is stated. The suitability of usage of mobile ad hoc networks is proved in this article as well.**

***Index Terms*— *remote and hazardous construction facilities, communication provision, mobile ad hoc networks, construction activity, emergency*.**

I. Introduction

The information exchange at the construction facilities is an issue of a great significance, especially from a perspective of the quality and safety management. The communication process at the construction facilities is based on the completion of the following tasks:

- operational cooperation assurance in the process of construction activity;
- monitoring and well-timed detecting of destructive factors of natural and anthropogenic origin leading to fire and explosion emergencies, destruction, flooding, impact of poisonous substances and radioactive radiation;
- efficient management and coordination of the personnel and resources in the process of emergency response and recovery at the construction facilities in order to minimize human losses and material damage [1].

All these tasks are complicated with a number of issues, which perplex the construction process especially when constructing the remote and hazardous facilities.

II. The specific nature of remote and hazardous construction facilities

A remote construction facility is a building constructed in the distant areas with poorly developed telecommunication infrastructure. Such geographically dispersed objects are being constructed, as a rule, at poorly acquired land areas such as mountains, industrial extracting companies, workers camps, research stations and camps, landing fields, spaceports.

A hazardous construction facility is a building constructed in conditions of destructive and adverse factors of both natural and anthropogenic origin such as fire and explosion emergencies, destruction, flooding, impact of poisonous substances and radioactive radiation. The following geographically dispersed objects could be aggregated to the class of hazardous construction facilities:

- buildings, constructed in the areas with a high level of probability of having an extreme impact of wind, earthquakes, flooding, landsliding and other natural extreme factors;
- tunnels, subway lines and other underground facilities being in the construction process;
- buildings and facilities constructed in the dangerous vicinity of industrial objects, where chemical, radioactive, flammable, toxic substances are produced, extracted, storage, transported and exterminated;
- objects being constructed in the vicinity of mining activity sites or mining excavations.

III. Benefits of the mobile ad hoc networks

In conditions of threats carried by destructive external impact and lack of the traditional telecommunication infrastructure the data transfer networks are of high demand. The system requirements of such networks are fast deployment, autonomous electricity supply for every component, high durability, capacity to transfer and receive messages in conditions of dynamically and rapidly changing topology (occasional processes of transferring, destruction, adding, turning-on and turning-off of the subassemblies). The characteristics mentioned above are incidental to the mobile ad hoc networks (MANET) [2–4]. The main benchmark feature of such networks is a randomly changing structure. The main peculiarities of MANET design are as follows:

- realization of decentralized network components control;
- lack of base stations (fixed subassemblies);
- every subassembly is designed to take on the router functional.

The MANET technology possesses great perspectives to be applied for communication provision at remote and hazardous construction facilities because of the benefits listed above. The usage of a self-organizing network is quite often a single opportunity to organize a decent informational exchange at construction facilities, as the deployment of the traditional

cellular, wired or radio communication is estimated to be laborious and expensive process. More than that, fixed receive/transmit subassemblies being set up make the network insufficiently mobile and durable or, in other words, incapable of effectively functioning in conditions of occasional displacements and extreme external impacts.

IV. The Specifics of the Mobile Ad Hoc Networks Usage at the Remote and Hazardous Under-Construction Objects

The personnel at the remote and hazardous under-construction objects should use compact personal ad-hoc devices which must be provided with autonomous electricity supply elements (rechargeable batteries). The following elements might also be included into the set:

- wireless headset – for communication sessions being held without operator's hands involved;
- surveillance cameras – for video information of the construction process transmitting to the operation-dispatching posts;
- equipment for monitoring of the external conditions at the under-construction facilities: sensors of water, potentially explosive gases, radioactive radiation, poisonous substances, crevices and fractures, ditches and falling objects detecting; sensors for the human body functionality analyzing (pulse, temperature and breathing measuring);
- equipment for network nodes positioning – required in order to get the coordinates of the construction workers in need of help and rescue.

Informational cooperation during the construction of the remote and hazardous objects based on the technology of the self-organizing network allows to conduct the following functions:

- monitoring the process of the construction works;
- remote management of the construction works;
- messages exchange for purposes of technical tasks solving;
- external conditions surveillance at the construction facilities in order to detect destructive factors in time and ensure the safety;
- distant control and coordination of the rescue and response process during the emergencies;
- informing the personnel and men in charge about the spots or coordinates of construction workers and their functional state;
- monitoring and control of the under-construction facilities and of their specific sections during the emergencies liquidation.

V. The Mobile Ad Hoc Networks Usage Issues

Despite the number of significant benefits, MANET has not gain widespread yet. The slowed down implementation tempo could be explained by the impact of the following factors which complicate the information delivery:

- shirt duration of connections as a result of a rapid network topology change and regrouping;
- high level of probability to distort the information due to radio noises waves;
- substantial packet transfer delays determined by the low radio channels' bandwidth;
- significant packets losses as a result of frequent network overloads, which occur due to the nonpermanent traffic [5].

The impact of the factors mentioned above turns up in decrease of the quantity of information and the increase of the information delivery time, which is unacceptable when constructing the remote and hazardous facilities. Plentiful and well-timed information is the basis of the constructors' life and health preserving, as well as the factor of construction damage reduction.

A simulation model was designed in MatLab + Simulink software environment to examine how the dynamic of ad hoc network topology affects information delivery characteristics. Its diagram is shown in the Fig. 1.

The model simulates controlled flow (CF) through ad hoc network channels. CF-packets are transmitted through the channels 1, 2, 3 and 4 which connect sending and receiving

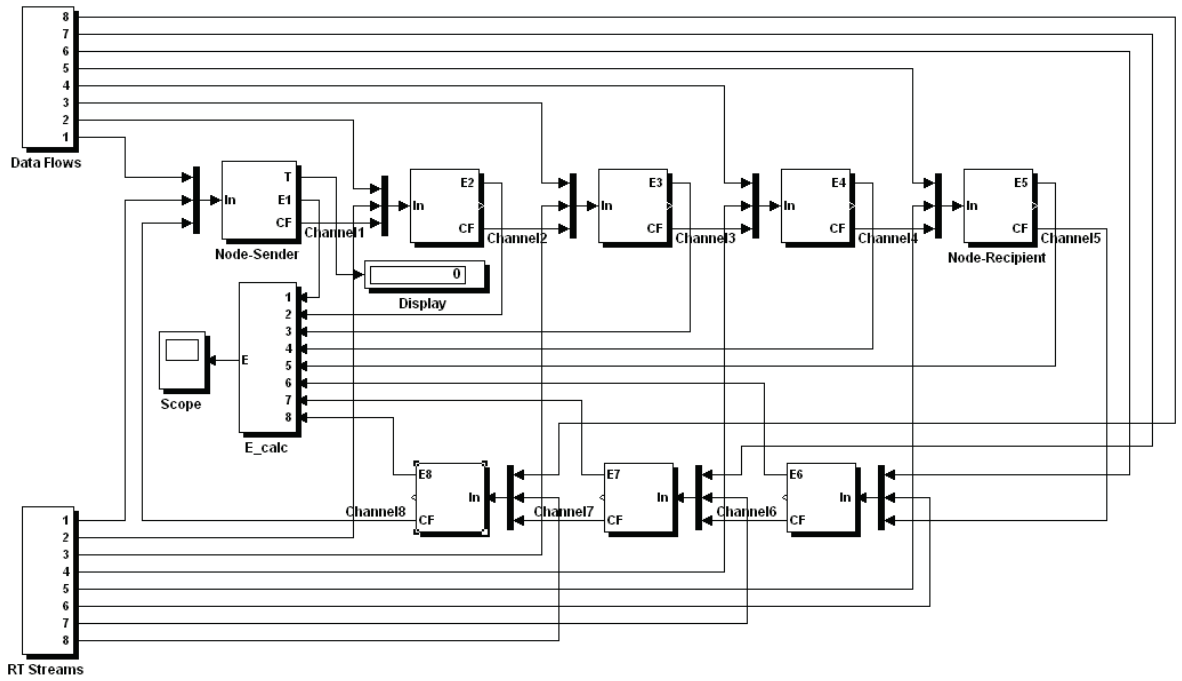


Fig. 1. Block diagram of a simulation model.

node. In the opposite direction, CF-confirmations are transmitted through the channels 5, 6, 7 and 8. The set of channels through which CF-packets and CF-confirmations are transmitted forms a closed cycle of CF-circuit flow.

Besides CF-packets and CF-confirmations, packets of other basic data and real time streams are transmitted. Due to the ad hoc network dynamic topology besides main streams, additional data and real-time streams are transmitted through CF-circuit channels. The presence of additional streams reduces channels bandwidth available for the transmission of CF.

The elements represented in the figure simulate the following processes. “Data Flows” – generating packets of basic and additional data flows; “RT Streams” – generating packets of main and additional real-time streams; “Node-Sender” – CF-Packets formation in the sending node, CF-packets arrival for transmission through the channel 1, arrival of main and additional data and real-time streams for transmission through the channel 1, CF-confirmations reception by sending node, buffering and discarding in the sending node; “Channel 1” – “Channel 4” – CF-packets transmission through CF-circuit channels; “Node-Recipient” is reception of CF-packets by receiving node, CF-confirmation formation in the receiving node, CF-confirmations arrival for transmissions through the channel 5, arrival of main and additional data and real-time packets for transmissions through the channel 5, buffering and discarding in the receiving node; “Channel 5” – “Channel 8” – CF-confirmations transmission through CF-circuit channels; “E_calc” is calculation of the current CF-circuit bandwidth available for the controlled flow transmission.

The element “Display” is used to display CF transmission time. Virtual receptionist “Scope” is designed to display the current values of the CF-circuit bandwidth available for transmission of the controlled flow.

On the basis of the represented model a series of simulation experiments was made in which various scenarios of applying ad hoc network for communication on dangerous construction sites were simulated. The original data contained in the TABLE I was used.

The simulation result revealed that due to the influence of nodes mobility and the destructive factors, time of the controlled flow transmission in an ad hoc network is 25.8% higher than the values of this magnitude in a fixed network topology. This is connected to a decrease of 19.2 % of the available network bandwidth channels with dynamic topology. Represented studies show that the application of ad hoc network messaging significantly reduces the efficiency of messaging in the process of information exchange on remote and dangerous construction sites.

TABLE I. INITIAL DATA FOR SIMULATION EXPERIMENTS

Parameter	Value
Channel bandwidth	1 Mbit/s
Probability of a bit error in a channel	0.00005
Bandwidth required to transmit video	256 Kbit/s
Bandwidth required to transmit voice	128 Kbit/s
Volume of messages transmitted by data flow	1 MB
Number of packets in the data flow	800
Packet sending duration	10 ms
Packet bit length	10 Kbit
Capacity of the channel queue buffer memory	80 KB

Having regard to the above, it can be affirmed, that the efficient communication at the remote and hazardous construction facilities is an up-dated task. Theoretical apparatus must be researched and developed, as well as perfection of the packet information transfer technology used in the conditions of the dynamically changing network topology.

VI. CONCLUSIONS

Currently the problem of establishing an efficient communication network at remote and hazardous construction facilities is of pressing need. The authors’ point of view is that the solution to this issue is the usage of wireless self-organizing networks. Further researches must have the main aim to develop and perfect the packet transfer methods and algorithms, which allow to advance the operativity of the information delivery in the random topology network.

The research has been carried out with the financial support of the Ministry of Education and Science of the Russian Federation (the unique identifier of the project is RFMEFI57815X0138).

REFERENCES

[1] I. Konstantinov, O. Pilipenko, K. Polshchykov and O. Ivaschuk, “The issue of communication in the process of prevention and liquidation of emergency situations at construction sites,” Building and Reconstruction, vol. 1, pp. 40–46, 2016.

[2] S. Basagni, M. Conti, S. Giordano and I. Stojmenovic, Mobile Ad Hoc Networking. IEEE Press, 2004.

[3] I. Konstantinov, K. Polshchykov and S. Lazarev, “Algorithm for Neuro-Fuzzy Control of Data Sending Intensity in A Mobile Ad Hoc Network For Special Purpose,” Journal of Current Research in Science, vol. 4, pp. 105–108, 2016.

[4] K. Polshchykov, “Synthesis of neuro-fuzzy systems of data flows intensity control in mobile ad-hoc network,” Proc. of the 23rd Int. Crimean Conf. “Microwave and Telecommunication Technology (CriMiCo)”, pp. 517–518, 2013.

[5] N. Rvachova, G. Sokol, K. Polshchykov and J. Davies, “Selecting the intersegment interval for TCP in Telecomms networks using fuzzy inference system,” Proc. of the 6th Int. Conf. “2015 Internet Technologies and Applications (ITA)”, pp. 256–260, 2015.

The Effect of the Device Mismatch on Parameters of GaN MMIC phase shifters

Aleksandr Timoshenko, Egor Belousov, Ksenia Molenkamp, Alexander Bakhtin
Telecommunication systems department
National Research University of Electronic Technology (MIET)
Moscow, Russia
timoshenko@edu.miet.ru

***Abstract*—The article presents the results of the design of GaN phase shifter for MIMO-based antennas. The optimization and simulation for each operating frequency 0.1–31.5 GHz was performed along with simulation of effect of capacitance mismatch on phase shifter parameters. The phase shifter is designed as a switchable high-pass filter that provides a 90° phase shift with minimum insertion losses at the operating frequency. Designed phase shifter works in the frequency bands up to 55 GHz**

***Index Terms*—Microwave, device mismatch, monolithic integrated circuits, wide band gap semiconductors, GaN, phase shifters, s-parameters**

I. Introduction

Given the current requirements for the electronic components [1-4], in recent years investigation of wide band gap semiconductors and devices based on them [5–16] have intensified (silicon carbide, gallium nitride, aluminum nitride, and others). GaAs and InP, used in most microwave applications, have a number of disadvantages [17], which contradict with obtaining the desired characteristics of devices for modern 4G mobile cellular communication systems, satellite communications systems, radar and military use [18, 19]. The most promising [20–26] wide gap material is gallium nitride (GaN) which has a band gap of about 3.5 eV, electron mobility and saturation velocity about 2000 $cm^2/V{\cdot}s$ and $2.7{\cdot}10^7$ cm/s and, accordingly, the thermal conductivity of 1.5 W/m·K [27]. The purpose of this article is to describe the design process of one of the basic microwave functional elements such as a phase shifter based on GaN transistors (HEMT) with a minimum gate length of 130/200 nm. Transistor model was verified with the measurement results of the experimental samples of GaN transistors. In addition, the effect of mismatch in capacitance values was considered in order to provide acceptable phase shifter parameters and, if required, compensation.

II. Review of GaN Phase Shifters

The single monolithic ICs of the GaN phase shifters are not frequently mentioned in the available resources, but there are implementations of these electronic components for the C-, X- and Ku-bands. In [28] a two-stage phase shifter was developed for C-band with device size of 30 × 30 mm, consisting of a 3-dB quadrature directional coupler and Wilkinson distributor. The results of measurements and simulations show a fixed phase offset of 45° ± 5° in the frequency band of 4–8 GHz, insertion loss less than 1 dB and more than 12 dB return loss at the central frequency of C-band. In [29] an adjustable voltage shifter was designed on surface acoustic wave (SAW) based on two-dimensional electron gas in AlGaN/GaN heterostructure, The operation of which was provided by changing the conductivity of the electron gas by applying a DC voltage to the structure. This structure is consistent with transistors and devices manufactured on GaN technology for the needs of signal processing and frequency control.

In [30] a highly reliable high power phase shifter was presented for a new type of 0.8 μm GaN HEMT technology, operating in X-band. The proposed phase shifter by 0°/45° with the high/low pass filtering shows low insertion loss of 2.5 dB, a good reflection loss and change in amplitude of not more than 0.5 dB for two-phase state throughout the operating range from 6 to 13 GHz, and the relative performance of phase is relatively constant across the operating band.

Company TriQuint has phase shifters [31] operating in the frequency range of 6–18 GHz (including X- and Ku- band) using 6-bit digital correction that provides operation with continuously adjustable time delay. Insertion loss varies between 3.5–10 dB with a mean square error in the amplitude of the order of 0.5 dB and phase error 3.5–10° [32].

III. Measurement and Correction of the GaN HEMT Transistor Model

A. Small signal transistor model for GaN HEMT

We used available HEMT models [33–39] to create a model for design of GaN IC prototype. The model's parameters were adjusted according to previous measurements of heterostructures of used technology [12, 40] and then corrected according of measurements of experimental transistors with varying width/length. In the development of small-signal model, all distributed elements were considered as lumped. We assumed that the characteristics of the circuit elements are not frequency dependent and the part of the elements can be neglected [41]. Small signal equivalent model for AlGaN/GaN HEMT [42], which includes the parasitic capacitance between all terminals and ground, conventionally

divided into two parts: the internal non-linear and dependent on the bias part, not including the parasitic elements associated with the terminal connections; linear and independent of the bias part, including parasitic elements. If we neglect some of the parasitic capacitances and the resistance transitions to convert the gate-source and gate-drain (R_{fgs} and R_{fgd}), the model can be simplified. It should be noted that the resistance R_i denotes the inside of the recombination of the donor under the gate depletion region; the delay time τ – time during which the depletion region responds to changes in the gate signal parameters.

Let us find the relation between the parameters of the internal components and *y*-parameters. Assuming that the transistor operates in a saturation region, the internal parameters y_{int} can be written as

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} (j\omega C_{GS} \parallel R_i^{-1}) + (j\omega C_{DG} \parallel R_{GD}^{-1}) + R_{fdg}^{-1} + R_{fgs}^{-1} \\ \times -(j\omega C_{GD} \parallel R_{DG}^{-1}) - R_{fdg}^{-1} \\ \frac{g_m e^{-j\omega t}}{1 + j\omega C_{GS} R_i} - (j\omega C_{DG} \parallel R_{DG}^{-1}) - R_{fdg}^{-1} \\ \times R_{DS}^{-1} + R_{fdg}^{-1} + j\omega C_{DS} + j\omega C_{GD} \parallel R_{DG}^{-1} \end{bmatrix}. \quad (1)$$

At low frequencies in the megahertz range the components, including capacitance, in y_{11} and y_{12} (1) tend to zero, therefore, R_{fgs} and R_{fgd} can be defined. The allocation of the real and imaginary parts of the components (1) makes it possible to determine the remaining parameters of the model.

B. Measurement and specification of s-parameters for GaN transistor model

Developed model [43, 44] allows embedding it into a modern CAD systems as a subcircuit and conduct modeling of the main parameters of devices and circuit elements. The results of simulation were compared with the results of experimental measurements at $V_{gs} = -1.5$ V, $V_{ds} = 10$ V. The measurements were performed on a set of experimental samples of multi finger transistors manufactured using nanotechnology complex NanoFab 100 (at the MIET premises), with widths of 0.2 to 2 mm and a gate length of 130 nm and 200 nm, formed by using a focused ion beam. The main dependences coincide, and the parameters variation was negligible. In a preliminary analysis of the obtained values of the resistance of the ohmic contacts was found to be equal to 0.36 Ω·mm (3.99×10^{-5} Ω·cm^2).

IV. GaN PHASE SHIFTER DESIGN AND PARAMETERS CALCULATION AND MISMATCH

When designing the phase shifter method of high-pass (HP) and low-pass (LP) frequency switchable filters were used [45] as shown on Fig. 1. Each of the filters parameters were calculated for a certain frequency at which it provides a fixed phase shift. The proposed scheme of phase shifter based on HPF is shown in Fig. 2. The circuit consists of the *LC* high-pass filter of the third order, consisting of the inductances L_1, L_2, and capacitance C_1, and the switches M_1 and M_2. For

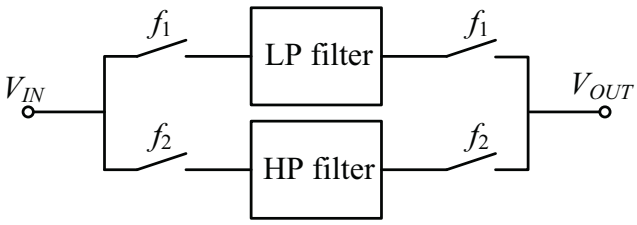


Fig. 1. The phase shifter section with switchable filters

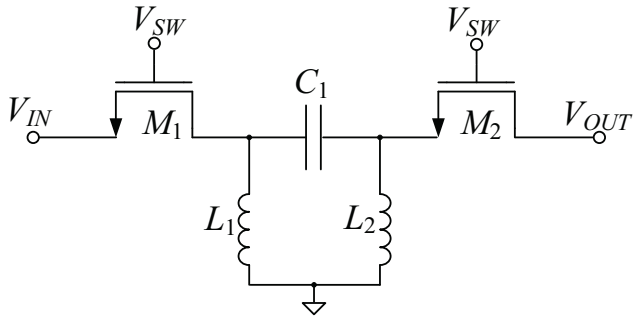


Fig. 2. The circuit level shifter on the basis ofHPF

different frequencies and phase offsets necessary calculations were carried out for the phase shift of 90° at the operating frequency. The calculated component values are shown in Table I, considering both inductors are the same for all frequencies.

For each set of elements parameters from the Table I the phase shifters were simulated and capacitance mismatch was obtained using the Monte Carlo method in *SpectreRF* to measure the operating frequency range and the magnitude of the parasitic amplitude conversion. For each device insertion losses and phase shift that can be achieved with the loss allowed in the specified frequency range were measured. The phase shift at the operating frequency was calculated as 360° minus the value at the phase-frequency characteristic at the operating frequency. The resulting values are also shown in

TABLE I. THE CALCULATED VALUES AND MEASURED PARAMETERS OF HPF ELEMENTS FOR DIFFERENT FREQUENCY BANDS, LOSS AND PHASE SHIFT OF 90°

#	Devices parameters			Measured parameters		
	Working freq., GHz	***C_1, pF***	***L_1, L_2, nH***	***Working freq. range, GHz***	***Insertion loss, phase shift 90°, dB***	***Phase shift on max. allowed losses***
1	1.246	2.56	6.2	0–3	0.1	0.5 dB, 135°
2	1.602	2	4.97	0.5–4	0.25	0.5 dB, 135°
3	2.4	1.3	3.3	0.5–4	0.25	0.5 dB, 135°
4	3.9	0.82	2.04	2–5	0.25	1 dB, 145°
5	6.225	0.52	1.28	2–10	0.25	1 dB, 145°
6	18.5	0.16	0.4	10–25	0.15	2 dB, 155°
7	30	0.09	0.22	10–55	0.2	2 dB, 155°

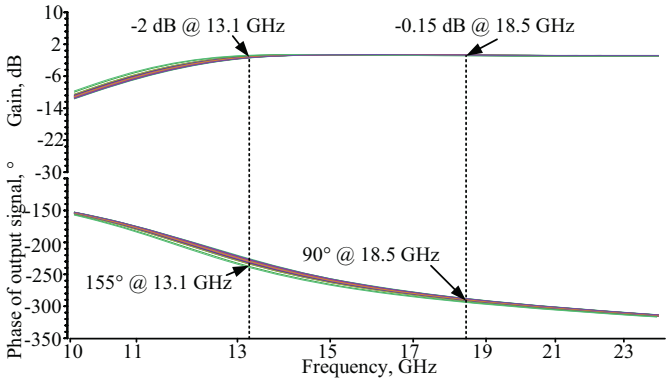


Fig. 3. The frequency response and phase shift parameters mismatch for the 10–25 GHz frequency range

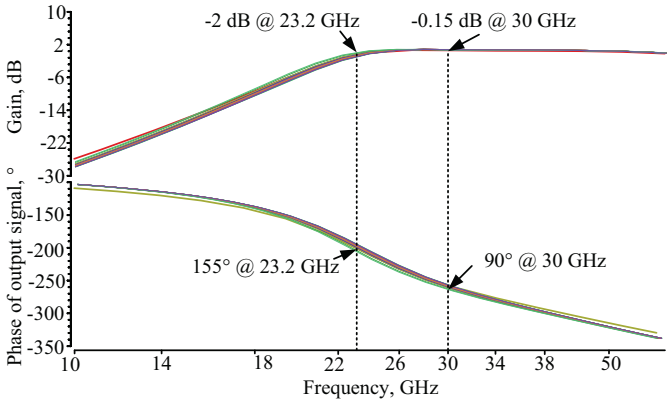


Fig. 4. The frequency response and phase shift parameters mismatch for the 10–55 GHz frequency range

Table I and on Fig. 3 and 4, with the example frequency ranges of 10–25 GHz and 10–55 GHz.

V. Conclusion

This article describes the design of the phase shifter based on GaN HEMT. A review of existing GaN IC phase shifters was made. The *SPICE* transistor model of GaN HEMT was modified and corrected after measurements, which allows the design of GaN circuits for chosen technology, using available software tools. The experimental samples of GaN transistors were fabricated.

The circuit solutions have been selected in the development of monolithic IC implementation of the phase shifter. The circuit simulation and optimization performed for each working frequency and temperature range. During simulation maximum phase shift and insertion losses were measured considering the effect of capacitance mismatch. For future research it is planned to develop a method of compensation of devices mismatch on phase shifter parameters.

Acknowledgment

Authors thank *Rohde & Schwarz* Representative in Moscow for effective collaboration. Results are obtained during applied research under grant No. 14.575.21.0097, unique ID RFMEFI57514X0097, from Ministry of Education and Science of the Russian Federation.

References

[1] M.E. Levinshtein, S.L. Rumyantsev, and M.S. Shur, Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, SiGe (New York: John Wiley & Sons, Inc., 2001), pp. 1–185.

[2] L. S. McCarthy *et al*., "GaN HBT: toward an RF device," in IEEE Transactions on Electron Devices, vol. 48, no. 3, pp. 543-551, Mar 2001.

[3] Danilin V.N., Zhukova T.A., Kuznecov Ju.A. i dr. Obzory po jelektronnoj tehnike. Ser.1, SVCh Tehnika. 2001. Vyp.1. 135 p. (in Russian)

[4] Danilin V.N., Zhukova T.A. i dr. Tranzistor na GaN poka samyj "krepkij oreshek" // JeLEKTRONIKA: NTB. 2005. No. 4. pp. 20–29. (in Russian)

[5] M. Ruff, H. Mitlehner and R. Helbig, "SiC devices: physics and numerical simulation," in *IEEE Transactions on Electron Devices*, vol. 41, no. 6, pp. 1040-1054, Jun 1994.

[6] Bougrov V., Levinshtein M.E., Rumyantsev S.L., Zubrilov A., in Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe // John Wiley & Sons, Inc., New York, 2001. pp. 1–30

[7] Y. S. Lee and Y. H. Jeong, "Applications of GaN HEMTs and SiC MESFETs in High Efficiency Class-E Power Amplifier Design for WCDMA Applications," *2007 IEEE/MTT-S International Microwave Symposium*, Honolulu, HI, 2007, pp. 1099-1102.

[8] J.W. Milligan, S.T. Allen, J.J. Sumakeris, A.R.Powell, J.R. Jenny, and J.W. Palmour, "Transition of High Power SiC MESFETs from 2-inch to 3-inch Production for Improved Cost and Producibility", GaAs MANTECH Conference Digest of Papers, Scottsdale Az., 2003.

[9] I. Angelov, H. Zirath and N. Rorsman, "A new empirical nonlinear model for HEMT and MESFET devices", *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 12, pp. 2258-2266, 1992.

[10] A.A.Kishhinskij. Usiliteli moshhnosti diapazona 0.8-2.5 GGc na SiC-tranzistorah. Materialy 16-oj Mezhdunarodnoj Krymskoj konferencii "SVCh-tehnika i telekom-munikacionnye tehnologii", 2006, pp. 171-172. (in Russian)

[11] B. Van Zeghbroeck, S.-S. Chang, R.L. Waters, J. Torvik, J. Pankove, GaN/SiC HBTs and related issues, Solid-State Electronics, Volume 44, Issue 2, 1 February 2000, Pages 265-270.

[12] K. Tsarik, K. Lavrentiev, V. Nevolin, V. Petukhov AlN sensor based on surface acoustic wave with signal amplification by AlN/GaN HEMT // ICCICCT, Dec. 2015. pp. 734–737.

[13] O. Katz, A. Horn, G. Bahir and J. Salzman, "Electron mobility in an AlGaN/GaN two-dimensional electron gas. I. Carrier concentration dependent mobility," in *IEEE Transactions on Electron Devices*, vol. 50, no. 10, pp. 2002-2008, Oct. 2003.

[14] M. Ruff, H. Mitlehner and R. Helbig, "SiC devices: physics and numerical simulation," in *IEEE Transactions on Electron Devices*, vol. 41, no. 6, pp. 1040-1054, Jun 1994.

[15] X. Hong and Y. Huang, "GaAs device Ohmic contact degradation assessment and analysis," *Millimeter Waves (GSMM), 2012 5th Global Symposium on*, Harbin, 2012, pp. 320-322.

[16] M. Shur, “GaN-based transistors for high power applications,”Solid-State Electron., vol. 42, p. 2131, 1998

[17] Bowick C. RF Circuit Design, 3rd Edition / Newnes, 288 p.

[18] I. Vikulov GaN-mikroshemy priemoperedajushhih modulej AFAR: Evropejskie razrabotki. // Jelektronika: Nauka, Tehnologija, Biznes 2009 №7. str. 90-97 (in Russian)

[19] Bakhtin, A., Semenova, A., Omelyanchuk, E., Teplyakov, I. On possibilities of capacity increasing of high speed radio downlink //_International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, January, 2015. – pp. 184-189.

[20] R. C. Fitch *et al*., "Implementation of High-Power-Density - Band AlGaN/GaN High Electron Mobility Transistors in a Millimeter-Wave Monolithic Microwave Integrated Circuit Process," in *IEEE Electron Device Letters*, vol. 36, no. 10, pp. 1004-1007, Oct. 2015.

[21] V. Camarchia *et al*., "Advanced GaN-based high frequency power amplifiers," *Wireless Power Transfer (WPT), 2013 IEEE*, Perugia, 2013, pp. 29-32.

[22] C. Andrei, O. Bengtsson, R. Doerner, S. A. Chevtchenko and M. Rudolph, "Robust stacked GaN-based low-noise amplifier MMIC for receiver applications," *2015 IEEE MTT-S International Microwave Symposium*, Phoenix, AZ, 2015, pp. 1-4.

[23] P. Choi, U. Radhakrishna, C. C. Boon, D. Antoniadis and L. S. Peh, "A Fully Integrated Inductor-Based GaN Boost Converter With Self-Generated Switching Signal for Vehicular Applications," in *IEEE Transactions on Power Electronics*, vol. 31, no. 8, pp. 5365-5368, Aug. 2016.

[24] P. Schuh, H. Sledzik, R. Reber, A. Fleckenstein, R. Leberer, M. Oppermann, R. Quay, F. van Raay, M. Seelmann-Eggebert, R. Kiefer, M. Mikulla GaN MMIC based T/R-Module Front-End for X-Band Applications / Microwave Integrated Circuit Conference European - EuMIC , 2008. – pp. 274 – 277

[25] M. Thorsell, M. Fagerlind, K. Andersson, N. Billstrom and N. Rorsman, "An X-Band AlGaN/GaN MMIC Receiver Front-End," in *IEEE Microwave and Wireless Components Letters*, vol. 20, no. 1, pp. 55-57, Jan. 2010.

[26] S. Masuda *et al*., "Millimeter-wave multi-chip module for GaN MMIC transceivers fabricated using multilayer ceramics technology," *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*, Seattle, WA, 2013, pp. 1-3.

[27] Ju.R. Efimenkov, V.O. Tkachenko, V.N. Danilin. Issledovanija po sozdaniju moshhnogo SVCh tranzistora s bar'erom Shottki, na GaN, s zatvorom, sformirovannym s pomoshh'ju jelektronnoj litografii. // Materialy V Mezhdunarodnoj nauchno-tehnicheskoj shkoly-konferencii Molodye uchenye – 2008, chast' 1, str. 148-151. (in Russian)

[28] V. P. Meschanov, I. V. Metelnikova, V. D. Tupikin and G. G. Chumaevskaya, "A new structure of microwave ultrawide-band differential phase shifter," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 42, no. 5, pp. 762-765, May 1994.

[29] J. Pedrós, F. Calle, R. Cuerdo, J. Grajal, Z. Bougrioua Voltage tunable surface acoustic wave phase shifter on AlGaN/GaN / Applied Physics Letters, 96, 123505 (2010). – pp.1–2

[30] T. N. Ross, K. Hettak, G. Cormier and J. S. Wight, "Design of X-Band GaN Phase Shifters," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 1, pp. 244-255, Jan. 2015.

[31] TriQuint launches GaN amplifiers and GaAs phase shifters and high-power limiters // Semiconductors Today, Dec. 2013. – Available at: http://www.semiconductor-today.com/news_items/2013/DEC/TRIQUINT_031213.shtml (accessed 05.04.2016)

[32] Phase shifters // TriQuint Inc., 2014. – Available at: http://www.triquint.com/products/all/control-products/phase-shifters (accessed 05.04.2016)

[33] S. Khandelwal, S. Ghosh, Y. S. Chauhan, B. Iniguez and T. A. Fjeldly, "Surface-Potential-Based RF Large Signal Model for Gallium Nitride HEMTs," *2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, New Orleans, LA, 2015, pp. 1-4.

[34] S. Khandelwal *et al*., "Unified Compact Model Covering Drift-Diffusion to Ballistic Carrier Transport," in *IEEE Electron Device Letters*, vol. 37, no. 2, pp. 134-137, Feb. 2016..

[35] J. C. Freeman, "Channel temperature model for microwave AlGaN/GaN power HEMTs on SiC and sapphire," *Microwave Symposium Digest, 2004 IEEE MTT-S International*, 2004, pp. 2031-2034 Vol.3..

[36] K. Jia, W. Sun and L. Shi, "A novel surface potential-based mobility degradation model of thin-oxide-MOSFET for circuit simulation," *Electron Devices and Solid-State Circuits (EDSSC), 2011 International Conference of*, Tianjin, 2011, pp. 1-2.

[37] Y. Tsividis 'Operation and modeling of the MOS transistor' –, 3 edition, Mcgraw Hill.

[38] A. Mangla, J. M. Sallese, C. Sampedro, F. Gamiz and C. Enz, "Modeling the Channel Charge and Potential in Quasi-Ballistic Nanoscale Double-Gate MOSFETs," in *IEEE Transactions on Electron Devices*, vol. 61, no. 8, pp. 2640-2646, Aug. 2014..

[39] U. Radhakrishna, D. Piedra, Y. Zhang, T. Palacios, D. Antoniadis 'Physics-based Compact Model of High Voltage GaN HEMTs: Experimental Verification, Field Plate Optimization and Charge Trapping'-, Electron Devices Meeting (IEDM), 2013 IEEE International , Dec. 2013.

[40] A. Timoshenko, A. Bakhtin, and K. Tsarik, “Power and RF GaN transistors for flexible antennas,” Proc. of the Int. Conf. on Power and Energy (CPE 2014), Shanghai, China, 29-30 November 2014. – pp.347–350.

[41] Berroth M., et. al. High-frequency equivalent circuit of GaAs FETs for large-signal applications// IEEE transactions on Microwave Theory and Techniques, 1991. – C. 224-229

[42] R. G. Brady, C. H. Oxley and T. J. Brazil, "An Improved Small-Signal Parameter-Extraction Algorithm for GaN HEMT Devices," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1535-1544, July 2008.

[43] E. Belousov, A. Timoshenko, K. Lomovskaya Survey on Designing a CAD Model for GaN HEMT Based on Measurements // Proceedings of IEEE EWDTS, Batumi, Georgia, September 26-29, 2015. – pp. 362–365.

[44] A. Timoshenko, K. Lomovskaya, E. Belousov Radio Frequency and Signal Integrity Parameters for GaN MMIC // Proceedings of the 2016 IEEE North West Russia Section Young Researchers in Electrical and Electronic Engineering Conference, St. Petersburg, Russia February 2-3, 2016 – pp. 495–499.

[45] Robert V. Garver. Broad-Band Diode Phase Shifters. IEEE Transactions on Microwave Theory and Techniques, vol. 20, no. 5, May 1972.

Real time Face Detection/Monitor using Raspberry pi and MATLAB

Ali Akbar Shah[1], Zulfiqar Ali Zaidi[2], Dr. Bhawani Shankar Chowdhry[3], Dr. Jawaid Daudpoto[4]
[1,2,3]Department of Electronics Engineering, [4]Department of Mechanical Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
shah.aliakber32@gmail.com

***Abstract*: In this paper, we are implementing facial monitoring system by embedding face detection and face tracking algorithm found in MATLAB with the GPIO pins of Raspberry pi B by using RasPi command such that the array of LEDS follows the facial movement by detecting the face using Haar classifier, tracking its position in the range assigned using the eigenfeatures of the face, which are detected by eigenvectors of MATLAB and by face tracking, which is been carried by geometrical transformation so that motion and gesture of the face can be followed. By doing so we are opening up new way of facial tracking on a live streaming by the help of Viola Jones algorithm and an IR camera.**

Keywords: RasPi, Matlab, Raspberry pi, facial tracking, Viola Jones, live streaming

I. INTRODUCTION

Face detection and tracking has been used for the purposes of surveillance, security, human computer interaction, etc. Various methods of face detection are reported in literature, they include Viola Jones, Haris corner, Principle component analysis, Haar classifier [1-2]. In this research,Haar classifier extracted from Viola Jones algorithm is used for the face detection. The Eigen features of the face for tracking its position are detected using MATLABand Raspberry pi. Computational models of face detections and face recognition have been traditionally used mostly for the purposes of criminal detection and recognition. These basic principles of image processing can be used for entertainment and gaming purposes as well. It can play an important and key role in helping us communicate between virtual and real world. This paper proposes gesture based tracking while using just a camera, the LED on the array of LEDs will glow to the exact position, where your face is located and not only this, it will follow and track your face within its prescribed limit.

Methods like Prinicipal Component Analysis (PCA), Independent Component Analysis (ICA), Local Binary Pattern (LBP), Linear Displacement Analysis (LDA), Sparse Representation based methods and Support Vector Machines are used for facial recognition and its features [2,3,4,5,6]. Some face recognition methods analyze the geometric features of facial images, such as location and distance between nose, eyes, and mouth [7, 8]. Some of the abovementioned methods are not so effective because of the illumination and facial expressions. To avoid this, a holistic image may be created to extract all features of the face. Hence, Viola Jones developed Haar classifier, which is a real time face detection algorithm [9].

Some parallel versions of face detection algorithm use Haar-like features [10-12]. The algorithm for face detection algorithm using GPU was proposed by Hefenbrock [10]. The major flaw in this algorithm was that it was not able to process a stream resolution of 640x480 in real time. Likewise, Obukhov's alogorithm was developed [11].

II. MATERIAL AND METHODS

A. Raspberry Pi

The Raspberry Pi which is a single board minicomputer developed in UK. Raspberry Pi 1 Model B is used in this research. It is a high specification, minicomputer embedded on a chip having 512 MB DRAM and a 100MB external port.Raspberry pi allows us to connect to any remote computer with MATLAB. The model B of Raspberry Pi used in this project has 4 USB ports, Ethernet port and audio out port. It's light and portable having 512 MB DRAM, power ratings of 700mA and a 5V via micro USB power source. Table. 1 lists the pinouts of the Raspberry Pi used in this research.

TABLE I. RASPBERRY PI PIN CONNECTIONS

1	3.3V	7	GPIO 4	13	GPIO 21	19	GPIO 10
2	5V	8	GPIO 14	14	GND	20	GND
3	GPIO 0	9	GND	15	GPIO 22	21	GPIO 9
4	5V	10	GPIO 15	16	GPIO 23	22	GPIO 25
5	GPIO 1	11	GPIO 17	17	3.3V	23	GPIO 11
6	GND	12	GPIO 18	18	GPIO 24	24	GPIO 8
25		GND		26		GPIO 7	

B. Block diagram:

The block diagram given in Fig. 1 represents the steps required for implementing the face recognition using Raspberry Pi and MATLAB. Each step is described in the following section.

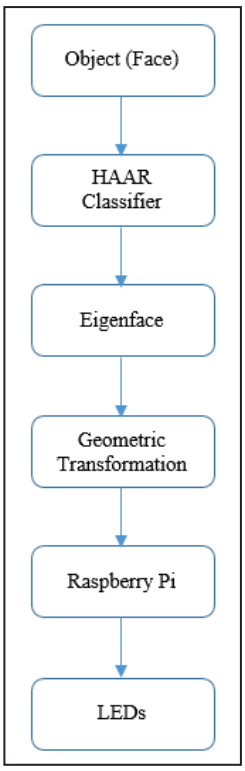


Figure 1. Block diagram

III. MATH

A. Haar Classifier

Fig. 2 contains five Haar patterns. The position and size of a pattern can change if its black and white rectangles have the same dimension and border, and each of them keep their relative positions. Dueto this, the drawn features are of small size. A 24×24 pixel image, for instance, has (43200, 27600), (43200, 27600) and 20736 features of category (a), (b) and (c)respectively, hence 162336 features in all. Hence we compute the formula sum of rectangular areas in the image using look up tables (for suppose, 4 look up tables) of an integral image as:

$$\text{Sum} = I(C) + I(A) - I(B) - I(D) \quad (1)$$

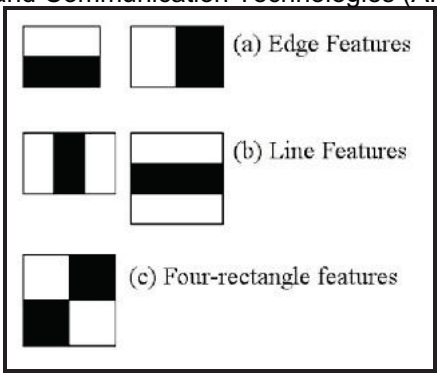


Figure 2. Haar Classifier Patterns

Where A, B, C and D belong to the integral image as shown in Fig. 3.

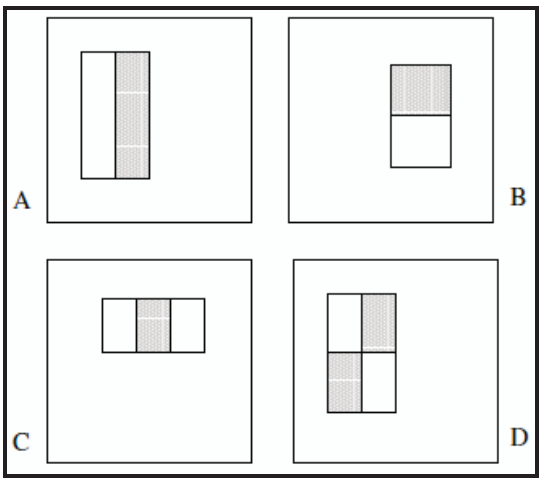


Figure 3. Integral Image

Thus, we are considering 4 look up tables for understanding equation (1) only, integral image can have more look up tables.

B. Eigenface

Eigenface is the name given to a set of eigenvectors when they are used in the computer vision problem of human face recognition [14]. The approach of using eigenfaces for recognition was developed by Sirovich and Kirby (1987) and used by Matthew Turk and Alex Pentland in face classification [15].

Thus, Eigenface uses the Eigen features for the recognition of the face. Eigen features include eyes, nose and mouth. Each integral image detected by Haar's classifier is then is further extracted by the conversion:

$$Y_k = U_k^T (G-Y) \quad (k=1,\ldots,5) \quad (2)$$

For each vector X (i = 1, …, M), the Euclidean distance between ω and X is calculated as follows:

$$e_{i,k}^2 = \|Y_k - Y_{i,k}^2\|^2 \quad (i=1,\ldots,M,\ k=1,\ldots,5) \quad (3)$$

The result is displayed in Fig. 4.

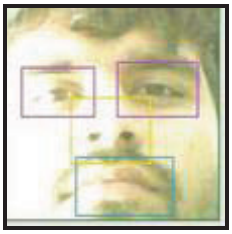


Figure 4. Facial Parts Recognition Using Eigenface Highlighted By bbox

As demonstrated in the figure, the Eigen features of the face are detected by applying eigenvectors and are highlighted by bbox boundaries. Hence, from this, we would fix our eigenfeatures value of the nose for performing the task of face tracking. Such that, whenever the face will be in motion, the bbox will highlight the nose using its boundary along with the motion on real time video streaming.

C. Geometric Transformation

We are performing face tracking operation by the help of geometric transformation. As discussed previously, we have detected face using Haar classifier of Viola Jones algorithm as well as have extracted eigenfeatures using eigenfeatures of eigenface. Now, for modifying the pixels of the image without changing its color, we are using geometric transformation. For geometric transformation, we are first of all changing the rgb image into gray for enhancing its contrast and are saving the old points detected in bbox using eigenfaceand allocating them in a pointer. Thus, the subject is in motion, therefore morphing of the new and old image is performed for modifying the pixels of the images and then again we will convert our gray image into rgb image.

Suppose (s, t) are represent the image co-ordinates of the previous image and (x, y) are representing the image co-ordinates of the new image. Function is used for matching the image co-ordinates of the two images. Such that,

Forward mapping:

$x=x(s, t)$ (4)

$y=y(s, t)'$ (5)

Such that equations (4) and (5) can be represented as,

$\mathbf{x}=x(\mathbf{s})$ (6)

Inverse mapping:

$s=s(x, y)$ (7)

$t=t(x, y)'$ (8)

such that equation (7) and (8) can be represented as,

$\mathbf{s}=s(\mathbf{x})$ (9)

let f(s, t) or f(s) represent the old image and h(x, y) or h(x) represent the new image.

$h(x, y)=f(s(x, y), t(x, y))$ or $\mathbf{h}(x)=f(s(\mathbf{x}))$ (10)

$f(s, t)=g(x(s, t),y(s, t))'$ or $\mathbf{f}(s)=g(x(\mathbf{s}))$(11)

The good thing about raspberry pi is that it allows us to connect to any remote computer having Matlab running on it. By setting just a command line on Raspberry pi (by using RasPi as a command of Matlab), it allows us to interface Matlab with the raspberry pi having LEDs connected to its output peripherals (GPIOs). First of all, we analyze the width of the frame of the video players in which our face detection and recognition is been performed then we are dividing that width with the number of LEDs connected to the Raspberry pi board. The width in which video stream has been played is 480 and we are using 4 LEDS connected to our GPIOs. Now by dividing 480 with 4 and we are getting the width range of each LED, which is 120. Now we are assigning the ranges of LEDs that are been used such that, if LED 1 covers the range of 120 then the LED 2 will have the range 120+120, that is 240 and so the same will be applied on LED 3 and 4, respectively. As shown in Fig. 5.

After that we are creating object using command on RasPi, that is; rpi = raspi();

By assigning the LED (ON) condition (according to their ranges set that we have calculated), with their GPIOs that LED will turn ON, if the face will enter its range.

Now, for suppose if we are positioning our face to the left, the bbox that is tracking our face using eigenface (eigenfeatures) will process the range of the width in which our face is located to the RasPi and the RasPi would turn on the LED of that range. If we shift our face to the other direction, the geometric transformation technique will update the pixels and the new range will be updated to RasPi. Same procedure will be repeated continuously.

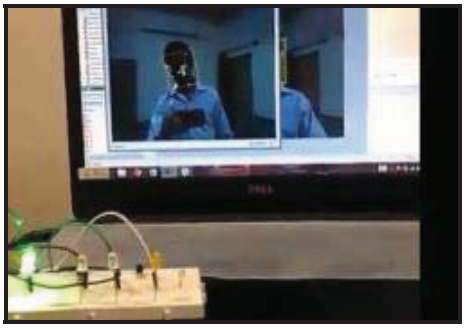


Figure 5. Facial Monitor by Interacting Face Tracking with LEDs using Raspberry Pi in Matlab.

IV. Conclusion and Future Work

In this paper, by embedding Raspberry PI with image processing techniques, we are able to show the position by turning on the LED of the range in which face is located. In future, we would like to use gesture for controlling devices. Moreover, this application is limited for single person face tracking purpose only. In future, we would like to expand it into multiuser application.

Acknowledgement

We would like thank our mentor, Prof. Dr. Bhawani Shankar Chowdhry for his support and guidance throughout our career and our second home, Mehran University Engineering and Technology for providing us different opportunity for exhibiting our talents

References

[1] P. Viola and M. Jones. Robust Real-time Object Detection. International Journal of Computer Vision, 57(2):137–154, 2002.

[2] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, “Face recognition by independent component analysis,” IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450–1466, 2002.

[3] R. Chellappa, J. Ni, and V.M. Patel, “Remote identification of faces: Problems, prospects, and progress,” Pattern Recognition Letters, vol. 33, no. 14, pp. 1849– 1859, 2012

[4] K. Etemad and R. Chellappa, “Discriminant analysis for recognition of human face images,” Journal of the Optical Society of America A, vol. 14, no. 8, pp. 1724– 1733, 1997.

[5] B. Heisele, P. Ho, and T. Poggio, “Face recognition with support vector machines: global versus component based approach,” in Proceedings of IEEE International Conference on Computer Vision, 2001, vol. 2, pp. 688– 694.

[6] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, 2008

[7] P.N. Belhumeur, J.P. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 7, pp. 711–720, 1997.

[8] Keun-Chang Kwak and W. Pedrycz, “Face recognition using an enhanced independent component analysis approach,” Neural Networks, IEEE Transaction son, vol. 18, no. 2, pp. 530–541, 2007.

[9] P. Viola and M. J. Jones, Robust real-time face detection, International Journal of Computer Vision, 57 (2004), pp.137–154.

[10] Daniel Hefenbrock, “Accelerating Viola-Jones face detection to FPGA-level using GPUs,” Proceedings of the 2010 18th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines, 2010, pp.11-18.

[11] Anton Obukhov, “Haar Classifiers for Object Detection with CUDA”, GPU Computing Gems. Emerald Edition, 2011, pp.517-544.

[12] Adam Herout, “Real-time object detection on CUDA,” Journal of Real-Time Image Processing, vol.6, issue 3, 2011, pp.159-170.

[13] Crow, F, "Summed-area tables for texture mapping", in Proceedings of SIGGRAPH, 18(3):207–212, 1984

[14] L. Sirovich and M. Kirby (1987). "Low-dimensional procedure for the characterization of human faces". Journal of the Optical Society of America A **4** (3): 519 -524.doi:10.1364/JOSAA.4.000519.

[15] M. Kirby and L. Sirovich (1990). "Application of the Karhunen-Loeve procedure for the characterization of human faces". IEEE Transactions on Pattern analysis and Machine Intelligence 12 (1):103-108. doi:10.1109/34.41390

The Model of Management Access to The Resources of the Closed Discretionary Information Computation Environment in The Form of Corporate Portal Network

Igor S. Konstantinov, Sergej A. Lazarev,
Oleg V. Mihalev, Vladimir E. Kiselev
Belgorod State University
Belgorod, Russia

Alexander V. Demidov
Orel State University
Orel, Russia

***Abstract*—In this article the mechanism is described of how to realize the user authorization model based on group policies, applied to the resources of the closed information computation environment, which is represented as an enterprise information portals network.**

***Index Terms*—access control, group policies, authorization model, user authorization, distributed information computation resources, distributed information computation environment, closed virtual environment.**

I. Introduction

The process of the distributed information computation environment (DICE) infrastructure development as the enterprise information portal network is presumed to involve realization of the unified access management policy based on the unified hierarchy of the user groups as the key task.

The description of the portals network could be found in the papers [1-5], where it is presented as a single entry point mechanism providing a unified user & administrator access rights differentiation policy. Such a mechanism constitutes a set of access check stations consolidated into a network operated by the portal network control center (PNCC). By the means of PNCC the unified information exchange policy is realized, and, among other things, feasibility of the authorized access to the secured information computation resources of the whole network as the management mechanism for a user session allocated for the necessary security level.

The suitable information computation access models have been contemplated in papers [6-8]. The modification of the discretionary access control, model (DAC) was conceived as the basis.

II. Task assignment. Resources access model

Traditional discretionary access control complies with the following requirements [9-10]:

1. All entities (including subjects) should be identified, i.e. a unique ID should be attributed to each subject.

2. An access matrix should be established, in which every matrix row matches a subject and each cell contains a rights list defining the entity access rights of the subject. Such a list represents a subset of the realized access rights set.

3. A subject possesses an entity access right upon only one condition – the relevant cell contains an exact access right.

Having regard to the above, it is possible to formalize the model of the access management system of the discretionary information computation system infrastructure.

Access management system should be defined as:

$$S = \langle C, N, R, U, D, G \rangle, \tag{1}$$

$C = \{C_k\}, k \in \overline{1, n_C}$ – set of network control centers,

$N = \{N_k\}, k \in \overline{1, n_N}$ – set of network nodes, while a particular node:

$$N = \langle R', U', D', G' \rangle. \tag{2}$$

The main terms should be defined as part of information computation resources access management model. All terms represented are accordant with directive document «Information access security. Terms and definitions. » by State Technical Commission attached to the President of the Russian Federation.

Access object – a unit of information computation resources in the automated system, access to which is defined by divisionary access rules.

Access subject – an entity or a process, actions of which are defined divisionary access rules. It is of high importance to estimate the dependence and the sanction attribute within a pair of access object and access subject.

Group policy objects are determined as a set of information computation resources:

$$R = \{r_k\}, k \in \overline{1, n_R}, \tag{3}$$

r_k – an information computation resource, n_R – the quantity of information computation resource required for system operating.

Group policy subjects are determined by the set of users:

$$U = \{u_k\}, k \in \overline{1, n_U}, \quad (4)$$

u_k – user, n_U – quantity of the registered users.

III. Main section

Traditionally the system user is considered to be a person or a device which is capable of sending system requests via telecommunications channels, and, more than that, capable of getting responses in the form of granting the access to the certain resource or abandon it. Concerning the particular system, it is stated that a user is a member of one and only one organization-member off the portals network. As for external users, who do not belong to any organizations-members of such a kind, a supplementary fictional organization is required. Every unregistered user is considered as a member of that fictional organization. It is needed to add that users of such a kind are equal in their access rights, as well as all of them are granted with only the lowest access rights level. A set of users registered as the members of the same supported organization forms a user's domain. Users' domain is bounded to the network control center and is operated by its administrator. Network control centers quintessentially are in the role of AAA-servers, ensuring storage of the authentication and other data belonging to the users of the domain.

Each user is associated with a user domain related to a set:

$$D = \{d_k\}, k \in \overline{1, n_D}, \quad (5)$$

d_k – user's domain, n_D – quantity of users' domains registered in the system.

For the effective access rights management users should be consolidated into groups according to the internal organizational hierarchy, for instance, that might be departments, services, branches. Network nodes should sustain the applicability of the consolidation and connection data. Following this goal, each information resource is affiliated with a certain group of privileges in the set:

$$G = \{g_k\}, k \in \overline{1, n_G}, \quad (6)$$

g_k – a group of privileges, n_G – amount of groups of privileges, registered in the system.

Access subjects and objects then are defined as tuple:

$$\forall u_k \in U : u_k = \left\langle i_k^U, d_i, g_j \right\rangle, \quad (7)$$

$I_U = \left\{ i_k^U \right\}, k \in \overline{1, n_U}$ – set of user's IDs;

$$\forall r_k \in R : \left\langle i_k^R, g_i \right\rangle, \quad (8)$$

$I_R = \left\{ i_k^R \right\}, k \in \overline{1, n_R}$ – set of information resources' IDs.

Conversely the relation F_A is of great importance, defining user's access to a resource:

$$\exists F_A : U \cdot R \to \{\text{true}, \text{false}\} \quad (9)$$

Discretionary access control model allows user u to commit an operation on information resource r if only group $g_i \in u$ matches together with group $g_j \in r$.

Nevertheless, the order of access configuration inheritance should be established as the viable tool for group access policy management. For completing this task privilege groups should be presented in the form of a directed graph. In this graph each group of privileges is matched with vertexes, while edges define access configuration inheritance:

$$\forall g_k \in G : g_k = \left\langle i_k^G, G_k \right\rangle, \quad (10)$$

$I_G = \left\{ i_k^G \right\}, k \in \overline{1, n_G}$ – set of privilege groups' IDs,

$G_k \subset G$ – set of inherited privilege groups.

Graph traversal forms a set $\widetilde{G} \subseteq G$ The user's u access to information resource r would be granted, if intersection of sets $\widetilde{G}_i$ and $\widetilde{G}_j$ is not an empty set. The first set $\widetilde{G}_i$ occurred while traversing the hierarchy graph of privilege group $g_i \in u$, while the second set $\widetilde{G}_j$ occurred while traversing the hierarchy graph of privilege group $g_j \in r$.

Due to distributed nature of the system, a number of restrictions should be implemented into the privilege groups' inheritance concept. More than that, a significant difference between inheritance links of groups should be taken into account, as the creation of a flexible access management mechanism is the main aim or the research.

$$\forall g_k \in G : g_k = \left\langle i_k^G, D_k, G_k, \overline{G}_k \right\rangle, \quad (11)$$

$D_k \subseteq D$ – set of users' domains permitted for groups, $G_k \subset G$ – set of privilege groups, included into the inheritance hierarchy, $\overline{G}_k \subset G$ – set of privilege groups, excluded from the inheritance hierarchy.

Inclusion links allow inheriting the access permissions belonged to privilege group hierarchy. Exclusion links realize the possibility of excluding privilege groups from the inheritance hierarchy without the necessity of making any kind of changes in the parental vertexes.

In general, it is possible to describe the relation F_G, defining a finite subset $\widetilde{G} \subseteq G$:

$$F_G(g_i) = \{g_i\} \cup \left(\bigcup_k^{G_i} F_G(g_k) \setminus \bigcup_k^{\overline{G}_i} F_G(g_k) \right) \quad (12)$$

Then the relation F_A, defining the user's u access permission for an information resource r, will be transformed into Eq. 13:

$$F_A(u,r) = \begin{cases} \text{true, for } F_G(g_u) \cap F_G(g_r) \neq \varnothing \text{ and } d_u \in D_r; \\ \text{false, for } F_G(g_u) \cap F_G(g_r) = \varnothing \text{ or } d_u \notin D_r. \end{cases} \quad (13)$$

Using binary coding of elements resource-intensive operations on sets might be simmered down to operations on Boolean values [4]. A logical scheme could be compiled using the symbolic representation of Boolean functions, taking the user's belonging to a privilege group as a predicate $\varepsilon(u,g)$ and taking the user's belonging to a user's domain as a predicate $\delta(u,d)$. Such a logical scheme defines an access permission following the principles enlisted and represented above:

$$\widetilde{F}_G(u_i, g_j) = \varepsilon(u_i, g_j) \vee \left(\bigvee_k^{G_j} \widetilde{F}_G(u_i, g_k) \wedge \overline{\bigwedge_k^{\overline{G}_j} \widetilde{F}_G(u_i, g_k)} \right), \quad (14)$$

$$\widetilde{F}_A(u_i, r_j) = (\vee_k^{D_r} \delta(u_i, d_k)) \wedge \widetilde{F}_G(u_i, g_r). \quad (15)$$

IV. Resulting deduction

The user would be granted with an access to an information resource, providing that the user is affiliated with one of the privilege groups from the inheritance hierarchy, while the user's domain is allowed for the privilege group in the root of the graph.

V. Conclusions

The examined approach is the basis for the realization of the authorization algorithm for users of the closed distributed environment in the form of the portal network. A modification of the discretion access model is used as the fundamental mechanism. The effective algorithm development has required an adaptation of a lightweight authentication algorithm in order to realize group access management policies.

Acknowledgment

The research concerning this issue was sponsored by the RF Ministry of Education and Science. The project ID is RFMEFI57514X0099.

References

[1] Lazarev, S.A., Demidov, A.V., 2010. The Concept of Construction of a Control System of an Information Exchange in The Network of Corporate Portals. Information Systems and Technologies, 4(60): 123–129.

[2] Lazarev S.A., O.A. Ivashchuk, I.S. Konstantinov, K.A. Rubcov, 2014. Mechanism of Information Exchange Management within Portal Network of Environmental Monitoring Subjects, International Journal of Applied Engineering Research, 9 (22), pp. 16789-16794.

[3] Lazarev, S.A, I.S. Konstantinov, O.V. Mihalev, V.L. Kurbatov, 2014, Analysis of the single session access model in the distributed portal network of the interacting parties of the informational space, Research Journal of Applied Sciences, 9 (11), pp. 771-773.

[4] Lazarev, S.A., I.S. Konstantinov, O.V. Mihalev, 2014. Realization of a single model session access in the distributed network portals. Vestnik komp'iuternykh i informatsionnykh tekhnologii (Herald of computer and information technologies), #6: 44-49.4.

[5] Lazarev S.A., Konstantinov I.S., Mihalev O.V., Kiselev V.E., 2015 Implementation of unified session access model in a closed virtual environment of distributed information-computational resource system as a secured portal network. Research Journal of Applied Science, 10 (10): 629-632.

[6] Lazarev S.A., A.V. Demidov, 2012. Features of a subsystem development for the system access management in respect of information exchange of corporate portal network. Information Systems and Technologies, # 4 (72): 103-110.

[7] Lazarev S.A., Konstantinov I.S., Mihalev O.V., Demidov A.V., Shateev R.V., 2015. The development of infrastructure security for distributed information computer environment based on secured portal network. International Journal of Applied Engineering Research, 10 (17): 38116-38120.

[8] Lazarev, S.A., Konstantinov I.S., Silaev P.P., 2015. Safety Mechanism for Multi-Factor Authentication with Digital Access Key Use in Closed Virtual Environment of Distributed Information Resources. International Journal of Applied Engineering Research, 10 (24): 44927-44923.

[9] Aksenov A.N., Aksenova M.V. Analysis and selection of the inter-site authorization mechanism. Contemporary science and education issues. – 2015. – № 2-1.

[10] Devyanin P.N. Analysis of the safety access and information flows management in the computer systems. – M.:Radio and communications, 2006. – 176 c.

Neuroevolution: problems, algorithms, and experiments

Sergey Rodzin, Olga Rodzina, Lada Rodzina
Southern Federal University
Rostov-on-Don, Russia
srodzin@sfedu.ru

***Abstract*—The article describes the problematic issues of neuroevolution, i.e. a promising approach for solving complex problems of machine learning, neural networks, adaptive management and multi-agent systems, evolutionary robotics, gaming strategies, and computer art. The authors have suggested neuro evolutional algorithm and presented experiment results on a standard task: the balancing trolley with two flagpoles of different lengths.**

***Index Terms*—Neuroevolution, reinforcement machine learning, optimization, evolutionary computation, fitness function**

I. Introduction

Neuroevolution is a promising approach for solving complex such problems of machine learning neural networks as adaptive management and multi-agent systems, evolutionary robotics, gaming strategies, computer art, and others. Implementation of supervised learning is almost impossible for these tasks.

Reinforcement learning is an intermediate type between such tasks as learning with the teacher on the precedents (regression, classification) and learning without a teacher, where you need to find patterns in the data (clustering, search for association rules). Reinforcement learning is a type of learning when it takes place in interaction with the environment, in other words, the algorithm makes some actions in the environment and sometimes gets the feedback (like games, negotiations, the scientific research process, etc.). Reinforcement learning is a compromise between research of unexplored areas and application of existing knowledge.

Neuroevolution algorithms are divided into the following three categories depending on the task: search of weights values in the neural network with a fixed structure; setting the neural network structure; setting activation functions of neurons; and various combinations of the above tasks.

The article is devoted to the problems of simultaneous connections and setting the weights of neural network structure.

II. Issues of neuroevolutionary algorithms

There are several fundamental questions related to the development of neuroevolutionary algorithms:

- evolutionary algorithms manipulate a variety of genotypes. In neuroevolution the term "genotype" means a representation (coding) of the neural network. How can we encode efficiently a network structure in the form of a genotype? Currently, each researcher uses its own encryption method. For example, in the direct coding, the genotype is equivalent to the phenotype, and neurons and connections are directly specified in the genotype. In contrast, in implicit coding we either specify the network settings (the number of layers and neurons) or use specialized grammar (grammar evolution, the evolution of the rules of rules usage);
- crossover problem is that the crossover operator can not be applied to the network if its genetic information is different in length. Besides, descendants generated by crossing over often have much worse value of fitness function than their parents;
- how to create a suitable initial population of neural networks? Generation of the initial population of neural networks with random topology is not the best approach, taking into account the cross-over problems. In other words, the effect of the crossover operator on the neuroevolution is destructive.

In the literature, a neural network which is produced as the evolution of network nodes and connections are called *TWEANNs (Topology & Weight Evolving Artificial Neural Networks)* [1]. Using *TWEANNs* was able to solve one of the most difficult test problems of reinforcement learning: the neural network manage the trolley without having information about speed; the trolley is connected via hinges flagpole with two different lengths and in a state of unstable balance [2]. The advantages of the algorithm are the following properties: independence on the type of activation function for neurons; no need for the training set; the ability to automatically search for the neural network structure. We must point out some shortcomings of the algorithm: the difficulties of assessing the neural network structure; it requires more memory than using gradient algorithms; the complexity of the organization structure of neural network research.

We should note the following research groups in the field of development neuroevolutionary algorithms: a group of neural network research (head R. Miikkulainen,

http://www.cs.utexas.edu/~risto/), the group of evolutionary complexity research (head K. Stanley, http://www.cs.ucf.edu/?kstanley), Intelligent Systems Laboratory (head D. Floreano), Laboratory of autonomous robots and artificial life (head S. Nolfi), Center of Computational Intelligence Research and Applications (head *X. Yao,* http://www.cs.bham.ac.uk/~xin), the group of researchers of adaptive systems optimization (head *Ch. Igel*), Swiss Institute of Artificial Intelligence (head *J. Schmidhuber,* http://www.idsia.ch/~juergen/). In Russia, neuroevolutionary algorithms research are conducted in the following research centers: Optoneuron Technology Center together with Keldysh Institute of Applied Mathematics (adaptive behavior), Siberian Federal University (classification and medical diagnostics), Southern Federal University (evolutionary neurocomputing), in Ulyanovsk State Technical University (Intelligent Decision Support System), Tomsk Polytechnic University (classification, image processing) [3-7]. The paper proposes a new algorithm for neuroevolution *TWEANNs* where the crossover operator is not used at all, and as the main operator, we use the mutation.

III. Evolutionary Synthesis Algorithm for TWEANNs Topology

Neurodevelopmental algorithm starts with a neural network without hidden neurons and goes in the direction of topology complication. Coding of the neural network should allow performing meaningful crossing over different topologies. The following two lists represent the genotype: a list of adjacent vertices (neurons), and a list of synaptic weights. The top is encoded by the following two components: the first one identifies it to the generations of the population, while the second one indicates the type of layer (input, hidden, output). List of synaptic weights includes pointers to adjacent vertices, the value of the weight on the flag (if the compound is used) and the "historical label" which serves to identify the connection for all generations in a population. This allows the operator to apply the crossover without duplication or deletion of genetic information from the parents.

Nevertheless, new offspring generated by crossover operator usually has a fitness function worse than the parental chromosomes. In [1] proposed to apply the crossover to individuals with small structural differences. However, this quickly leads to a topological monotony of obtained neural networks.

The authors propose an approach where the crossover is not used. In this approach, the population is regarded as the central object of neuroevolutional algorithm. The authors proceed from the assumption that evolution is primarily a process of adaptation at the behavioral level. This level of abstraction does not provide recombination. Therefore, there is no crossover operator. A mutation in the proposed neuroevolutional algorithm is the only alternative solutions operator.

Let us consider the encoding of neural networks as a way to develop a reliable method for generating neurostructures with a few changes and only in a limited area. We use the biological concept of the operon. Operon is a group of functionally linked genes whose activity is ordered, and depends on external conditions and on the activities of other genes. Based on this concept, as an operon, we understand a certain subnetwork consisting of a subset of neurons (nodes) and a subset of connections between neurons.

Then genotype, as a string encoding consisting of characters, represented as a subset of the operons that resembles the following:

$$string = \{operon_1, operon_2, ..., operon_N\} \quad (1)$$

$$operon_i = \{\{node_j | \, j \in [1: J]\}, \{edge_k | \, k \in [1: K]\}\} \quad (2)$$

where N is a maximum number of operons($i \in [1: N]$); $node_j$ is the identification number of the j-th neuron included in the set $operon_i$; J is the number of neurons included in $operon_i$ with their own identification numbers; $edge_k$ is the k-th connection in $operon_i$; K is the number of connections in $operon_i$.

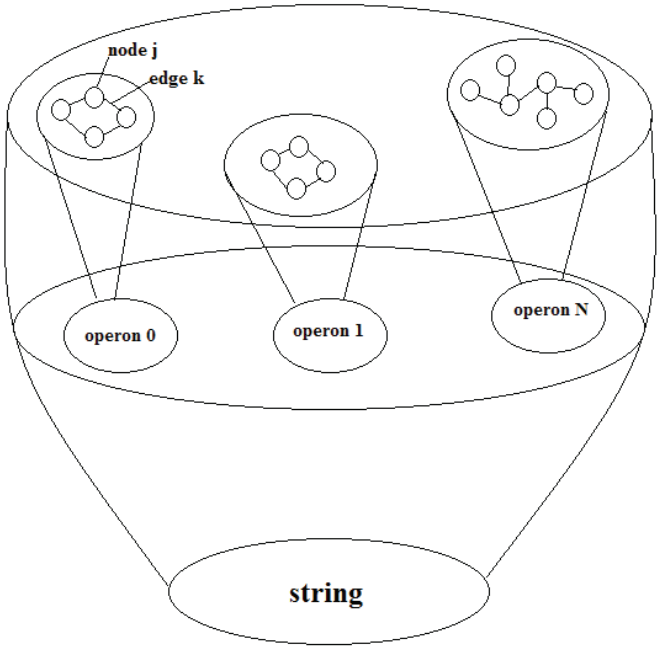


Fig. 1. Coding of neural network: string-operon-(node+edge).

The initial population consists of individuals having only one $operon_0$, which contains only the input and output nodes and connections between them. Here nodes are absent in the hidden layer.

Let us consider two kinds of mutation operator, which are used in the proposed neuroevolution algorithm for adding the nodes and connections in the neural network.

Mutation operator for adding nodes is applied to each operon with constant probability p_m. One of the synaptic connections is randomly selected for removing. Next, the neuron of the hidden layer and synaptic connections associated with it are added instead of removed synaptic connections. If one of the ends of the removed synaptic connection is connected to the node of $operon_0$, then, the added neuron of the hidden layer and its connections will become a new operon. For example, we have $operon_0$ with two inputs x_1, x_2 with weights w_1, w_2 accordingly, as well as with output neuron y_1 that has a sigmoid as an activation function.

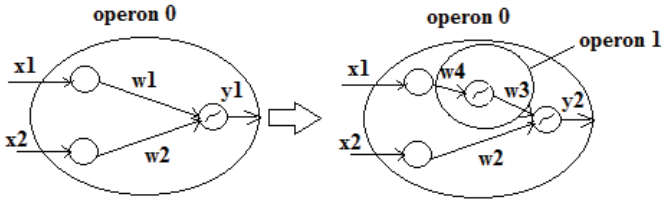


Fig. 2. Mutation for adding nodes.

After removal of the randomly selected synaptic connection with weight w_1 we add the neuron of the hidden layer and two connections with weights w_3 and w_4 accordingly, where w_4 is the weight value of the input x_1, w_3 is the weight of the connection between the input of the new neuron of the hidden layer and the output neuron y_2 operon0. Added neuron and its connections become a new $operon_1$.

To change the output $operon_0$ after the mutation, we must be sure that the values of y_1 and $y_2$are equal. Therefore, the following condition on the assumption that S is a sigmoid function should be carried out:

$$S[w_1x_1+ w_2x_2] = S[w_3S[w_4x_1] + w_2x_2] \quad (3)$$

Hence we find that:

$$w_1x_1 - w_3S[w_4x_1] = 0 \quad (4)$$

Furthermore, under the condition that $w_1 = w_3$ and $S[x] = 1/(1 + exp^{\beta(\alpha-x)})$, we obtain the value of the fitness function $f(x_1)$ after the mutation operator:

$$f(x_1) = x_1 - 1/(1 + exp^{\beta(\alpha-w4x1)}), \quad (5)$$

where in order to simplify the results of experiments the following values were established: $w_4 = 1$, $\alpha = 0{,}5$, $\beta = 5$.

Mutation operator for adding connections is also applied to each operon with a constant probability

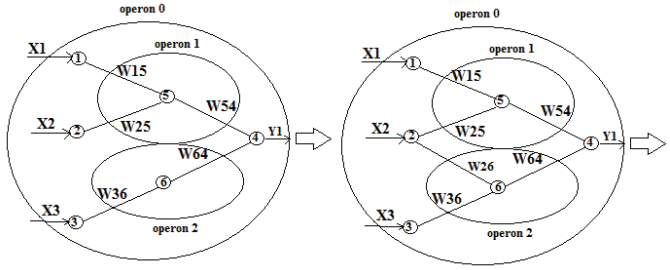


Fig. 3. Mutation for adding connections.

Some node is randomly selected and added to the connection to the node in $operon_i$ or $operon_0$. Weight of the new connection is set to 0, so it does not alter the transmitted signal.

IV. The Task of Balancing Trolley with Two Flagpoles

The authors have evaluated the effectiveness of the proposed neuroevolutional algorithm on such standard task as balancing trolley with two flagpoles [8]. Two flagpoles are connected to the moving trolley by hinges (Fig. 4).

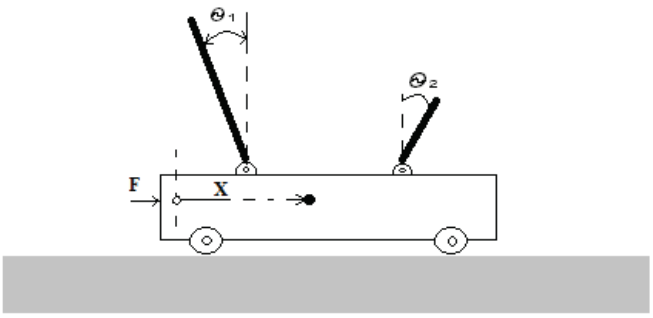


Fig. 4. Trolley with two flagpoles.

A neural network has to manage the movement of the trolley if we want to balance both the flagpole as long as possible. System status is determined by the position x and the speed v_x of the trolley, as well as by the angle γ_1 of deviation from the vertical of the first flagpole and its angular velocity ω_1, and by the angle γ_2 of deviation from the vertical of the second flagpole and its angular velocity ω_2. However, in order to simplify the information about v_x, ω_1 and ω_2 are not used as input data.

Flagpoles have different lengths. Computer modeling was performed with four different ratios of the low and high length flagpole 1:10, 1:5, 1:3 and 1: 2,5. It is well known that the task of balancing flagpoles becomes more difficult when the ratio of the lengths of flagpoles is coming to 1. All other details were the same as in [1] for the computer simulation. The authors used for comparison the data given in [2] for neuroevolutional algorithm *NEAT* (*Neuro-Evolution of Augmenting Topologies*). The reason for choosing the *NEAT* algorithm for comparison is the following: it solves the problem of balancing the flagpole on the cart is 25 times better than the algorithm using cellular encoding [9, 10].

The parameters of the proposed neuroevolutional algorithm and the *NEAT* in the experiments had the following values: the size of the population is 1000; the size of the tournament selection is 20; the weight of connections by using a mutation operator for adding nodes is 1; sigmoidal function parameters $\alpha = 0{,}5$, $\beta = 5$.

The authors have conducted a series of 10 experiments and obtained the following results.

The problem with a probability close to 1 is successfully solved by compared algorithms with the ratio of the lengths between the low and high flagpoles as 1:10 and 1:5. However, further alignment of flagpoles lengths we see as the result that the *NEAT* algorithm became worse. In particular, the most difficult conditions is when the ratio of the lengths of flagpoles 1:2,5. In this case, the balance of the control task is successfully solved for *NEAT* algorithm only once in ten experiments. The proposed algorithm, the likelihood of a successful solution was 0,8.

However, the *NEAT* algorithm with the ratio of the lengths between the low and high flagpoles like 1:10 and 1:5 have found the solution for about 40 generations. This result was significantly better than for the proposed neuroevolutional algorithm. When the length ratio between the low and high

flagpoles as 1:3 and 1:2,5, *NEAT* resolve it within approximately 200 generations, the proposed algorithm - during approximately 350-400 generations.

The most successful case are as follows: the ratio of the lengths of the low and high flagpole 1:10. In this case, the algorithm *NEAT* added 6 units and 11 compounds. With the same ratio between the lengths the proposed algorithm gives a comparable result. However, when the ratio between lengths as 1:2,5, the network complexity obtained by the *NEAT* algorithm grows rapidly and reaches a size of more than 50 nodes and over 300 connections. At the same time, the complexity of the neural network topology generated by the proposed algorithm for all four ratios between the lengths of the low and high flagpoles remains virtually unchanged: it is about 10 nodes and 20 connections. Thus, simulation results confirm the hypothesis about the benefits of generating neurostructures by small mutational changes in a limited area.

V. Conclusion

The authors believe that the novel result of the study is the neuroevolutional algorithm that allows you to "grow" the neural network for solving tasks of machine learning with reinforcements. The key feature of the proposed algorithm is a unique opportunity to change the architecture of the neural network by simultaneously adding nodes as well as compounds. The proposed algorithm does not use the crossover operator. Evolution of the network occurs as small mutational changes in a limited area. The advantages of the algorithm are the follows: independence on the type of activation of neuronal functions; no need for the training set; the ability to automatically find a suitable neural network architecture.

The researchers have conducted experiments on the difficult task: the trolley management for balancing of two flagpoles with different lengths fortified on it. The results of the developed algorithm had been compared with the similar algorithm – the NEAT. The developed algorithm is superior to analogue on the following parameters: in the most difficult conditions in the ratio of the lengths of flagpoles 1:2,5 balance control problem is successfully solved by the NEAT algorithm with probability 0,1 and the proposed algorithm with a probability 0.8. Similarly, the complexity of the networks produced by the algorithm HEAT is growing rapidly bit in the proposed algorithm it is almost the same. In addition, the algorithm does not use as an input the trolley speed information and the angular velocity of the flagpoles. The authors consider that the algorithm holds promise for use in evolutionary robotics, searching for game strategies, computer art, and other areas where you want to perform actions in the environment, only occasionally getting feedback.

Acknowledgment

The study was performed as part of the Government assignment of Ministry of Education and Science of Russia (Project № 8.823.2014) and partly supported in by the grant from the Russian Science Foundation (project # 16-07-00335) in the Southern Federal University

References

[1] Stanley K.O., Miikkulainen, R. Evolving Neural Networks Through Augmenting Topologies // Evolutionary Computation. 2002. № 10(2). Pages 99–127.

[2] Stanley K.O.: http://www.cs.ucf.edu/~kstanley/

[3] Kureychik V.V., Kureychik V.M., Rodzin S.I. The theory of evolutionary computation. Moscow: Fizmatlit, 2012. 260 pages.

[4] Tsoy YU.R., Spitsyn V.G. The evolutionary approach to the setting up and training of artificial neural networks // Neuroinformatics. 2006. Vol 1. № 1. pp. 34-61.

[5] L.A. Zinchenko, V.M. Kureychika, V.G. Red'ko. Bionic information systems and their practical application. Moscow: Fizmatlit, 2011. 288 pages.

[6] Rodzin S., Rodzina O.N. Proc. of the 6th IEEE Int. Conf. on Cloud System and Big Data Engineering (Confluence'2016, India). AI and Soft Computing. 2016. pp. 1-5.

[7] Rodzin S., Rodzina L. Theory of bionic optimization and its application to evolutionary synthesis of digital devices // Proc. of the 14th IEEE east-west design & test symposium (EWDTS'14), 2014. pp. 147–152.

[8] Gomez F., Miikkulainen R. Solving non-Markovian control tasks with neuroevolution // Proc. 7th Int. Joint Conf. on Artificial Intelligence, Morgan Kaufmann, San Francisco, 1999. pp. 1356–1361.

[9] Gruau F. Genetic synthesis of Boolean neural networks with a cell rewriting development al process // Proc. of the IEEE Workshop on Combinations of Genetic Algorithms and Neural Networks, 1992. pp. 55-74.

10. Moriguchi H., Honiden S. CMA-TWEANN: Efficient Optimization of Neural Networks via Self-Adaptation and Seamless Augmentation // Proc. of the 14th Annual Conf. on Genetic and Evolutionary Computation, 2012. pp. 903-910.

A Heuristic Algorithm for the Band Collocation Problem

Arif Gursoy
Department of Mathematics
Ege University
Izmir, Turkey
arif.gursoy@ege.edu.tr

Mehmet Kurt
Department of Mathematics and Computer Sciences
Izmir University
Izmir, Turkey
mehmet.kurt@izmir.edu.tr

Hakan Kutucu
Department of Computer Engineering
Karabuk University
Karabuk, Turkey
hakankutucu@karabuk.edu.tr

Urfat Nuriyev
Department of Mathematics, Ege University
Izmir, Turkey
Institute of Control Systems of ANAS
Baku, Azerbaijan
urfatnuriyev@gmail.com

***Abstract*—In this paper we present a heuristic algorithm for the The Band Collocation Problem (BCP) which may have some applications in the field of telecommunication. First, we give the definition the BCP. Second, we explain how we create the problem instances with known optimal solutions as a library. Third, we propose the heuristic algorithm. Then, we analysis and interpret the results of the proposed algorithm on the problem instances with known optimal solutions. Finally, we suggest new ideas about the BCP and its solution approaches.**

***Index Terms*—Band Collocation Problem, Heuristic Algorithm, Telecommunication Applications**

I. Introduction

The Band Collocation Problem (BCP) which is a combinatorial optimization problem aims to determine the wavelength assignments of information packages in a single fiber optic cable to minimize hardware costs in telecommunication networks. The BCP and its combinatorial, nonlinear and linear mathematical models have been proposed by Nuriyev, Kutucu, Kurt and Gursoy [1-3]. Readers may refer to these papers for technical details and mathematical models.

As an extended version of the Bandpass problem, which is introduced by Babayev, Bell and Nuriyev, the BCP is NP-hard [4]. In this paper, we propose a heuristic algorithm to solve the BCP.

The paper is organized as follows. We first present in Section 2 the definition of the BCP with an illustrative example. In Section 3, we create a library of the problem instances with known optimal solutions for the BCP. In Section 4, we improve a simple and very fast heuristic algorithm with an analysis of its time complexity to solve the BCP. We discuss the computational results of the algorithm on the problem instances with known optimal solutions in the library in Section 5 and finally, we suggest new ideas and techniques related to solution algorithms of the BCP that is the subject of future investigations in Section 6.

II. Definition of The BCP

Given a binary matrix A of dimension mxn which represents the data traffic in a fiber optic network, $B_k = 2^k$, called a B_k-*Band*, where ($k=0,1,2,\ldots,t=floor(\log_2 m)$) and a cost c_k of the B_k-*Band* ($2c_{k-1} \geq c_k$) for all k. A B_k-*Band* in a column includes 2^k consecutive entries. Each non-zero element of the matrix A must belong to a B_k-*Band*. However, a B_k-*Band* may have zero elements. Besides, there can not be a common element for any two B_k-*Bands*. That is, B_k-*Bands* in the same column cannot overlap. The BCP consists of finding an optimal permutation of rows of the traffic matrix, which produces the minimum total cost of B_k-*Bands* in all columns.

Consider, as an example, matrix A with m=10 and n=4 is shown in Figure 1. If $c_0=1000$, $c_1=1600$, $c_2=2560$ and $c_3=4100$ for B_0-*Band*, B_1-*Band*, B_2-*Band* and B_3-*Band*, respectively, then the total cost to cover all 1's in all columns by *Bands* is 16280 as it can be seen in Figure 2. Note that B_3-*Band* is not used.

In the BCP, the rows of the matrix can be relocated (renumbered). The question is if there is a row permutation providing less total cost or not if exists what it is. Figure 3 shows a row permutation which has less total cost, 12780. In this relocation of the rows, there is only one B_0-*Band* instead of six used in the original matrix and note that one B_3-*Band* is used.

It is clear that exact solution algorithm of the BCP requires brute-force, i.e., $m!$ permutations. This number grows faster than exponentially with m (see Stirling's formula). That is, it is not reasonable to try to find the optimal solution by exhaustive search. Therefore, we develop fast performing heuristic algorithm of polynomial complexity to generate near optimal or optimal solutions.

III. The BCP Library and Creating Problem Instances With Known Optimal Values

For the BCP, there are many criteria that affect the efficiency of the solution algorithms such as the number of rows, the number of columns, the density of the number of 1's in the matrix. However, increasing rate between the costs c_{k-1} and c_k is also another important factor.

r \ c	1	2	3	4
1	1	1	0	1
2	0	1	1	0
3	0	0	1	0
4	0	1	0	1
5	1	0	0	1
6	0	1	0	1
7	0	1	0	0
8	1	0	1	0
9	0	1	0	1
10	0	0	1	0

Figure 1: A binary matrix

r \ c	1	2	3	4
1	1	1	0	1
2	0	1	1	0
3	0	0	1	0
4	0	1	0	1
5	1	0	0	1
6	0	1	0	1
7	0	1	0	0
8	1	0	1	0
9	0	1	0	1
10	0	0	1	0

Figure 2: Positions of B_k-*Bands*

r \ c	1	2	3	4
10	0	0	1	0
3	0	0	1	0
8	1	0	1	0
2	0	1	1	0
5	1	0	0	1
1	1	1	0	1
6	0	1	0	1
9	0	1	0	1
7	0	1	0	0
4	0	1	0	1

Figure 3: A relocation of the rows and the location of the last bands

Therefore, we determined 72 matrices with different number of rows, number of columns and density of number of 1's named from T1 to T72 for the problem instances. Furthermore, we considered that the balancing of the number of 1's in all columns was important for determining efficiency of the solution algorithms. This point defines the matrix types. If all columns have almost the same number of 1's, then the matrix is balanced. This matrix type is called M1. If the number of 1's in columns are irregular such has some columns have 5, some columns have 20, some columns 12, that is unbalanced. We call this type of matrix M2. We then decided to create some matrices randomly. We call these matrices M3. Eventually, we have 72 matrices and 3 different types for each matrix: T1-M1, T1-M2, T1-M3, T2-M1, T2-M2, …, T72-M1, T72-M2, T73-M3.

Another important parameter is increasing rate of c_k values. We assume that $2c_{k-1} \geq c_k$. But this assumption alone is not sufficient to analyse efficiency of the solution algorithms. Some cost values with different rates may lead the different optimal solutions. For example, we can have the following two configurations satisfiying the assumption:

(case 1) $c_0=1000,\ c_1=1950,\ c_2=3810$ and $c_3=7430,$
(case 2) $c_0=1000,\ c_1=1500,\ c_2=2250$ and $c_3=3380.$

As it can be seen, the increasing rate between c_k and c_{k+1} in case 1 is more than in case 2.

Let us consider the second column in Figure 2 and Figure 3. In Figure 2, for case 1, total minimum cost is 5900 (two B_1-*Bands* for {1,2} and {6,7}, two B_0-*Band* for {4} and {9}), however, for case 2, the cost is 4500 (two B_2-*Bands* for {1,2,3,4} and {6,7,8,9}).

In Figure 3, for case 1, total minimum cost in the second column is 5810 (two B_0-*Bands* for {4} and {6}, one B_2-*Band* for {7,8,9,10}), however, for case 2, it is 3380. That is, if we have case 1, then the optimization rate will be 1.5% and if we have case 2, then it will be 24%. The main result is that the optimization can be varied over a wide range of values with different increasing rates. Therefore, we determine six increasing rate for each matrix to analyze the solution algorithms in further details. Given the initial cost c_o, c_k is calculated for all $k=1,2,\ldots,t=floor(log_2 m)$ as follows:

$$c_{k+1}=(2-\rho) \times c_k \qquad \rho=0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4 \text{ and } 0.5$$

In the library, problem instances are named as Tx-My-Rz where, x is the matrix number, y is the matrix type and z is the increasing rate.

It is obvious that if all 1's in each column is consecutive, then total cost is minimum regardless of the matrix type and the increasing rate. Therefore, in order to create problem instances with known optimal cost value, we prepared 72x3=216 matrices with the properties mentioned above and then relocated their rows. Totaly, we obtained 216x6=1296 problem instances with known optimal value having six different increasing rates.

The Band Collocation Problem Library is available online at http://www.izmir.edu.tr/bps .

IV. A Heuristic Algorithm

In this section, we propose a heuristic algorithm with two different versions. In both versions, the columns are sorted in either ascending or descending order by the sum of the non-zero elements in the columns. The parameters used in the algorithms are as follows:

- m: the number of rows in the matrix,
- n: the number of columns in the matrix,
- A: A Boolean matrix of dimension mxn,
- $k = log_2 m$,
- C: A cost vector of length $k+1$ including $c_o, c_1, \ldots, c_k$

Heuristic Algorithm1:

```
Sort_Columns_Ascending (A);
Boolean_Order(1, m, 1, 1);
BC_Cost_Calculate(A, C);
```

Heuristic Algorithm2:

```
Sort_Columns_Descending(A);
Boolean_Order(1, m, 1, 1);
BC_Cost_Calculate(A, C);
```

Sort_Columns_Ascending and Sort_Columns_Descending procedures sort matrix *A* (relocate the columns of the matrix) in ascending and descending order by the sum of each column, respectively. These procedures runs in *O(nlgn)* time since they use the well-known quicksort algorithm.

The BC_Cost_Calculate procedure takes matrix A and cost vector C as input, and uses a dynamic programming algorithm proposed by Nuriyeva in [5] to evaluate band collocation cost of the matrix after relocations of the rows . This procedure runs in *O(knm)* time.

The core of the heuristic algorithm is the Boolean_Order procedure. It takes top, bottom, column_id and value as input. The top and the bottom are used to determine the range of the rows to be applied, the column_id indicates a column to be performed, and the value may be either 0 or 1. If the value is 0, then Boolean_Order relocates the rows with zero values in the corresponding column on the top of between top and bottom. If the value is 1, then Boolean_Order relocates the rows with non-zero values in the corresponding column on the top of between top and bottom.

```
1  Boolean_Order (top, bottom, column_id, value)
2     location_index = 1;
3     for i = top to bottom;
4        if (A[i][column_id] == value)
5           swap(i, top + location_index++);
6     middle = top  + location_index - 1;
7     newvalue = (value + 1) mod 2;
8     if (middle - top > 1 && column_id < n)
9        Boolean_Order(top, middle, column_id + 1, newvalue);
10    if (bottom - middle + 1 > 1 && column_id < n)
11       Boolean_Order(middle + 1, bottom, column_id + 1, value);
```

The initial call of the Boolean_Order starts with the whole rows in the first column. After the initial call, the procedure calls itself recursively exactly twice for each column, one for the upper part the other for the lower part. In this procedure, the for loop in lines 3-5 selects entries which are the same with the value and move them to the front of the rows. Line 6 determines distinction between 0's and 1's, and line 7 calculates the one's complement of the value. In lines 8-11, the current range are separated out two sub-ranges and two recursive calls are executed. The Boolean_Order procedure runs in *O(nm)* time since it handles each entry of the matrix only once.

V. Test Results of the Heuristic Algorithm

We performed two versions of the proposed heuristic algorithm. The test instances are taken from the BCPLib in [6]. The heuristic algorithm with ascending order gives better results than with descending order. The average gap for ascending order and descending order versions is 7.04% and 7.31%, respectively. Since there is no significant difference between the gaps in both versions, we gave the test results for ascending order. The algorithm has solved to optimality 32 instances out of 1296 and the CPU time varies from 0.02 to 1.92 milliseconds.

As it can be seen in Figure 4 and 5, the gap increases as the number of rows and columns increase. However, In Figure 6, we can observe that the proposed algorithm gives better results when increasing ratio between costs of the Bands is high. In Figure 7, we see that algorithm give better results in low and high densities. As a result, we can say that the algorithm gives better results on matrices which are generated randomly.

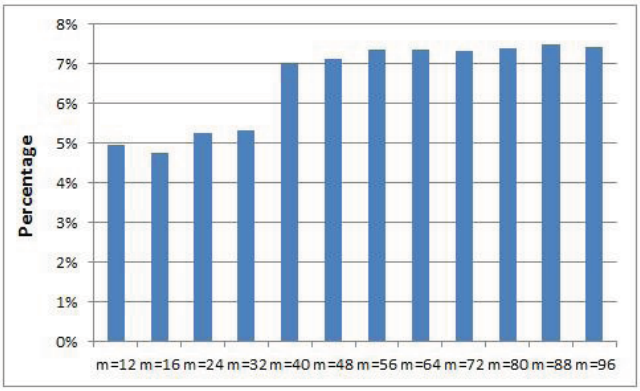


Figure 4: Average gap by the number of rows.

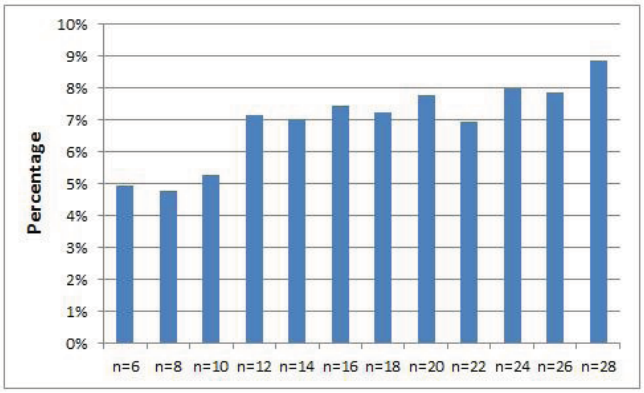


Figure 5: Average gap by the number of columns.

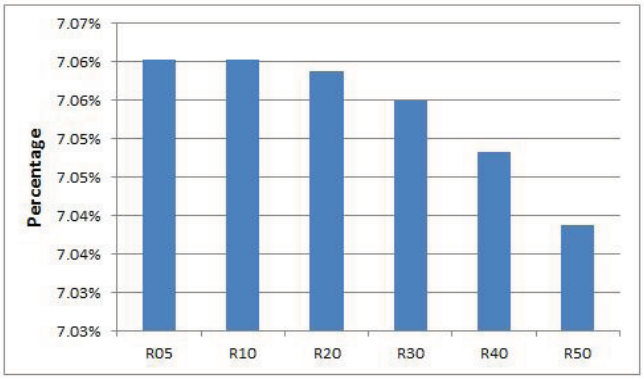


Figure 6: Average gap by the increasing ratio of costs.

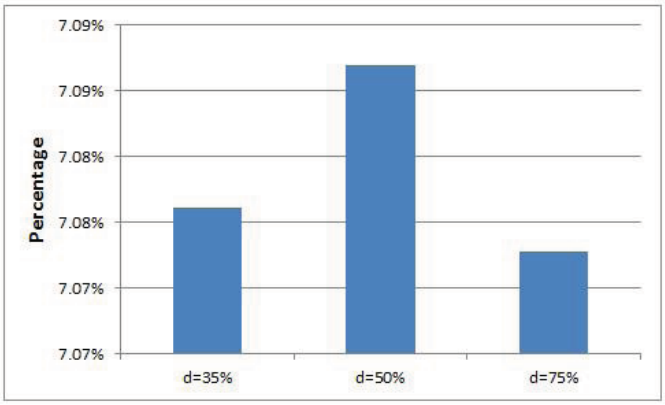


Figure 7: Average gap by the density of 1's.

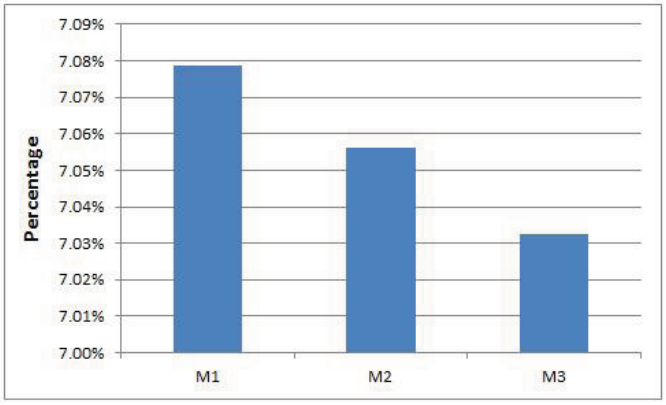


Figure 8: Average gap by the matrix types.

VI. CONCLUSION

In this paper, we developed a heuristic algorithm for the Band Collocation Problem and created 1296 problem instances with known optimal solutions. Then we presented our test results on these instances. In general, the gap results are very promising but not enough. However, the implementation of the recursive algorithm is very fast in runtime.

We are going to study on row similarity degree (RSD) for the new heuristic algorithms. After defining RSD, we would like determine row groups of sizes 2, 4, 8, etc. which would be in sequential order. Besides, we are going to study on sequential degree between groups (SDG) for choosing the groups which would be in a sequential order. Furthermore, metaheuristic algorithms are of particular interest and also merit study.

ACKNOWLEDGMENT

This work is supported by Scientific and Technological Research Council of Turkey – TUBITAK 3001 Project (Project No: 114F073).

REFERENCES

[1] U. Nuriyev, M. Kurt, H. Kutucu and A. Gursoy, "The Band Collocation Problem and Its Combinatorial Model", The International Conference "Mathematical and Computational Modelling in Science and Technology" (ICMCMST 2015), pp. 140-141, Izmir, Turkey, 2015.

[2] U. Nuriyev, H. Kutucu, M. Kurt and A. Gursoy, "The Band Collocation Problem in Tele1communication Networks", The 5th International Conference on Control and Optimization with Industrial Applications, pp. 362-365, Baku, Azerbaijan, 2015.

[3] A. Gursoy, H. Kutucu, M. Kurt and U. Nuriyev, "A Binary Integer Linear Programming Model for The Band Collocation Problem", Theoretical and Applied Aspects of Cybernetics, Proceedings of the 5th International Scientific Conference of Students and Young Scientist TAAC'2015, pp. 131-134, Kyiv, Ukraine, 2015.

[4] D.A. Babayev, G.I. Bell and U.G. Nuriyev, "The Bandpass Problem, Combinatorial Optimization and Library of Problems", Journal of Combinatorial Optimization, 18, pp. 151-172, 2009.

[5] F. Nuriyeva, "On a generalized sequential partially covering problem", Appl. Comput. Math., 15(2) pp. 240-243, 2016.

[6] http://www.izmir.edu.tr/bps/ (Last Access Date: 15.05.2016).

Figuratively-Semantic Support in Precedent-Oriented Solving the Project Tasks

Petr Sosnin
Computer Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
sosnin@ulstu.ru

***Abstract*—The paper deals with semantic modeling a mental imagery caused by processing a new project task in real-time designing a system with the software. In the processing of the new task, the designer applies a framework "model of precedent" and its iterative filling by content with using a feedback caused with a stepwise refinement of the task statement. For understanding the current conditions of the task statement, the designer uses a figuratively-semantic support. Combining the indicated means is implemented in the toolkit WIQA.**

***Index Terms*—Automated designing, conceptual experiment, precedent, question-answering, stepwise refinement**

I. Introduction

For designing a software intensive system (SIS), one of the features is an unpredictable appearance of a new project task when designers collaboratively solve a set of delegated tasks in the real time. Before this situation, these tasks have been appointed to designers in accordance with used rules of managing from a tree of tasks of the SIS project. Moreover, tasks were distributed among designers so that, in current time, some of them are solving more than one task. These working conditions lead to the fact that tasks (delegated to any designer) should be solved not only in parallel but pseudo-parallel modes also. Hence, tasks can form a queue at a workplace of any designer who must rationally interact with this queue. So, any new task must be included into the tree of tasks and, even, into the corresponding queue of tasks if it is necessary.

However, all told above concerns organizational actions that indicate the dynamic specificity of the tree of tasks. An answer to the question about the structure and content of life cycle of the new task is more important. This paper provides an answer to this question. The offered answer is based on a precedent-oriented approach, an implementation of which is coordinated with multitasking in conditions of pseudo-parallel interactions with queues of tasks. The use of stepwise refinement in the work with the new task and iterative creating the model of the corresponding precedent allow including this work in the multitasking mode of the execution at a workplace of the designer who discovered this task.

The stepwise refinement is a powerful paradigm for developing a complex program (task) from a simple program (subtasks) with incrementally adding details [1]. The value of stepwise refinement is caused by the fact that its application is coordinated with the understanding that for a specific task, the designer can build several variants each of which presents the solution of this task at a certain level of abstraction. By performing stepwise refinement, the designer determines interim (conceptual) solutions of the task, beginning from the highest level of its abstraction.

The offered version of the stepwise refinement applies in the context of the precedent model that uses analogies with "intellectually processed conditioned reflexes" [2]. These analogies take into account intertwining the processes of left and right hemispheres of human brains that were a reason for an inclusion a figuratively-semantic support in the life cycle of the new task.

For the use of this feature of the work with the new task and other indicated above, the designer applies the toolkit WIQA (Working In Questions and Answers) that supports conceptual designing of Software Intensive Systems (SISs) [2].

II. Preliminary bases

The proposed version of stepwise refinement has following history. At its first stage, our investigations were concentrated on creating the statement $ST(Z_i, t)$ for the new task Z_i that unpredictably appears in collaborative designing the system with the software. Conducted investigations have led to the development a set of instrumental means described in our paper []. These means support:

1. Including the work with the new task in a multitasking mode that governs the use of the Agile management and human-computer interruptions.
2. Interacting with the statement $ST(Z_i, t)$ as with a dynamic object beginning with "zero-state" $ST(Z_i, t_0)$ when its textual units are absent.
3. Question-answering in the work with textual units of the statement when they are formed.
4. Translating any textual unit in its Prolog-like description.
5. Reflecting the textual unit in its figuratively semantic model.
6. For investigated textual unit of the statement $ST(Z_i, t)$, iterative coordinating the Prolog-like description with the figuratively semantic model.
7. Conceptually algorithmic programming indicated actions.

Thus, at the first stage of the developed approach, we have organized and tested only the version of controlled

intertwining the textual and graphical components of the statement ST(Z_i, t) in conditions of multitasking.

Out conceived approach has suggested the use of the figuratively semantic support in precedent-oriented solving the task Z_i. This feature of the approach was investigated and realized at the second stage, the results of which was described in the paper [].

This stage was aimed at following actions:

1. Analyzing the precedent model of the task from the viewpoint of its development managed by iterative creating the statement ST(Z_i, t).
2. Specifying the steps of the development of the precedent model in terms of increments that decrease the uncertainty of corresponding textual units of the statement.
3. Creating the precedent model with the use of conceptual experiments.
4. Improving and evolving the set of means that support the offered approach.

As a result, iterative actions of the designer were applied for all types of increments used in creating the model of the precedent. It should be noted, in the paper [], stepwise refinement was not mentioned.

The third stage integrates all novelties of the offered approach by the use of stepwise refinement as the specific mechanism of management in the work with the new task in conditions specified above. This paper presents the approach from the viewpoint of its total integrity.

III. Related Works

For our study, the nearest group of related works concerns the stepwise refinement paradigm in its applications. In this group we mark the use of the stepwise refinement for object-oriented programming [5], heap-manipulating of code [1], developing logic programs [6] and implementing the refinement with perfect developer [7]

The experimental side of our study concerns works [8] and [9] in the subject area "Conceptual Development and Experimentation (CD&E)".

Indicated papers prompt a number of solutions that specify intellectual processing the results of conceptual experimenting and the use of conceptually algorithmic programming in iterative forming the precedent model.

In related works, the important place occupies studies of the mental imagery and diagrammatic reasoning. In this group, we mark papers [10] and [11].

These papers help to build the specialized graphical editor that supports three types of visualized images each of which allows expressing as static images so their dynamics.

It should be noted, all papers indicated in this section were used as sources of requirements for developing the set of instrumental means provided the offered version of the figuratively semantic support in solving the new tasks. Any of these papers concerns only a part of the offered approach to interrupting in conditions of multitasking.

IV. Appearance of the New Task

When a person mind is active, it implicitly and explicitly creates signs, the definite composition of which can be estimated by thinking as discovering a situation indicating on the appearance of a new task. If the new task is important, then such situation plays a role of a reason that initiates the interruption of the current activity of the person for switching on the discovered task. By another word, the reason initiates self-interruption by the person for switching on the work with the new task.

So, the new task starts an own life cycle in the mind of the designer. More definitely, it explicitly begins with a composition of signs (signals), for example, as a set of words (w_1, w_2,…, w_k) that appeared in the left hemisphere of designer's brains. These words open the possibility to define a "zero-state" of the new task.

The scheme presented in Fig.1 demonstrates the operational conditions where the new task is discovered and registered with the list of essential keywords.

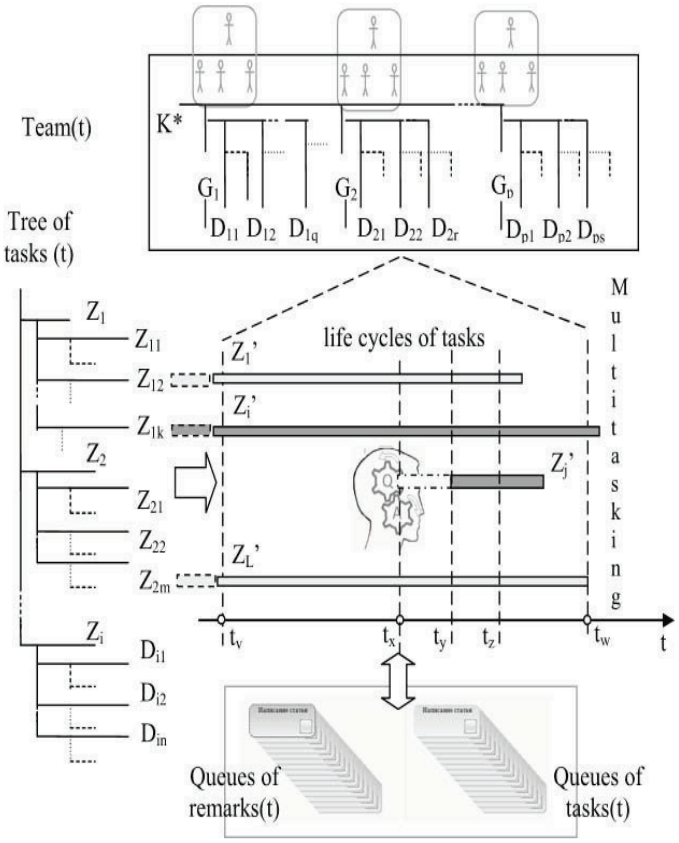


Fig.1. Pseudo-parallel activity of the designer

After that, this task should be included in the collaborative process of conceptual designing the features of which are defined by reflecting the operational space of designing on the semantic memory. In this memory, designers implement the stepwise refinement as for the tree of the task as whole (shown in Fig.2) so for any task by the use of iterative creating the model of the corresponding precedent.

.

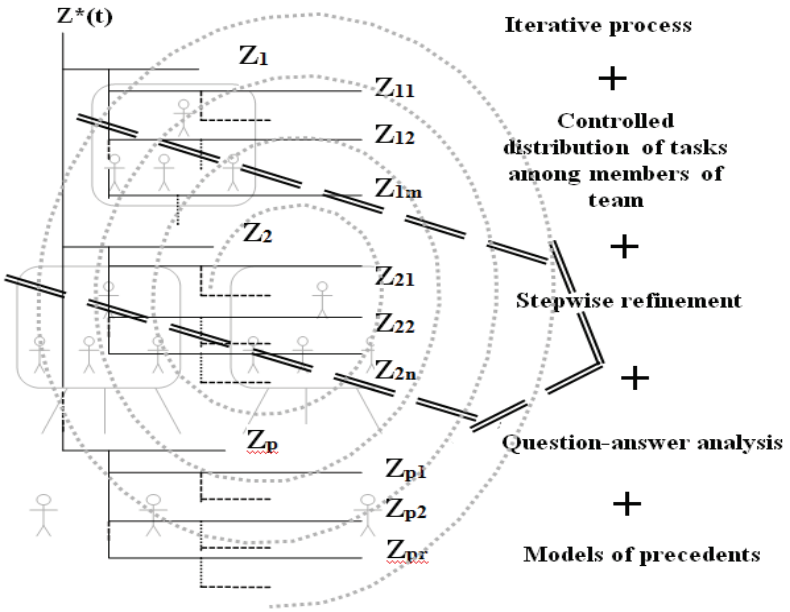


Fig.2. Multitasking environment of designing

The scheme underlines that the centerpiece in the reflection occupies by the tree of project tasks dynamics of which defines the stepwise refinement. Any new task develops current state of the tree. Moreover, for each task of the tree, the responsible designer conducts a stepwise refinement of the question-answer type.

V. STEPWISE REFINEMENT

One important specificity of the used multitasking environment is the use of models of precedents that are generated in designing and accumulated in the Experience Base. The designer creates any of such unit with applying the normative scheme (framework, FP) that is shown in Fig. 3.

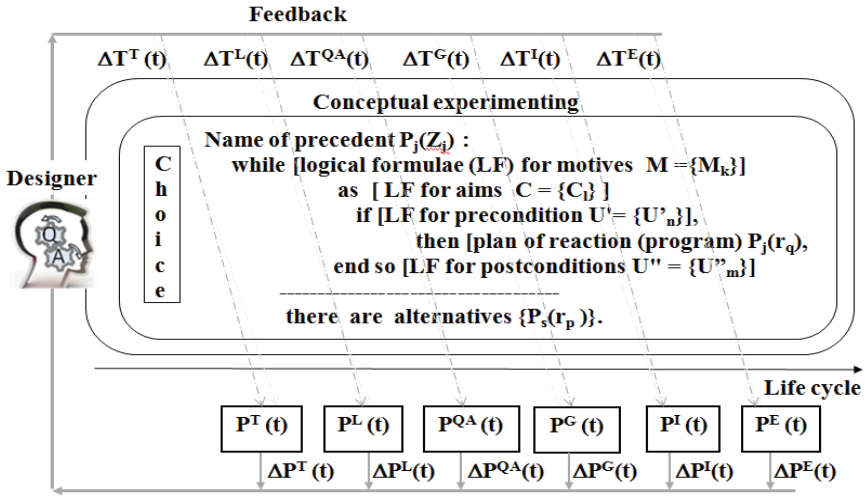


Fig. 3. The model of the precedent

The scheme reflects the structure of the created model MP(Z) and its step by step development in terms of the life cycle. The structure includes following subordinated models: textual model P^T of the solved task; its model P^{QA} in the form of registered QA-reasoning; logical formulae P^L of modeled regularity; graphical (diagram) representation P^G of precedent; pseudo-code model P^I in QA-program form; and executable code P^E.

Note, in the step by step development of MP(Z), the reason for any step is caused by increments of textual units that are created in forming the statement of the corresponding project task.

Let us clarify this feature of the model MP(Z) in detail. As told above, the task Z and its model MP(Z) begins their life with zero-states when even the initial statement of the task is absent.

In our profound conviction, the initial statement must be formulated with the short text that most abstractly (but in the sufficient measure) expresses the essence of the task.

After creating the initial statement of the task, the designer turns to an analysis of its text and implementing the other normative actions of the used technology. During these actions, step by step, the statement S(Z, t) will be enriched while its uncertainty will decrease.

The enrichment will be caused with generating increments $\Delta S(Z_j, t_1)$, $\Delta S(Z_j, t_2)$, …, $\Delta S(Z_j, t_{K-1})$ prioritization of which essentially determines the characteristics of the task being solved. This feature indicates the necessity of managing the development of the task statement.

For managing, the offered approach uses following solutions:

- Using the stepwise refinement way for decreasing the uncertainty of the task statement $S(Z_j, t)$;
- Coordinating the statement development with the process of creating the corresponding model of the precedent.

The first of these solutions is oriented on the use of question-answering in its application for discovering of uncertainty portions, their coding by appropriate questions and decreasing by corresponding answers.

The feature of the second solution is the use of iterative creating the precedent model in the real-time work of the designer with the new task. The designer creates any component of this model on the base of the task statement in its current state S (Zj, t). It may happen that steps of such creation will be sources of information for enriching the state S (Zj, t) or its corrections. This new informational unit can have not only the textual form. They can also be diagrams, pictures, tables, formula and algorithmic units.

So, in its turn, any changing the state S (Zj, t) can be a reason for corrections any of the components P^L, P^{QA}, P^G, P^I, P^E. In other words any increment $\Delta S^x(Z_j, t)$ can lead to the increment $\Delta P^Y(Z_j, t)$ where $Y \in (T, L, QA, G, I, E)$. In its turn, any increment $\Delta P^Y(Z_j, t)$ can be the source of a textual increment $\Delta T^Y(Z_j, t_1)$, that can be a reason for corrections $\Delta S^x(Z_j, t)$ and so on. Thus, components P^T, P^L, P^{QA}, P^G, P^I, P^E are results of the iterative development in coordination with the feedback that is shown in Fig. 3.

VI. Features of Figuratively-Semantic Support

As told above, any step of creating the model MP(Z) is managed by the next increment $\Delta T^Y(Z_j, t_m)$ that should be coordinated with the current state of corresponding components from the set including P^T, P^L, P^{QA}, P^G, P^I, P^E in their states. It is possible that such coordination will require the use of modeling the activated mental imagery that is important or corresponds to the applied normative diagram (for example, the diagram of the UML-set). Analysis and estimations of such necessity have led us to the development of the specialized graphical editor presented in Fig. 4. Any model of the corresponding mental image is bound to its programmed code that helps to build this model repeatable if it will be useful.

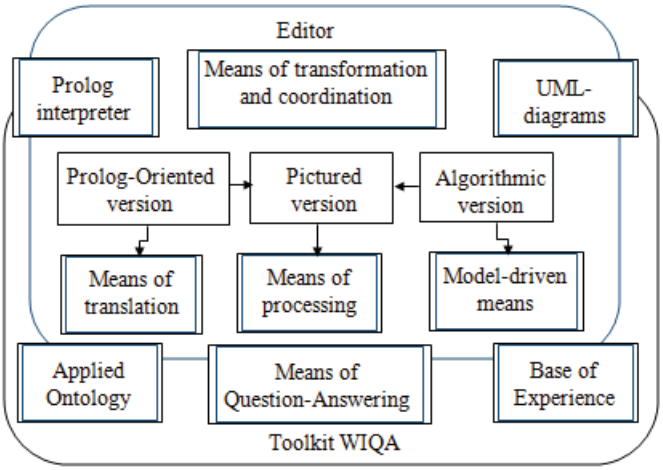


Fig. 4. Structure of the specialized editor

The editor supports following versions of semantic-loaded images:

1. Prolog-oriented versions in the form of the figuratively semantic schemes (FS-schemes) that are coordinated with their Prolog-like descriptions (will be clarified below).
2. Pictured versions any of which can be drown by the designer with the use of units of the chosen paltry.
3. Conceptually algorithmic versions examples of which express the programmatic point of view on the precedent submodel P^I.

Let us clarify the place and role of the FS-Scheme. For implementing the graphical support, the textual unit T_i is translated previously in a Prolog-like description $T^*_i(\{C_{ij}\})$ that includes a set of predicates $\{C_{ij}\}$ each of which presents the corresponding simple sentence embedded in T_i. After that, the description $T^*_i(\{C_j\})$ must be stored in the semantic memory of the toolkit WIQA, and it will be visually accessible to the designer in an interface environment where the designer has the possibility of correcting the description. Then, the Prolog-like description will be interpreted as a declarative program that is written in an extension of the pseudo code language L^{WIQA} embedded in WIQA. Operators of this extension have the syntax that corresponds to the syntax of predicates. Such transformations are iteratively implemented in an operational surrounding that is presented in Fig. 5.

This figure includes clarifying labels. The FS-scheme can be plugged to the ontology for the scheme specification and its informational enriching. Interactions with the ontology help to open the actual variants of the systematization registered in the ontology for nodes of the built scheme.

The scheme also reflects (by the spiral) the dynamics of iterative coordinating the text unit with its FS-scheme. The basis of such coordination is the feedback of FS-scheme with corresponding prolog-like description.

Any correction at the scheme leads to the correction of the text equivalent. This mechanism is useful when translating complex sentences into simple clauses described at their prolog-like versions.

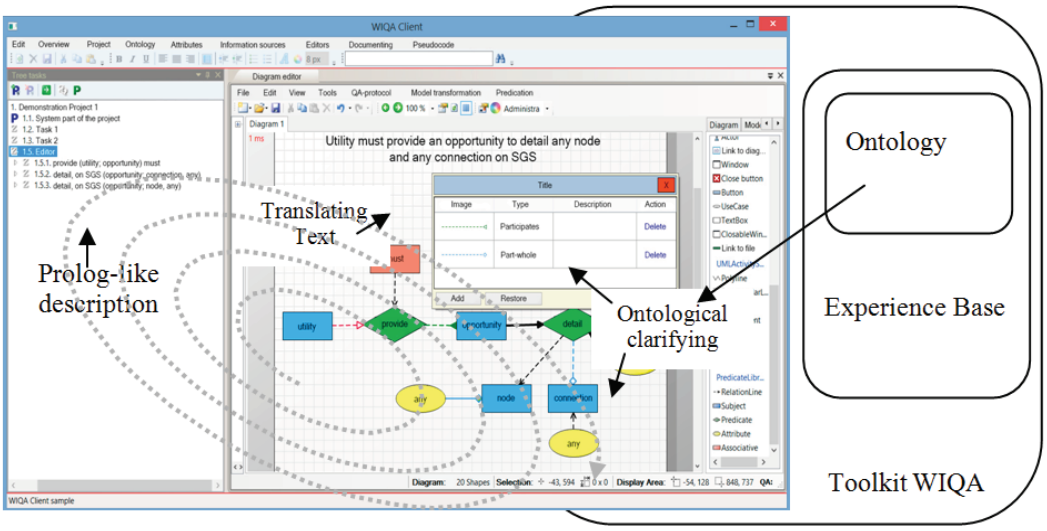


Fig. 5. Iterative coordinating the text and FS-scheme

Note, that the aggregation of the text T, its prolog-like version, and the coordinated graphical model express the understanding that the creator of these constructs embeds in the result of transformations. Achieving the necessary understanding is a fundamental aim of conceptual experimenting.

VII. CONCLUSION

The following conclusions can be drawn from the described study focusing on an unpredictable appearance of a new task at a time when a team of designers develops the system with the software:

1. The appearance of new tasks is typical for real-time designing such systems, and it should be managed with using the agile-means.
2. The use of stepwise refinement mechanisms is a rational way of working as with planned tasks so with new tasks.
3. The work with tasks can be improved by iterative creating the human-oriented models of precedents with using figuratively-semantic support.

By other words, real-time solving the new project task will be more successful if the designer creates its precedent model on the base of analogies with intellectual processing the conditioned reflexes. In the described approach, such position has led to the inclusion of conceptual experimenting in designer's actions.

The approach supports real-time intertwining internal mental processes with their models in a computerized environment, in particular, intertwining mental imagery with images that are semantically useful loaded.

For implementing such activity, the specialized graphical editor has been developed as the extension of the toolkit WIQA with embedded Ontology and Experience Base. The editor supports the creation and use of three types of images with transformations among them. Moreover, any image of any type can be transformed to the executed program. Indicated possibilities help to organize controlled intertwining for achieving the useful aims, first of all, for achieving the necessary level of understanding.

REFERENCES

[1] K. R. M. Leino M. and K. Yessenov, "Stepwise refinement of heap-manipulating code in Chalice", Formal Aspects of Computing, Vol. 24, Issue 4, 2012, pp 519-535
[2] P. Sosnin, "A Scientifically Experimental Approach to the Simulation of Designer Activity." In The Conceptual Designing of Software Intensive Systems," IEEE ACCESS, vol. 1, 2013. pp. 488-504.
[3] P. Sosnin, M. Galochkin, A. Luneckas "Figuratively Semantic Support in Interactions of a Designer with a Statement of a Project Task", Advances in Intelligent Systems and Computing, vol. 451, 2016, pp 257-267.
[4] P. Sosnin, "Processing of a New Task in Conditions of the Agile Management with Using of Programmable Queues", LNCS, vol. 9789, 2016, pp 554-569.
[5] R. Colvin, L. Groves, I. J. Hayes, D. Hemer, R.Nickson and P. Strooper, "Developing Logic Programs from Specifications Using Stepwise Refinement", Lecture Notes in Computer Science, Volume 3049, 2007, pp 66-89
[6] G. Carter, R. Monahan and J.M. Morris, "Software refinement with perfect developer". In: Aichernig BK, Beckert B (eds) Third IEEE international conference on software engineering and formal methods (SEFM 2005). IEEE Computer Society, New York,2005, pp 363–372
[7] A. Tafat, S. Boulmé and C. Marché, "A refinement methodology for object-oriented programs". In: Beckert B, Marché C (eds) Formal verification of object-oriented software, papers presented at the international conference, pp 143–159.
[8] MCM-0056-2010, "NATO Concept Development and Experimentation (CD&E) Process". Brussels: NATO HQ, 2010.
[9] V. Jaros, "Concept Development and Experimentation in project STRUCTURE." In Cybernetic Letters.Brno: Expertia, o.p.s., no. 2, 2011, pp.36-41.
[10] F.W. Mast, E.N. Tartaglia, M.H. Herzog, "Frontiers in Psychology", vol. 3, art. 360, 2013
[11] Michael H.G. Hoffmann. "Cognitive Conditions of Diagrammatic Reasoning" Georgia Tech's School of Public Policy Working Paper Series, 24, 2007.

* This work was supported by the Russian Fund for Basic Research (RFBR), Grant #15-07- 04809a

CPTrustworthiness: New Robust Model for Trust Evaluation in Cloud Computing

Fargana Abdullayeva
Institute of Information Technology of ANAS
Baku, Azerbaijan
farqana@iit.ab.az

***Abstract*—In this paper trustworthiness evaluation issue of cloud provider is considered. A new architectural approach for the trust assessment system is proposed. According to the approach the cloud provider's trust level is calculated by aggregating provider's risk and reputation values. For the demonstration of the capabilities of the proposed method the experimental tests are conducted in Matlab program environment.**

***Index Terms*—Cloud computing; trust model; reputation; cloud provider; CPTrustworthiness.**

I. Introduction

The emergence of cloud technologies paradigm is considered as yet another revolutionary process in Internet environment [1]. According to the definition of this technology given by NIST organization, resources here are presented as a service to users via Internet. For example, SaaS (Software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service). Based on these technologies, users gain access to common computer resources (network, storage, program applications) pool on-demand [2].

In general, cloud environment is constructed on multi-provider and multi-service scenarios. As the access of each cloud is a management mechanism, providing access of one cloud (domain) users to the resources of another cloud is considered as an important issue. For this purpose, availability of management mechanisms of flow of user identifiers between clouds is necessary. One of these mechanisms is the federative management mechanism serving to provide secure mutual communication of clouds.

The main goal of federative management is to share and distribute the identity information among different domains within the framework of the specified safety policy. For this reason, a federative model enables the provision of secure access to other domain resources of roaming users of a particular domain and request without any additional login process [3].

The most popular federated identity management mechanism is single sign-on (SSO). This technology allows users to authenticate at a single domain and gain access to multiple ones without providing additional information.

Ample amount of research has been conducted on the development of federated management mechanisms [4]. In this mechanisms approaches to ensure dynamic federation of entities do not provided. In other approaches models for providing dynamic federation of clouds are proposed [5, 6, 7]. But these models unable to operate as a single system, which allows making decision of the federation. Dynamic federation of entities are providing through the use of trust models [8].

Trust is a fundamental issue to address scalability. Moreover, the flexibility of every federation framework is tied to the underlying trust model, often poorly defined or even out of the specifications scope. For this reason, new enhanced techniques are required to achieve dynamic federation. Furthermore, the significance of research on this topic has been recently highlighted to the point of stating that "If dynamic federation negotiation and trust management in IdM systems could be achieved, it would revolutionize the Internet marketplace" [9].

In [8] closest model to our work was offered. However, this work by applying MAUT (Multi-Attribute Utility Theory) theory seeks to determine the trust level based on aggregation of two same type (trustworthiness, assurance) factors. This complicates the process of determination how much the provider is safe. Here an aggregation method also is chosen not successfuly. For the same values of two variables the model also loses its accuracy.

In this paper a new dynamic trust model for the identity federation in the cloud computing is proposed.

II. Trust Evaluation System Architecture

Trustworthiness evaluation of cloud provider is usually conducted by multiple attributes. Therefore, an evaluation of the trustworthiness have been resolved through multi-criteria decision-making methods. In this paper approach assessing trustworthiness of cloud provider (CPTrustworthiness) via the combination of multisource risk and reputation data is proposed. CPTrustworthiness helps discover more trustworthy cloud provider more accurately. CPTrustworthiness model has been described in figure 1 and solves cloud provider trustworthiness evaluation problem in three phases.

Firstly, in the risk attributes block of model risk factors that may affect the violation of the security of the cloud provider's infrastructure are selected, weight ratios to these factors are assigned. Secondly, based on the Mamdani type

fuzzy inference algorithm the risk value of provider's infrastructure is calculated. In this phase, model aggregates reputation scores given to cloud provider as well as calculates the value of his reputation. Finally, the trustworthiness value of cloud provider will be released by combining multi-source assessment data.

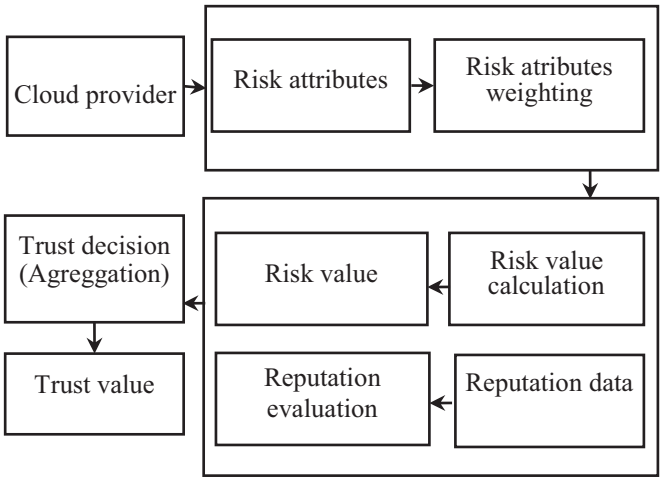


Figure 1. Robust CPTrustworthiness model for trust evaluation

In the proposed architecture for dynamic federation management, the Trust Decision is a key component.

The function of this module is deciding whether to initiate or not a federation with another entity.

The inputs for this module are the risk and reputation values of the entity (figure 2). Then, the decision procedure consists of aggregating two metrics in order to obtain a final meaningful number that can be compared to a decision threshold.

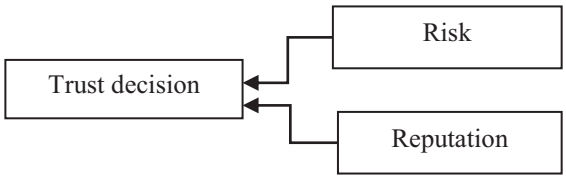


Figure 2 . Trust metrics

Here risk and reputation values are obtained via following algorithms [5, 6].

III. Cloud Provider's Risk Value Calculation

Here cloud provider's risk priority vectors are formed and on the basis of this vector fuzzy logic based risk value calculation is provided.

AHP algorithm is interpreted as follows:

Step 1. *Development of decision hierarchy.* Our study suggests a five-layered hierarchical structure [5]. The objective, placed in level 1 of the hierarchy is the provider's risk value. Second level of hierarchy includes 2, and third level includes 5 factors, that enter into these decision: data security and privacy risks, organizational risks, technical risk, compliance and audit risks, physical security risks. The next layers of the hierarchy are sub factors of the main factors.

Step 2. *Establishment of comparison matrix for each layer.* In this step, carried out the establishment of dominance rates matrix, based on a 9-point system ranging from 1 to 9.

$$A = \left[a_{ij}\right]_{n\times n} \qquad (1)$$

where, a_{ij} – upper diagonal elements of the comparison matrix, $a_{ji} = \frac{1}{a_{ij}}$ lower diagonal elements of the comparison matrix.

Step 3. *Establishment of normalized pairwise comparison matrix.* Normalized comparison matrix is determined by dividing each element of the matrix A by its column total. Assume the sum of j -th column elements is $\sum_{i=1}^{n} a_{ij}$, then

$$a_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \qquad (2)$$

Step 4. *Calculation of weight vectors for the factors.* Weight vectors of factors are determined by averaging the elements on each row of normalized comparison matrix. Weight ratio of row i is calculated as follows:

$$w_i = \frac{\sum_{j=1}^{n} a_{ij}}{n} \qquad (3)$$

where n - is the number of factors.

Step 5. *Calculation of principal Eigen value.* Principal Eigen value is obtained from the summation of products between each element of weight vector and the sum of columns of the decision matrix A .

$$\lambda_{\max} = \sum_{i=1}^{n} w_i \times a_{ij} \qquad (4)$$

Step 6. Calculation of consistency index (CI) and consistency ratio (CR).

$$CI = \frac{\lambda_{\max} - n}{n-1} \qquad (5)$$

$$CR = \frac{CI}{RI} \qquad (6)$$

where, n - is the number of factors, RI is random consistency index and is determined by Saaty as in following Table I.

TABLE I.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

Having obtained the risk priorities vector of cloud provider, we are able to calculate the risk value of cloud provider according to fuzzy logic of fuzzy set theory.

Fuzzy risk assessment

Fuzzy logic inference process for risk assessment can be described as a system which contain following blocks (Figure 3).

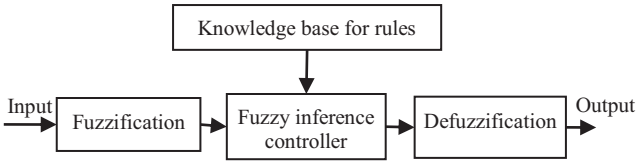


Figure. 3. Procedures of Fuzzy logic for cloud risk assessment

In this paper Mamdani type fuzzy logic inference algorithm is used. Mamdani type fuzzy logic inference model mainly contains the following five assessment steps:

Step 1. *Fuzzification.* In this step determination of main parameters which become necessary for risk assessment is performing. Due to uncertainty nature of these parameters their measurement are too complex. Therefore, the measure of each parameter is shown by linguistic terms and transforming to the appropriate fuzzy number.

In this study, as shown in Figure 4, we adopt triangular membership function.

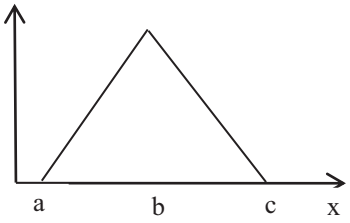


Figure. 4. The triangular membership

A triangular membership function is specified by three parameters $\{a, b, c\}$:

$$f(x;a,b,c)=\begin{cases}0, & x\le a\\ \dfrac{x-a}{b-a}, & a\le x\le b\\ \dfrac{c-x}{c-b}, & b\le x\le c\\ 0 & c\le x\end{cases} \qquad (7)$$

By using the defined membership functions, we replace the input values with a set of linguistic values and assign a membership degree for each linguistic value using triangular membership functions.

Step 2. *Construction of fuzzy rules.* A fuzzy rule can be defined as a conditional statement in the form: "IF x is A THEN y is B" where x and y are linguistic variables and A and B are linguistic values determined by fuzzy sets on the universe of discourses X and Y, respectively. In this study, the Fuzzy logic system is represented with three fuzzy sets Low, Medium, High. These fuzzy sets determine the shape and location of the membership functions. As shown in Figure 5, we give the fuzzy sets of cloud security level risk.

Step 3. *Inference.* The inference engine makes decisions based on fuzzy rules. In other words, in this step calculation of output parameters for the rules are conducting here. For example, rule output parameter $B_i'(y)$ for the i-th rule "*If x is A_i then y is B_i*" is represented by following formula

$$B_i'(y)=\sup_{x\in X}(T(A'(x),T(A_i(x),B_i(y)))) \qquad (8)$$

where $A'(x)$ is the system input parameter, x is elements of the universal X set of input parameters of the system and y is elements of the universal Y set of output parameters of the system.

In this study, the inference engine for main block makes decisions based on 15 fuzzy inference rules as shown in figure 5.

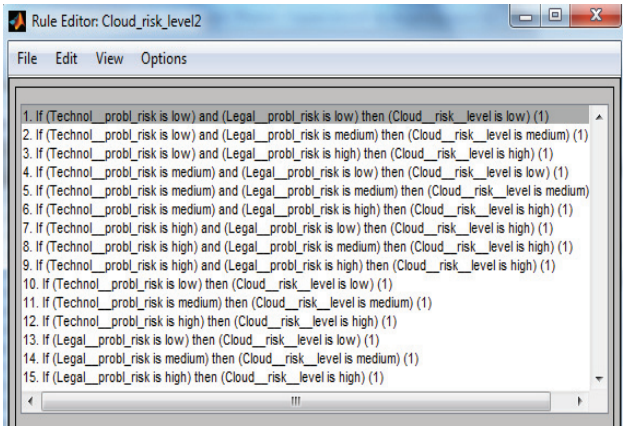


Figure 5. The fuzzy rulers defined for cloud security level risk subsystem

Step 4. *Agregation.* Single output of rule knowledgebase is obtaining by aggregating of $B_i'(y)$ output parameters of all rules and calculating by the following formula

$$B_{out}'(y)=S\left(B_n'(y),S\left(B_{n-1}'(y),S\left(\ldots,\ S\left(B_2'(y),B_1'(y)\right)\right)\right)\right) \qquad (9)$$

Step 5. *Defuzzification.* In this step, implementation of transformation of the linguistic value of cloud risk level into crisp risk values is carried out. We adopt the most common defuzzification method, called center of gravity to obtain the risk value of cloud provider with a value in the range $[0, 1]$.

$$y_{out}=\int\frac{B'_{out}(y)\,zdz}{B'_{out}(y)\,dy} \qquad (10)$$

IV. Cloud Provider's Reputation Value Calculation

In the environment where the number of entities is large, the decisions are specified by the group (collaborative) making decisions based on the specific criteria. The main idea of collaborative method is enabling the full use of knowledge and the intellect of group members for the purpose of efficient decision-making based on the criteria specified. In the problem stated, the group members consist of entities listed in SP trust list which possess a particular knowledge regarding IdP and act as experts in the system. In suggested model, those experts are substituted with intellectual agents.

Assume that n number of intellectual agents assessing IdP and m number of criteria are present. Then, intellectual agent-based multi-criterion collaborative model can be described as in Table II [6].

TABLE II. THE DRAFT OF INTREP MODEL

Entity listed in SP trust list / Identity providers	Intellectual agent 1			Intellectual agent 2			...	Intellectual agent n		
	c_1	...	c_m	c_1	...	c_m	...	c_1	...	c_m
1										
...										
N										

Here, intellectual agents assign different scores to IdP based on $X=(c_1,c_2,...,c_m)$ criteria set as usual. For example, those scores can be constituted of high scores, low scores and average scores according to provider's behaviour information.

In order to include the different opinions of intellectual agents and enable the use (participation) of all opinions in the process of assessment those opinions are fuzzified. For this purpose, the following fuzzy triangle value is used in order to describe the opinions of intellectual agents:

$$\widetilde{Z}_i=(z_{i1},z_{i2},z_{i3})= (lows\,core, averagescore, high\,score) \quad (11)$$

here, i is a numerical index of intellectual agent.

As noted, here individual intellectual agents generate different opinions using different parameters. In the research work suggested, an intellectual agent produces $Y=f(x)$ output score based on X assessment criteria set and this score is used as a reputation score. Here, the classification of provider as reliable or malicious is carried out according to frontier score T. According to federalization policy, only the identity provider with reputation score exceeding frontier score T is allowed to be contacted.

Assume that i^{th} intellectual agent has generated k number of different scores based on X criteria set for $IdP_1: f_1(X_A), f_2(X_A),..., f_k(X_A)$. In order to provide the participation of all scores during the assessment a fuzzy opinion is formed as a triangular fuzzy number. In this case, the lowest, average and the highest scores among the $\mathbf{k}$ number of score scores are used as left, center, and right belonging degrees. It is to be noted that in order to determine the degree of belonging other fuzzification methods can be applied. In this case, several other scores are not considered which results in loss of useful information. Hence, in order to include all scores in fuzzy assessment, the average value of scores are accepted as predicted score. Using this fuzzification method, intellectual multi-agents can carry out a fuzzy assessment for each provider. Hence, fuzzy triangle score for intellectual multi-agents can be described as following:

$$\widetilde{Z}_i=(z_{i1},z_{i2},z_{i3})=\left(\left[\min\left(f_1^i(X_A),f_2^i(X_A),...,f_k^i(X_A)\right)\right],\left[\sum_{j=1}^{k}\frac{f_j^k(X_A)}{k}\right],\left[\max\left(f_1^i(X_A),f_2^i(X_A),...,f_k^i(X_A)\right)\right]\right) \quad (12)$$

The opinions of multi-agents are fuzzified with this method and the problem of provider assessment is posed as fuzzy decision-making process. Assume that p number of intellectual agents is present, and $\widetilde{Z}=\psi(\widetilde{Z}_1,\widetilde{Z}_2,...,\widetilde{Z}_p)$ – is an aggregation of p number of fuzzy opinions, $\psi(\bullet)$ - is an aggregation function. There are existing innumerable aggregation functions than can be used in aggregation of fuzzy scores. Some of them are linear, whereas some of them are not. Using linear additive procedures, the aggregation of fuzzy opinions of p number of group members are carried out as following:

$$\widetilde{Z}=\sum_{i=1}^{p}w_i\widetilde{Z}_i=\left(\sum_{i=1}^{p}w_iz_{i1},\sum_{i=1}^{p}w_iz_{i2},\sum_{i=1}^{p}w_iz_{i3}\right) \quad (13)$$

here, w_i is the weight of i^{th} fuzzy opinion, $=1,2,...,p.$

For the aggregation of weights of opinions given by the agents, they must be normalized in a particular way. Otherwise, malicious entities may assign high weights to other malicious entities, and low weights to trustworthy entities. Here, the weights are normalized as following:

$$\sum_{i=1}^{p}w_i=1 \quad (14)$$

The next stage is the provision of aggregation of fuzzy opinions produced by different multi-agents. For this purpose, the method of least squares is applied. The main idea of this approach is based on the minimization of the sum of squares of distances from one fuzzy opinion to another. In special case, the square of the distance between $\widetilde{Z}_i$ and $\widetilde{Z}_j$ is calculated as follows:

$$d_{ij}^2=\left(\sqrt{\left(w_i\widetilde{Z}_i-w_j\widetilde{Z}_j\right)^2}\right)^2=\sum_{l=1}^{3}\left(w_iz_{il}-w_jz_{jl}\right)^2 \quad (15)$$

According to rule (15), the optimization model serving to the minimization of sum of squares of the distances between all weighted fuzzy opinions pairs can be describes as following:

$$D=\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p}d_{ij}^2=\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p}\left[\sum_{l=1}^{3}\left(w_iz_{il}-w_jz_{jl}\right)^2\right] \quad (16)$$

here, $\sum_{i=1}^{p}w_i=1, \quad (17)$

$w_i\geq 0, i=1,2,...,p. \quad (18)$

Using Lagrange multiplier method, Lagrange function is constructed as follows based on (16) and (17) formulas:

$$L(w,\lambda)=\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p}\left[\sum_{l=1}^{3}\left(w_iz_{il}-w_jz_{jl}\right)^2\right]-2\lambda\left(\sum_{i=1}^{p}w_i=1\right) \quad (19)$$

By differentiating the equation (19) with w_i weight, the following formula is obtained:

$$\frac{\partial L}{\partial w_i}=2\sum_{j=1,j\neq i}^{p}\left[\sum_{l=1}^{3}\left(w_iz_{il}-w_{jz_{jl}}\right)z_{il}\right]-2\lambda=0 \quad (20)$$

here, $i = 1,2,..., p$.

Equation (20) can be simplified as follows:

$$(p-1)\left(\sum_{l=1}^{3} z_{il}^2\right) w_i - \sum_{j=1, j\neq i}^{p}\left[\sum_{l=1}^{3} z_{il} z_{jl}\right] w_j - \lambda = 0,\ i = 1,2,..., p. \quad (21)$$

Assume that $W = (w_1, w_2, ..., w_p)^T$, $I = (1,1,...,1)^T$,

here, T superscript indicates the transposed matrix, and assume that,

$$b_{ij} = (p-1)\left(\sum_{l=1}^{3} z_{il}^2\right), i = j = 1,2,..., p,$$

$$b_{ij} = -\sum_{l=1}^{3}(z_{il} z_{jl}), i, j = 1,2,..., p; j \neq i.$$

In this case, the matrix containing these elements can be described as following:

$B = (b_{ij})_{p\times p} =$

$$= \begin{bmatrix} (p-1)\left(\sum_{l=1}^{3} z_{1l}^2\right) & -\sum_{l=1}^{3}(z_{1l} z_{2l})... & -\sum_{l=1}^{3}(z_{1l} z_{pl}) \\ -\sum_{l=1}^{3}(z_{2l} z_{1l}) & (p-1)\left(\sum_{l=1}^{3} z_{2l}^2\right)... & -\sum_{l=1}^{3}(z_{2l} z_{pl}) \\ -\sum_{l=1}^{3}(z_{pl} z_{1l}) & -\sum_{l=1}^{3}(z_{pl} z_{2l})... & (p-1)\left(\sum_{l=1}^{3} z_{pl}^2\right) \end{bmatrix} \quad (22)$$

Based on matrix B and the conditions above the formulas (21) and (17) can be written as below:

$$BW - \lambda I = 0, \quad (23)$$

$$I^T W = 1. \quad (24)$$

Similarly, equation (16) can be expressed in a matrix form as $D = W^T B W$.

The following are obtained based on formulas (23) and (24):

$$\lambda^* = 1/(I^T B^{-1} I) \quad (25)$$

$$W^* = (B^{-1} I)/(I^T B^{-1} I) \quad (26)$$

After the completion of the aggregation process of weights, a fuzzy mean opinion is formulated using (13) formula.

In order to obtain the crisp value for reputation score, fuzzy triangle $\tilde{Z} = (z_1, z_2, z_3)$ is defuzzified to crisp value by using the center method.

$$z = \frac{\int_{z_2}^{z_3} x\mu_z(x)dx}{\int_{z_2}^{z_3} \mu_z(x)dx} = \frac{\int_{z_1}^{z_2}\left(x \bullet \frac{x - z_1}{z_2 - z_1}\right)dx + \int_{z_2}^{z_3} x \bullet \frac{z_3 - x}{z_3 - z_2}}{\int_{z_1}^{z_2}\left(\frac{x - z_1}{z_2 - z_1}\right)dx + \int_{z_2}^{z_3}\left(\frac{z_3 - x}{z_3 - z_2}\right)dx} = \frac{(z_1 + z_2 + z_3)}{3} \quad (27)$$

V. Trustworthiness Evaluation

Trustworthiness of the cloud provider is obtained via aggregation of reputation and risk values. Therefore, the following aggregation method is used.

Assume cp_n is the n[th] cloud provider, $r_1, r_2, ..., r_n$ - provider's risk values, $rp_1, rp_2, ..., rp_m$ are provider's reputation values. n[th] cloud provider's cp_n trust value is calculated as follows:

$$cp_n = \frac{\sum \omega \times r_n}{I} + \frac{\sum \omega_j \times rp_m}{J - I - 1}, \quad (28)$$

here ω_i and ω_j are appropriate weights of reputation and risk attributes and $\sum_{i=1}^{I} \omega_i + \sum_{j=1}^{J} \omega_j = 1$. Weight provides robustness of the model against attacks. Equation (28) defines the aggregation method of risk and reputation values under the constraint $\sum_{i=1}^{I} \omega_i + \sum_{j=1}^{J} \omega_j = 1$. Here I is the number of risk parameters, in our case four risk parameters are used. J is the number of reputation parameters, in our case ten reputation parameters are used.

VI. Experimental Verification of the Suggested Method

Assume that the provider's reputation and risk attributes get values in [0, 1] interval. By applying equation (28) over these values the relationship between risk and reputation parameters can be build as following aggregation matrix.

TABLE III. Reputation and Risk Parameters Relation Matrix

rp\r	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.01	0.0083	0.0080	0.0079	0.0076	0.0075	0.0074	0.0072	0.0071	0.0063	0.0061	0.0051
0.1	0.0110	0.0107	0.0106	0.0103	0.0102	0.0101	0.0099	0.0098	0.0091	0.0088	0.0078
0.2	0.0158	0.0155	0.0154	0.0151	0.0150	0.0149	0.0147	0.0146	0.0138	0.0136	0.0126
0.3	0.0126	0.0123	0.0122	0.0119	0.0118	0.0117	0.0115	0.0114	0.0107	0.0104	0.0094
0.4	0.0218	0.0215	0.0214	0.0211	0.0210	0.0209	0.0207	0.0206	0.0198	0.0196	0.0186
0.5	0.0149	0.0146	0.0145	0.0142	0.0141	0.0140	0.0138	0.0137	0.0130	0.0127	0.0117
0.6	0.0148	0.0145	0.0144	0.0141	0.0140	0.0139	0.0137	0.0136	0.0128	0.0126	0.0116
0.7	0.0147	0.0144	0.0143	0.0140	0.0139	0.0138	0.0136	0.0135	0.0128	0.0125	0.0115
0.8	0.0138	0.0135	0.0134	0.0131	0.0130	0.0129	0.0127	0.0126	0.0118	0.0116	0.0106
0.9	0.0132	0.0129	0.0128	0.0125	0.0124	0.0123	0.0121	0.0120	0.0113	0.0110	0.0100
1	0.0083	0.0080	0.0079	0.0076	0.0075	0.0074	0.0072	0.0071	0.0063	0.0061	0.0051

As we can observe in the values represented in table 3, there are cases in which very low values of risk combined with high level of reputation the provider's trustworsiness is high, high value of risk combined with low value of reputation the provider's trustworsiness is low.

Surface model allowing decision making about provider's trustworthines via this agregator is represented figure 6.

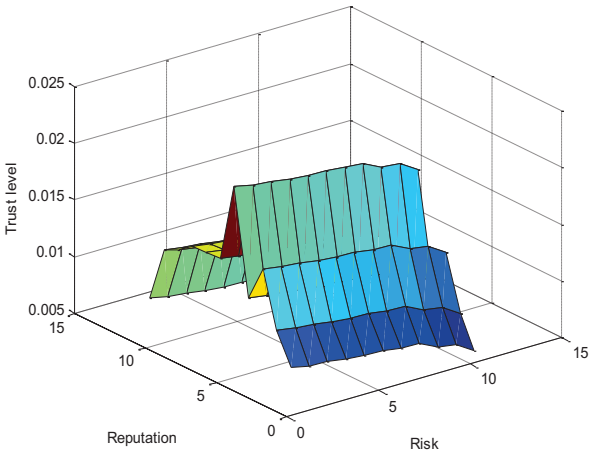


Figure 6. Surface model for trust decision making

As we can observe in the surface model represented in figure 6 there are cases in which very low values of risk combined with high level of reputation tead to final acceptable values. For example, for r=0.2 and rp=0.7 provider's trustworthiness is equal to 0.0143.

In the calculations the following weight ratios were used (table IV)

TABLE IV. WEIGHT RATIOS

ω_i	0.29	0.25	0.134	0.85	0.062	0.049	0.039	0.033	0.024	0.02	0.013
ω_j	0.1	0.19	0.2	0.091	0.16	0.073	0.06	0.051	0.04	0.033	0.01

ACKNOWLEDGEMENTS

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – **Grant № EİF-2013-9(15)-46/16/1**

REFERENCES

[1] R. Alguliev, F. Abdullaeva, "User profiles and identifying user behavior in the cloud computing environment," Universal journal of communications and network, 2014, vol. 2, no 5, pp. 87-92.

[2] SP. 800-145. The NIST definition of cloud computing, 2011, 7p.

[3] E. Maler, D. Reed, "The Venn of identity: Options and issues in federated identity management," IEEE Security & Privacy, 2008, vol. 6, no 2, pp. 16-23.

[4] D.W. Chadwick, "Federated identity management. Foundations of security analysis and design, Lecture Notes in Computer Science, 2009, vol. 5705, pp. 96-120.

[5] R. Alguliyev, F. Abdullayeva, "Development of risk factor management method for federation of clouds," Proc. of the IEEE international conference on connected vehicles and expo (ICCVE), 2014, pp. 24-29.

[6] R. Alguliyev, F. Abdullayeva, "Development of novel robust reputation evaluation method for dynamic federation of clouds," Social and basic sciences research review, 2016, pp. 56-62.

[7] P. Cabarcos, F. Almena´rez, A. Lo´pez, D. Dı´az-Sa´nchez, R. Guerrero, "A metric-based approach to assess risk for ''On Cloud'' federated identity management, Journal of network and systems management, 2012, vol. 20, no 4, pp. 513-533.

[8] P. Cabarcos, F. Almenárez, F. Mármol, A. Marín, "To federate or not to federate: A reputation-based mechanism to dynamize cooperation in identity management, Wireless personal communications, 2014, vol. 75, no 3, pp. 1769-1786.

[9] Identity and access management for Networks and Services; Dynamic federation negotiation and trust management in IdM systems, ETSI, 2011, 18 p.

Virtual cloud network laboratory based on IaaS with automatized creation of network topology on demand

Yu.A. Ushakov, P.N. Polezhaev, L.V. Legashev, A.E. Shukhman, I.P. Bolodurina
Department of Mathematics and Computer Science
Orenburg State University
Orenburg, Russian Federation
shukhman@gmail.com

***Abstract*—The article describes an approach to create virtual cloud network laboratory. The topology is created with the visual editor. Then a script for network deployment is created. It is based on the scheme description. For implementation of this approach the cloud system based on OpenNebula has been created. The experiment has shown the efficiency of using cloud systems in IT education. Even the small cloud datacenter has provided virtual classes for studying operating systems, network administration, cloud services and creating various cloud services and networks.**

Keywords—virtual laboratory, IT educaton, IaaS.

I. INTRODUCTION

The present day university IT education implies students' access to virtual laboratories. Virtual machines joined in the network are used for practical study of server OS such as Windows and Linux.

The most popular virtualized networking projects (such as GENI, VNET, VNRMS) are considered in the paper [1]. However, we offer a simpler solution for training. It uses a local cloud data center.

Any class requires a certain set of nodes and a topology of their connections. Managed routers and switches used in such networks make access formation complicated. Non-standard topologies and components exclude the use of standard physical schemes. The standard hypervisors (MS Hyper-V, VMware vSphere, Xen, KVM) and cloud aggregators (OpenNebula, OpenStack) provide no possibility to create and deploy the topology on demand using conventional methods.

Cloud services provide the use of any needed number of virtual machines connected to the networks by Open vSwitch (OVS). The use of OpenFlow in cloud infrastructure allows to create any topologies independent from physical constraints. The Open vSwitch is integrated in the topology as a managed switch under the same controller. Nevertheless, such technologies are not supported widely enough as yet.

The system described here is meant for automatized creation and deployment of laboratory and industrial virtual networks with any topology based on OpenNebula cloud platform.

II. PROBLEM DESCRIPTION

The problem of automatized creation and deployment of virtual infrastructure includes creating and editing topology schemes according to current cloud infrastructure possibilities, and converting the schemes into sets of command scripts to deploy infrastructures on demand.

Automatization methods for V-Labs topology creation are presented in some papers. The UNetLab project [3] provides the virtual topology based on single VM. This project uses Qemu, Dynamips and Open vSwitch to emulate real topology and web-based SSH for access to devices. This approach is limited to local memory and CPUs within the single VM. Our project also uses similar Web-based approach for visual topology creation, but we simplified interface for old computer compatibility.

Sometimes, linked virtual nodes must allocate virtual space on different hosts within the common network. Most open source cloud tools use only a real network (VLANs, VXLANs) for segment connection. We have tested GRE based method within the heterogeneous topology [4] to create borderless common network space without any hardware limitation. This method uses virtual GRE links over any real network topology.

A complex management system for both virtual and real networks is presented in our previous paper [5]. This system is based on SDN approach. Within an OpenFlow converged network the controller can make routing decisions without any overhead. In [6] authors use OpenFlow for real-time IP monitoring and routing without MAC address binding. The traffic control method is described in [7]. It works up to L7 and provides protocol-based routing. It can resolve QoS and balancing problem in high-load cloud network to optimal route all the critical network protocols.

The visual design problem with a certain set of components having preset constraints has been considered in a number of researches. The simplest component for use and integration is Draw2D touch component [2] supported in all the browsers. The use of the component allows to store the schemes in JSON format and visually edit them in the browser (Fig. 1).

The description of this example is given in Fig. 2. Due to the simple scheme description and graphic representation, it is possible to build more complex schemes with many interconnected components.

The research has been supported by the Russian Foundation for Basic Research (projects 15-07-06071 and 16-07-01004), and the President of the Russian Federation within the grant for young scientists and PhD students (SP-2179.2015.5).

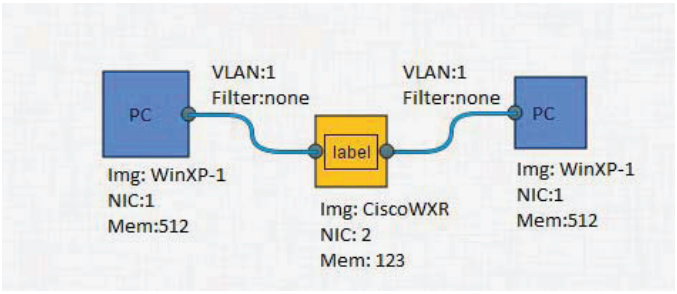


Fig. 1. The example of simple network.

```
{
  "type": "draw2d.shape.basic.Rectangle",
                  //Rectangle PC1
  "id": "354fa3b9-a834-0221-2009-abc2d6bd852a",      //ID
  "userData": {IMG:WinXP-1, NIC:1, MEM: 512},       //Load parameters
  "ports": [{                                       //Ports
    "type": "draw2d.Port",                          //Type
    "id": "ebfb35bb-5767-8155-c804-14bd48789dc21",  //ID
    "userData": {name:eth0,speed:100tx,vlan:1},     //Settings
    "name": "eth0",                                 //Name
   ]},{
  "type": "draw2d.shape.basic.Rectangle",           //Rectangle PC2
  "id": "354fa3b9-a834-0221-2009-abc2d6bd852a",      //ID
  "userData": {IMG:WinXP-1, NIC:1, MEM: 512},       // Load parameters
  "ports": [{
    "type": "draw2d.Port",                          //Type
    "id": "ebfb35bb-5767-8155-c804-14bd48789dc22",  //ID
     "userData": {name:eth0,speed:100tx,vlan:1},    //Settings
     "name": "eth0",
                  //Name
   ]},{
  "type": "draw2d.Connection",                      //Link
  "id": "69f7e40c-4586-0d7f-8817-c798f8c07969",
  "userData": {Type:OVS, VLAN:1, Filter:none, Duplex:yes}, //Link settings
  "source": {
    "node": "ebfb35bb-5767-8155-c804-14bd48789dc21"},// Source
  "target": {
    "node": "ebfb35bb-5767-8155-c804-14bd48789dc22"}}//Destination
```

Fig. 2. The JSON description of network topology

The second problem is to create a script for network deployment based on the scheme description. XML RPC is supported in most platforms, and it is used for interaction with OpenNebula. Python extension OCA is used on the server side for the scheme editor and ScriptMaker.

The script to make the infrastructure uses one.template.instantiate() method for creating virtual machines. The information about the parameters of network cards, their number and settings is written in the command of creation. All the virtual machines are created in the “Hold” state. The size of memory and network card parameters can be changed after the virtual machine has been created.

The routing module is used to establish the links between objects based on the object distribution on servers and the network scheme. This module sets up the OpenFlow rules for switches and OVS to provide the topology function. The rules are sets at follows:

<IN_PORT = value> <additional filters> <rewrite rule> <forward port >.

This way to set rules is called pro-active (it implies pre-routed traffic). Each switch and OVS gets its set of rules to distribute and isolate traffic flows. There is no high bandwidth requirements preset, the OpenFlow controller may be used to route all the flows through the rule:

< IN_PORT = value ><forward CONTROLLER>,

which allows dynamic processing of all flows on the controller.

Ryu SDN framework has been chosen as the platform for OpenFlow, due to possibility to configure tables through REST.

III. IMPLEMENTATION DESCRIPTION

For implementation of our approach, the cloud system based on OpenNebula has been created (Fig. 3). All the servers have two network cards: one for the service network and the other for the cloud network. The latter is linked to the OpenFlow switch and by default runs in the regular switch mode with dynamic VLAN.

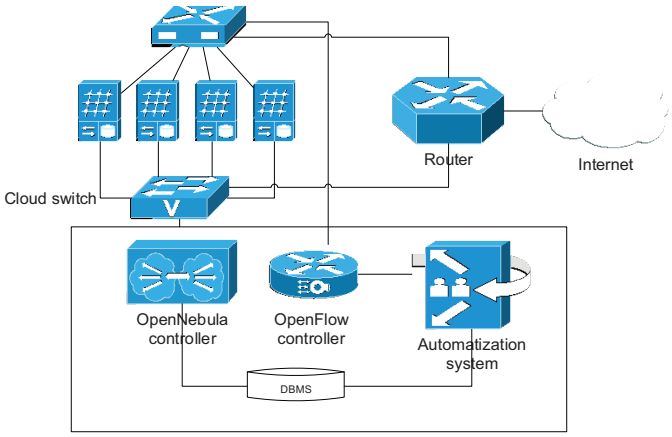


Fig. 3. The scheme of the cloud system

Each server has a virtual switch Open vSwitch (ovs-br0) for connection to virtual machines (Fig. 4). The switch functions as the real one with OpenFlow routing.

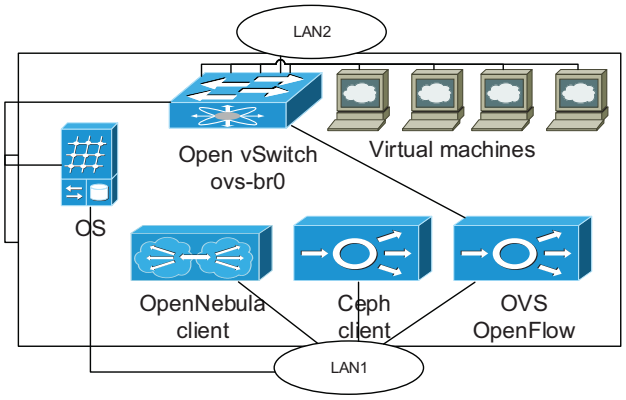


Fig. 4. The JSON description of network topology

From the point of view of controller, each route is a sequence of input and output ports (Fig.5). For routing the Dijkstra algorithm is used on the information, which controller receives through the LLDP protocol.

The network starts after the virtual machines have been deployed and routing rules have been appointed. For access to virtual machine consoles the VNC protocol is used (by default for KVM) through a browser with NoVNC extension.

When viewing the network scheme a double click on a VM opens the window with the console for access to the VM. The access through NoVNC is granted on port 443. Nginx runs on it and forwards traffic to NoVNC. The further access is performed by unique tokens from OpenNebula database, which allows to register real IP addresses and server ports in NoVNC. Each user has access through his tokens only.

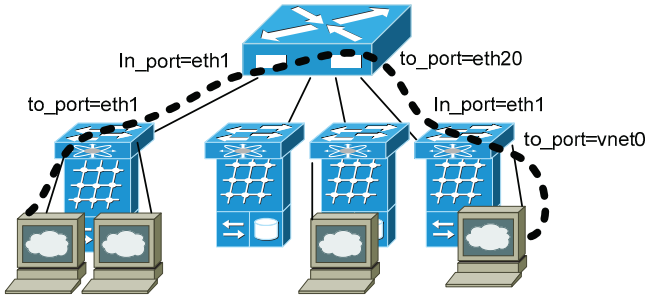


Fig. 5. The route through the network

IV. Access Control System

As virtual laboratory infrastructure is to be used for educational purposes, it needs a hierarchy of accounts with distributed access permissions to scheme components.

LDAP has been chosen for account information storage due to its common use. Besides LDAP is integrated with OpenNebula, so there is no need for additional authorization modules. The hierarchy of accounts has the following structure «Administrator» - «Teacher» - «User». Administrators can create or delete any schemes, resources, accounts and virtual machines within their control competence. Teachers can create or delete schemes, accounts, and resources (including virtual machine templates) within their control groups. Users can load, run and stop schemes within quotas granted to them. For automatized start, it is planned to use the class timetable as well as manual start.

V. Experiment Description

For our experimental research, we used the university educational cloud data center. The center includes four two-processor servers and eight one-processor servers with 32GB RAM each. Both the service network and the cloud network are implemented by HP 3500 switches supporting OpenFlow. The cloud is hypervised by OpenNebula, all the servers have the Open vSwitch bridge with OpenFlow for interconnection between the network and virtual machines. The Ceph four-server distributed storage is used. Each server has three 1Tb disks with RAM cashing. In addition, KVM save cashing is used for better speed. All the networks function at 1Gb/s, the links between switches provide 10Gb/s.

The first experiment was carried out to test storage performance for different configurations, determining the optimal choice. The virtual machines were started in automatic mode, starting times and Ceph storage latency were measured simultaneously. The start was performed with "onevm resume" command, the running time was registered with the monitoring script based on "onevm list" command. Fig.6 shows the experiment results for 1 through 50 virtual machines.

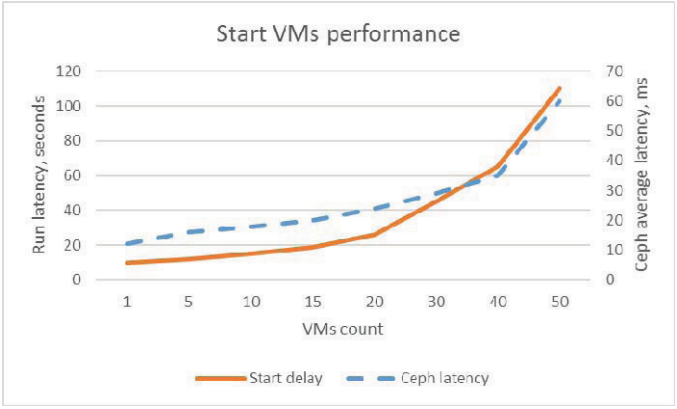


Fig. 6. The start VMs performance

As can be seen in Fig.6 run latency grows exponentially as VMs count grows. It is related to disk storage latency growing. The cloud implies 8 disks on 4 servers that can provide theoretical read performance up to 800Mb/s.

Fig.7 shows the storage throughput at different numbers of read/write processes.

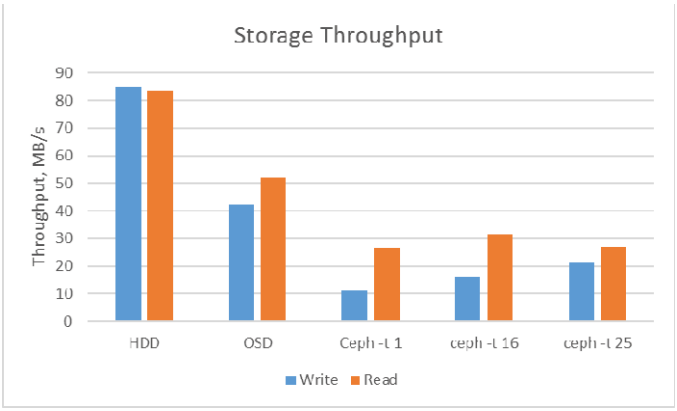


Fig. 7. The storage throughput.

As can be seen the storage throughput is significantly depends on access method. The experiment embraced only one client with several flows. If the number of clients is growing, the read performance grows proportionally, but the write performance remains the same (Fig.8).

Fig. 8. The multi-client throughput.

This explained by the distributed read method and the logging write method. The use of SSD for cashing and logging may provide a several times better performance.

VI. Conclusion

The experiment has shown the efficiency of using cloud systems in IT education. Even the small cloud datacenter (12 low-performance servers) has provided virtual classes for studying operating systems, network administration, cloud services and creating various cloud services and networks.

We are planning to carry on further research using a high-speed storage network with a 10G switch and SSD for cashing.

References

[1] Chowdhury, N.M.M.K.; Boutaba, R. Network virtualization: state of the art and research challenge //IEEE Communications Magazine, 2009– vol. 47– n.7– PP. 20-26.

[2] Draw2D touch URL: http://www.draw2d.org/draw2d/Konnov A.L., Legashev L.V., Polezhaev P.N., Shukhman A.E. Concept of Cloud Educational Resource Datacenters for Remote Access to Software // Proceedings of 11th International Conference on Remote Engineering and Virtual Instrumentation (REV), Polytechnic of Porto (ISEP) in Porto, Portugal from 26-28 February 2014 – 2014 – pp-246-247.

[3] UNetLan Official Site. Documentation http://unetlab.com/docs/

[4] Polezhaev P., Shukhman A., Ushakov Yu. Network Resource Control System for HPC based on SDN // Proceedings of 14th International Conference, NEW2AN 2014 and 7th Conference ruSMART 2014, St. Petersburg, Russia. Lecture Notes in Computer Science. – vol. 8638. – PP. 219-230.

[5] Polezhaev P., Shukhman A., Konnov A. Development of educational resource datacenters based on software defined networks // Proceedings of 2014 International Science and Technology Conference “Modern Networking Technologies (MoNeTec)”, Moscow, Russia, 2014. – PP. 133-139.

[6] Stabler G., Rosen A., Goasguen S., Kuang-Ching Wang Elastic IP and security groups implementation using OpenFlow. // VTDC '12 Proceedings of the 6th international workshop on Virtualization Technologies in Distributed Computing Date - PP. 53-60.

[7] Gorja P., Kurapati R. Extending open vSwitch to L4-L7 service aware OpenFlow switch. // 2014 IEEE International Advance Computing Conference (IACC) - PP. 343-347.

Informational and Control System Configuration Generation Problem with Load-Balancing Optimization

E.V. Melnik
Informational Technologies and Control Processes department
Southern Scientific Center
of the Russian Academy of Sciences (SSC RAS)
Rostov-on-Don, Russia

A.B. Klimenko
Scientific Research Institute of multiprocessor computing systems
Southern Federal University
Taganrog, Russia
Anna_klimenko@mail.ru

***Abstract*— This paper is devoted to the informational and control systems (ICSs) configuration generation problem. A set of configurations for the ICSs with performance redundancy is developed on system design stage. The problem is multicriteria with respect to key ICS design issues, including graceful system degradation, tasks relocation, etc. Within this paper the load-balancing criterion is added due to its influence onto the gamma-percentile operating time to failure. Then the solutions of the resulting multiobjective problem are evaluated and compared, and some conclusions are made.**

***Index Terms*— Informational and control systems, hazardous environments, autonomous objects, load balancing, configuration generation.**

I. Introduction

Contemporary informational and control systems (ICSs) are used in a wide range of a mechatronical objects, which operate in complex hazardous environments. For instance, aircrafts operate in a lack of time conditions: latency in monitoring and control can lead to system failure and casualties. Another example of "unfriendly" hazardous environment is the open space. Here the long terms of maintenance absence is actual. So, ICSs for such objects must be designed and developed with application of high reliability requrements. One aspect of the reliability is the possibility of recovery, in particular, through reconfiguration. Reconfiguration is possible when there is a redundancy in the system: for instance, until nowadays, the main type of redundancy in aircrafts/spacecrafts is structural. Almost all of modern aircraft such as Boeing 777 and Airbus A320/330/340 have used triplex- or quadriplex-redundant activation systems, flight control computer and databus systems [1-3]. As to the performance redundancy[4-6], it must be mentioned, that such a way to design the reliable ICSs has multiple advantages[4-6], but needs more complex recovery procedure. In the scope of this paper reconfiguration of the ICS is the procedure which relocates the control and monitoring tasks (CMTs) from the faulted computational nodes to the operational ones. In the lack of time conditions it can be done by choosing some predefined configurations; in case of no configuration available some heuristic procedures can be applied [7,8]. Here it must be mentioned, that the distributed dispatching of the ICS is also important, due to multiple drawbacks of the centralized architecture[8]. But, resuming, in general case the set of ICS configurations must be prepared and tested on the design stage, and the chosen objective function is crucial.

One of the configuration generation issues is the load balancing parameter. Latest studies show that it is important [4,6] and influences onto the gamma-percentile operating time to failure (up to 30%). Such influence can be crucial for the mechatronical objects, which operates without maintenance possibilities (spacecrafts, autonomous robots, etc.). The question is how to incorporate the load balancing objective function into the configuration generation problem and what results can be achieved.

The outline of this paper includes the brief review of the reconfiguration within the performance redundancy using, a configuration generation problem formalizing via multiobjective function. Then the load-balancing criterion is added, and the solutions of such multicriteria optimization are evaluated and compared.

II. Performance Redundancy and Decentralized Monitoring and Control

The structural redundancy considers the set of basic (operational) elements and a set of reserve elements. When the node from the basic set fails, all control and monitoring tasks from this node can be relocated to the reserve node. Performance redundancy considers all computational units (CUs) as operating elements with some performance reserve.

In the case of CU failure (software or hardware) CMTs from the faulted node are launched by the operational ones (fig.1). Advantages of performance redundancy are explained in details [4-6,9]. Briefly, they are: resource utilization improving, gamma-percentile operating time to failure increasing, and, for particular configurations, the number of

possible recoveries increases. With the performance redundancy the decentralized ICS dispatching increases the reliability and scalability of the system. Any kind of central control element can become a bottleneck of the system, so the failure of central control element leads to the system failure without possibility to recover. When the ICS dispatching subsystem is decentralized, the system consists of equal control elements, and in case of failure the operational nodes make the collective decision about further operating. As shown in fig.1, each CU is represented by software agent, which is launched via CMT task. Agents are aware of each other, and when one agent does not broadcast the sign of its "presence" into communicational network (CN), other agents must begin the recovery procedure (the question of malfunctioning agents notification is out of scope of this paper).

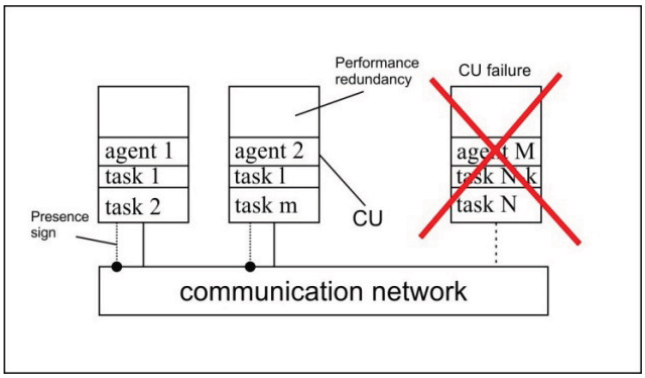


Fig. 1. Performance redundancy and decentralized dispatching.

In practice, possible configurations are designed and tested on the design stage of the ICS creation. So, the problem of configuration generation must be solved and evaluated

III. Problem Formalization

Let the input data be the following:

- A set of CMTs $G=\{x_i\}$, $i=1...N$, where x_i – the size of task i, N – the number of tasks.
- Let $G = G_c \cup G_{nc}$, $G_c \cap G_{nc} = \varnothing$, where G_c – a subset of critical MCTs, G_{nc} – a subset of non-critical ones. Non-critical CMTs can be stopped or eliminated from the system during reconfiguration. The number of critical CMTs is N_c, the number of non-critical CMTs N_{nc};
- Let G_f be the set of CMTs from the faulted CU. $G_f \subseteq G$, G_p is the performing tasks, $G_p \subseteq G$, $G_p \cap G_f = \varnothing$.
- A planned completion time for the set G is T_{plan}.
- Number of CUs is M with the equal performance p.

Let' s take into consideration that we have to allocate the CMTs from the set G_f within the system of operational CUs, on which the tasks from the set G_p are allocated with the constraint of completion time *Tplan*. Let the resource allocated by CU j for the subtask i be λ_{ij}. The tasks allocation before the failure is described by matrix R:

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{1M} \\ \dots & \dots & \dots \\ r_{N1} & \dots & r_{NM} \end{vmatrix} \quad (1)$$

$$\text{where } r_{ij} = f\left(\frac{x_i}{\lambda_{ij} p}\right),$$

$$f\left(\frac{x_i}{\lambda_{ij} p}\right) = \begin{cases} \dfrac{x_i}{\lambda_{ij} p}, & \text{if } x_i \text{ is running on CU } j, \\ 0, & \text{otherwise.} \end{cases}$$

Let the failure occurred on the CU with number d. The column d of matrix R is deleted, so there are $M-1$ columns and $N-|G_f|$ lines in the new matrix R_f. Renumber the elements of R_f in the following way, saving the indexes from the matrix R in the upper positions:

$$R_f = \begin{vmatrix} r_{11}^{ij} & r_{12}^{ij} & r_{1(M-1)}^{iM} \\ \dots & \dots & \dots \\ r_{11}^{Nj} & \dots & r_{(N-G_f)(M-1)}^{NM} \end{vmatrix} \quad (2)$$

R_f describes the system state before the reconfiguration and contains the allocation of the operational tasks among the operational CUs. R_r will be the allocation of the task set G on the $M-1$ CUs. Formally, the subset G_f will be added to the G_p with the number of CUs = $M-1$:

$$R_r = \begin{vmatrix} r_{11}^{ij} & r_{12}^{ij} & r_{1(M-1)}^{iM} \\ \dots & \dots & \dots \\ r_{11}^{Nj} & \dots & r_{(N)(M-1)}^{NM} \end{vmatrix}; \quad (3)$$

$$r_{ij}^{kl} = f\left(\frac{x_i}{\lambda_{ij} p}\right) g(x_i)$$

$$g(x_i) = \begin{cases} 0, \text{if } x_i \in G_{nc} \text{ and eliminated from the system,} \\ 1, \text{otherwise.} \end{cases} \quad (4)$$

Let's consider matrix Φ.

$$\Phi = \left[\varphi(x_{1j}^{kl}) \quad \varphi(x_{2j}^{kl}) \quad \ldots \quad \varphi(x_{Nj}^{kl})\right] \tag{5}$$

where

$$\varphi(x_{1j}^{kl}) = \begin{cases} 0, & l = j, \\ \xi, & otherwise \end{cases}.$$

k,l – the saved indexes of matrix R, j – the number of CU in matrix Rr, ξ – the integer number.

The matrix Φ describes wether the CMT xi was relocated from CU $_{l}$ to CU $_{j}$.

The first objective function can be written in the following manner:

$$F_1 = \sum_{i=1}^{N} \varphi(x_{ij}^{kl}) \to MIN \tag{6}$$

The maximum number of non-critical tasks running equals the maximum summa of all g(xi) in the matrix R'.

The next objective function component can be presented as following:

$$F_2 = -\sum_{i}^{N} g(x_i) \to MIN \tag{7}$$

Herewith the time constraint must be satisfied:

$$\forall j : \sum_{i=1}^{N} r'_{ij} \le T_{plan}, j \in [1...M] \tag{8}$$

Let's put the current multicriteria optimization problem to the following form:

$$F = \sum_{i=1}^{N} \varphi(x_{ij}^{kl}) \to MIN \tag{10}$$

$$-\sum_{i}^{N} g(x_i) \le \mu; \quad \forall j : \sum_{i=1}^{N} r'_{ij} \le T_{plan}, j \in [1...M]$$

$x_i > 0;$, where μ is the integer number.

Here, in expression (10), the two-criteria configuration generation problem is formalized. In other words, we need to find CMTs allocation with minimum relocations with respect to preferable graceful system degradation (keeping non-critidal tasks running if it is possible).

As to the load balancing criteria, we need to add one more objective function to our multiobjective function. It can be presented in the following way.

$$F_3 = MAX(\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il}) \to MIN, \forall k,l, \lambda_{ik} \subset R_r \tag{11}$$

With expression (11) multicriteria configuration generation problem can be written as follows:

$$F = \sum_{i=1}^{N} \varphi(x_{ij}^{kl}) \to MIN \tag{12}$$

$$MAX(\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il}) \le \gamma; \; -\sum_{i}^{N} g(x_i) \le \mu; \quad \forall j : \sum_{i=1}^{N} r'_{ij} \le T_{plan}, \; j \in [1...M]$$

$x_i > 0; \; 0 < \lambda_{ij} < 1$, where $0 < \gamma < 1$ is the assumed level of load dispersing, μ is the constraint for the eliminated non-critical tasks.

IV. Simulation Results and Discussion

Within the current research, two types of simulation have been made. The first simulation estimates the solution quality of the problem (12) in dependent on constraint values. The simulated annealing with the quenching temperature scheme has been used due to its fast convergence. The second simulation allows to estimate the influence of the load-balancing component in the multicriteria objective function and to compare it with the objective function solutions(10).

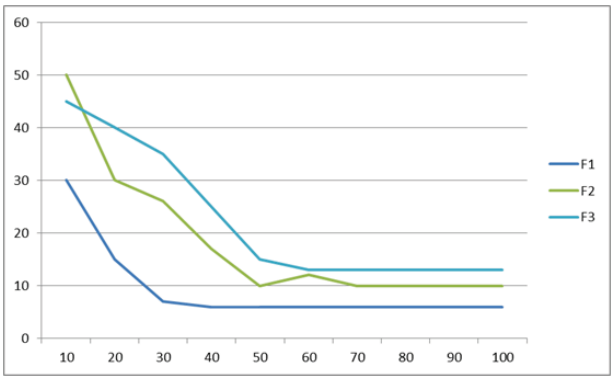


Fig. 2. F1 – $\gamma = 0{,}5$; F2 – $\gamma = 0{,}3$; F3 – $\gamma = 0{,}2$. $\mu = 10$.

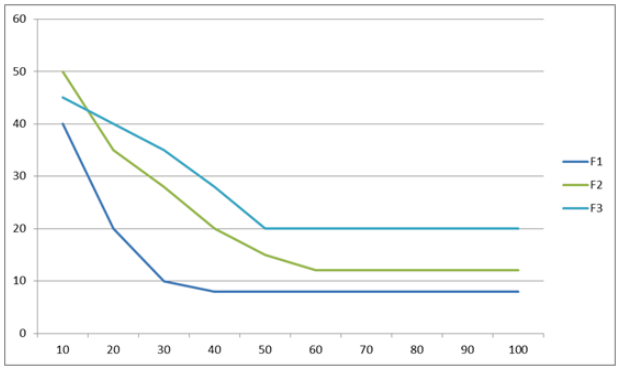


Fig. 3. F1 - $\gamma = 0{,}5$; F2 - $\gamma = 0{,}3$; F3 - $\gamma = 0{,}2$. $\mu = 5$.

It is seen (fig.2, fig.3) that with constraint strengthening the solution quality becomes worse. Nevetherless, the results of simulation allow to concern the simulated annealing as prospective from the time of solution obtaining point of view. The local minima can be reached within the 20-60 object function calls. It must be mentioned here, that in[10] one-criteria objective function was explored, and the local minimas were obtained within 20 object function calls.

The next simulation allows to evaluate the influence of the load balancing component to the solutions obtained. The number of CUs is 6, the number of CMTs with random size is10. We assumed that the performance of the CUs is big enough (in other words, the performance redundancy is in stock). On the fig.4 the task were relocated without load balancing, and some loading “pikes” are seen. On the fig.5, with load balancing component, the “pikes” are much more smoother.

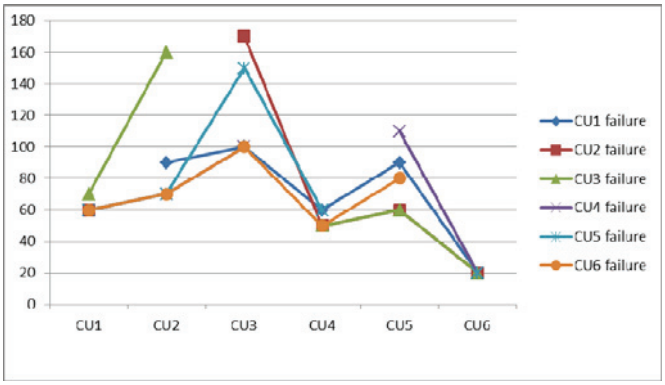


Fig. 4. Configuration generation problem solution without load-balancing component

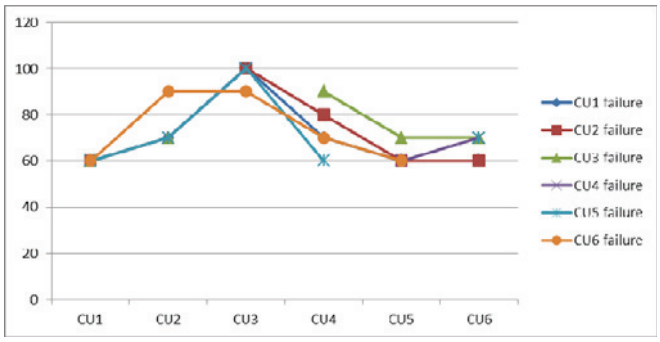


Fig. 5. Configuration generation problem solution without load-balancing component

CU load dispersing with the load-balancing optimization has the decreasing tendency. Assuming results of [6], the increasing of gamma-percentile operating time to failure is up to 10% depending on faulted node.

V. Conclusions

Performance redundancy and decentralized dispatching increase the ICS reliability and recovery potential. On the other hand, they make the recovery procedure more complex in comparison with the structural redundancy and requires new recovery methods and techniques for design and implementation. In general, the problem of configuration generation must be solved on the system design stage, and estimated and tested further. Besides, the load balancing an important aspect of the problem due to its influence to the gamma-percentile time to failure.

Within this paper the problem was formalized as multicriteria one, and the load balancing criteria was incorporated. The simulatioin results allow to estimate the influence of the load balancing component to the solutions with respect to the gamma-percentile time to failure.

Resuming, the main result of this research is that the load-balancing optimization as one of the multicritreria objective function component allows to obtain solutions with acceptable load dispersing and thus improve the gamma-percentile time to failure.

Acknowledgment

The reported study was funded by SSC RAS projects 0256-2014-0008, 0256-2015-0082 within the task 007-01114-16 PR and by RFBR project 14-08-00776-a.

References

[1] Y. Zhang, J.Jiang. Bibliographical review on reconfigurable fault-tolerant control systems. Annual Reviews in Control, 32(2):229–252, 2008.

[2] G.F. Bartley. Boeing B-777:fly-by-wire flight controls. In C.R. Spitzer, editor, The Avionics Handbook. CRC Press, Boca Raton, FA,2001.

[3] D.Biere, C.Favre, P. Traverse. Electrical flight controls, from Airbus A320/330/340 to future military transport aircraft: A

family of fault-tolerant systems. In C.R. Spitzer, editor, The Avionic Handbook. CRC Press, Boca Raton, 2001.

[4] E.V. Melnik, G.V.Gorelova. Simulation options for redundancy in distributed information and control systems with a decentralized organization. Proceedings of SFU. Ser. Technical science. №3.pp.184-193. 2013.

[5] E.V. Melnik, D.Ya.Ivanov. Principles of organization of the decentralized network-centric information management systems. Herald of computer and information technologies. №4.pp.25-30. 2013.

[6] E.V. Melnik, G.V.Gorelova. Effect processor computational load balancing devices in highly distributed information management system. Mechatronics, Automation, Control. pp. 29-35. 2012

[7] E.V. Melnik, V.V.Korobkin, A.B.Klimenko. Fault-tolerant architecture for the hazardous object information control systems. Application of information and communication technologies – AICT2015 (IEEE catalog number CFPI556H-PRT): conference proceedings (Rostov-on-Don, Russia 14-16 October 2015). – Rostov-on-Don: SFedU. – 2015. – P.274-276.

[8] E.V. Melnik, A. B. Klimenko. Experimental Study of Simulated Annealing Algorithm with Rapid Decrease in Temperature to Form a Distributed Configuration Information and Management System. Journal of Computer and Information Technology. № 3 (129). pp.16-21. 2015.

[9] I.A. Kalyaev, E.V. Melnik. Decentralized systems of computer control. Rostov n / D: Publishing SSC RAS. 196p. 2011.

[10] E. Melnik, A. Klimenko, V.Klimenko. The parallel simulated annealing-based reconfiguration algorithm for the real time distributed control fault-tolerance providing. Application of information and communication technologies – AICT2015 (IEEE catalog number CFPI556H-PRT): conference proceedings (Rostov-on-Don, Russia 14-16 October 2015). Rostov-on-Don: SFedU.– pp.277-280. 2015.

A Better Heuristic for the Minimum Connected Dominating Set in Ad Hoc Networks

Onur Ugurlu
Department of Mathematics
Ege University, Izmir, Turkey
onurugurlu@mail.ege.edu.tr

Deniz Tanir
Department of Mathematics
Ege University, Izmir, Turkey
tanirdeniz35@gmail.com

Elnur Nuri
Department of Mathematics
Ege University, Izmir, Turkey
nurielnur@gmail.com

***Abstract*— Since no fixed infrastructure and no centralized management present in Wireless Ad Hoc Networks (WANETs), a Connected Dominating Set (CDS) representing the network is widely used as a virtual backbone. Given a graph, a CDS is a subset of vertices such that every vertex in the graph is either in the subset or adjacent to a vertex in the subset and the subgraph induced by the subset is connected. A smaller virtual backbone (a smaller size CDS) incurs less communication overhead. However, finding a minimum size CDS is NP-hard. Thus, it is important to design effective algorithms for the minimum CDS (MCDS) problem. In this article, a new efficient heuristic name as 2-Lenght Betweenness Heuristic for the MCDS problem is proposed. Comprehensive simulation results demonstrate that the proposed heuristic algorithm finds better solutions than the existing approach.**

***Index Terms*— Wireless Ad Hoc Networks, Minimum Connected Dominating Set Problem, Heuristic Algorithms.**

I. INTRODUCTION

WANETs have many real life applications in many different areas such as the health-care industry, food industry, military application [1], etc. Unlike wired networks, WANETs have no physical infrastructure. Thus, generating routing protocols is to one of the essential problems in WANETs. A virtual backbone increases the bandwidth efficiency, reduces the total energy consumption and the communication overhead [2] in WANETs. A MCDS can be optimal as a virtual backbone in such networks [3].

A dominating set (DS) D for $G=(V,E)$ is a subset of V such that every vertex in not in D is adjacent to at least one vertex in D. If the D is connected, then D is a CDS. The MCDS problem consists of finding the smallest CDS in a graph. The problem belongs to NP-hard [4].

WANETs can be modeled as an undirected graph $G = (V,E)$, in which V is the set of mobile hosts which have transmission ranges, and E represents the set of edges where an edge $(u,v) \in E$ if and only if u and v are within their transmission ranges. As a result, the network can be modeled as a unit-disk graph (UDG), where all the vertices have same transmission range [5].

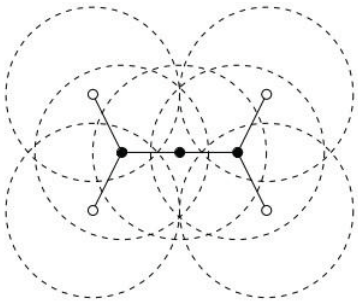


Fig. 1. MCDS on a unit disk graph.

It is quite obvious that the efficiency of virtual backbone technique depends primarily on finding a minimal size CDS. But, finding a minimum size CDS in a UDG is also NP-hard [6]. Therefore, it is important to design effective heuristic and approximate algorithms for the MCDS problem for real life applications. There are several approximation algorithms [7,8,9,10] and heuristic algorithms [11,12,13,14] have been proposed for the MCDS problem in the literature. Most of these approaches can be divided into two groups; Maximum Independent Set (MIS)-based algorithms and Tree-based algorithms. In MIS-based MCDS algorithms [7,8,9] a maximal independent set of the given graph is found at the first step. Then, a local search approach is implemented to vertices (connectors) to connect the vertices in the maximal independent set and construct a CDS. Tree-based CDS algorithms [11,12,13] focus on building a CDS by expanding a small trivial CDS. This type of algorithms begin with a subset of vertices called initiators and expands the CDS tree from each of the initiators. A comprehensive survey on the construction algorithms of MCDS problem was presented in [15]. Almost all of the currently existing algorithms for MCDS start with an empty solution and try to construct the CDS by adding vertices at each iteration according to some certain criteria. The only exception is the algorithm was proposed by Butenko et al. [14]. However, addition-based algorithms may

have some disadvantages such as more energy consumption and additional set up time.

In this paper, a new subtraction-based heuristic algorithm which we name as 2-Length Betweenness Heuristic (2LBH) is proposed for the minimum CDS problem. We compare our algorithm with the existing subtraction-based (BCOP) algorithm [14]. The experimental results indicate that 2LBH can find better solutions than BCOP. Also, we share our test instances which are used in experiments [16]. Thus, researchers can use the instances to compare the effectiveness of their algorithms. This article is structured as follows. 2LBH algorithm is described in the following section. In Section III, 2LBH algorithm is compared with BCOP algorithm and experimental results are given. In the last section, we conclude the paper.

II. The Proposed Algorithm

A. Preliminary Knowledge

The neighborhood of a vertex: For each $v_i \in V$, the neighborhood of v is de£ned by $N(v_i) = \{u \in V / v_i$ is adjacent to $v_i\}$.

The degree of a vertex: The degree of a vertex $v_i \in V$, denoted by $d(v_i)$ and is defined by the number of neighbors of v_i.

Betweenness Centrality [17]: Let σ_{st} denote the number of shortest paths from $s \in V$ to $t \in V$ and $\sigma_{st}(v_i)$ denote the number of shortest paths from s to t that some $v_i \in V$ lies on. Betweenness Centrality of v_i is as follows:

$$C_B(v_i) = \sum_{s \neq v_i \neq t \in V} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \qquad (1)$$

2-Length Betweenness: Let p_{st} denote the all 2 length path between $s \in V$ and $t \in V$. if $v_i \in V$ *lies on a* p_{st} then $p_{st}(v_i) = 1$, otherwise $p_{st}(v_i) = 0$; denote the number of shortest paths from s to t that some $v_i \in V$ lies on. 2-Length Betweenness of v_i is as follows:

$$2LC_B(v_i) = \sum_{s \neq v_i \neq t \in V} \rho_{st}(v_i) \qquad (2)$$

B. 2LBH Algorithm

As we already mentioned in the previous section, MCDS algorithms mostly start with an empty solution, then they try to find vertices to be a part of CDS mostly according to a greedy criterion and adding the vertices to the solution. However, BCOP algorithm uses the inverse method; it begins with a feasible solution which contains every vertex in the graph. At each step, BCOP tries to remove a vertex from CDS, which does not make CDS an unfeasible solution. To choose which vertex to remove, the algorithm uses the degree of the vertices as greedy criteria. At each iteration, the algorithm chose the vertex of smallest degree which is proper to be removed. This procedure continues until there is no removable vertex.

The proposed algorithm starts with a solution CDS which contains every vertex in the graph, and recursively removes vertices from this solution until there is no removable vertex in the graph like BCOP. However, unlike BCOP, we use 2-Length Betweenness as greedy criteria to find vertices which will be removed from CDS. Moreover, we use the static greedy technique to prevent information loss in the network while selecting vertices. The proposed 2LBH is as follows:

Input: A graph with *n* vertices.
Output: *CDS*.

1. $CDS \leftarrow V$
2. **for** i=1 **to** n
   Compute the 2-Lenght Betweenness $B[i]$ for all v_i
   end for
3. Sorts the vertices in descending order according to B
4. **for** i=1 **to** n
   if $CDS \backslash \{v_i\}$ is a feasible solution **then**
   $CDS \leftarrow CDS \backslash \{v_i\}$
   end if
   end for

The biggest advantage of the subtracting algorithms is to maintain a feasible solution at any iteration of its execution; in this way, there will not be any need for setup time requirements. Also, the proposed algorithm simple to implement.

C. Computational Complexity

The time complexity of the algorithm 2LBH can be obtained as follows: In step 2, calculating all the 2-Length Betweenness for all vertices based on the neighbors information requires at most $O(n^3)$. Sorting array B with quicksort require $O(nlogn)$ time in step 3. In final, we must check if removing a vertex disconnects make the solution unfeasible n times. We use bread first search algorithm for this procedure. Thus, step 4 requires $O(n.(m+n))$ time. Consequently, the total time complexity of 2LBH can be given by $O(n^3)$.

III. Experimental Results

In order to evaluate the performance of 2LBH for the MCDS problem, comprehensive computational experiments were carried out both on random and deterministic graphs. 2LBH was implemented by the authors in C++ and compiler with gcc version 4.9.2 with –o3 optimization. The experiments were carried out on Intel Core i7-4700HQ processor with 32GB of RAM. The reading of the graph and creating the adjacency matrix were not included in running times since they are common to all algorithms. To benchmark 2LBH, the algorithm was compared with existing subtraction algorithm BCOP. Random graphs were generated in square units of a 2-D simulation area, and a random number of vertices *(n)* are placed on this area. We generated random graphs in a 100x100 square units. We varied the number of mobile hosts (*n*) from 10 to 500. In the network, two vertices are connected through an edge, only if the distance between them is equals or less than the radius of the mobile host's transmission area (*r*), and all vertices which are in the network have the same *r*. This state

that the network is a unit-disk graph. For each n, we set r to three different values: 25,50,75. It is clear that when r increases, the density of the graphs increases also. The density (d) of a graph measures the ratio of how many edges the generated graph contains to the maximum number of edges a graph which has the same number of vertices with generated graph. For each n and r, we created 5 random graphs. Recall that only connected graphs were taken into consideration for the experimentation. The test instances can be downloaded via [16].

TABLE I. RESULTS OF COMPUTATIONAL EXPERIMENTS ON RANDOM UNIT DISK GRAPHS

n	*R*	*d*	*BCOP*	*2LBH*	T_{BCOP}	T_{2LBH}
10	25	0,30	5,2	5,2	1,00	1,00
	50	0,54	2,2	2,2	1,00	1,00
	75	0,76	1,4	**1,2**	1,00	1,00
25	25	0,16	10,8	10,8	1,00	1,00
	50	0,49	3	**2,8**	4,80	2,80
	75	0,75	2	**1,4**	2,80	1,00
50	25	0,15	11,8	11,8	8,20	1,00
	50	0,47	3,8	**3,4**	1,40	10,00
	75	0,83	1,6	**1**	12,00	16,00
75	25	0,16	**11,6**	12	30,00	10,00
	50	0,47	3,6	3,6	46,00	34,00
	75	0,80	1,4	**1**	76,00	52,00
100	25	0,16	13,2	**12**	68,00	1,00
	50	0,36	3,6	3,6	106,20	92,40
	75	0,81	2	**1**	92,40	126,20
125	25	0,14	12,8	**12,6**	94,40	1,00
	50	0,51	3,6	3,6	236,00	156,00
	75	0,51	1,6	**1**	194,00	208,00
150	25	0,15	12,6	12,6	200,00	42,40
	50	0,47	3,8	3,8	326,00	186,00
	75	0,79	1,8	**1**	416,00	312,00
200	25	0,16	13	**12,6**	574,00	156,00
	50	0,49	4	4	868,00	576,00
	75	0,80	1,8	**1**	936,00	926,00
250	25	0,16	**12,6**	13,2	976,00	25,00
	50	0,50	3,8	3,8	1750,00	109,00
	75	0,81	1,8	**1**	1874,00	1686,00
500	25	0,16	**12,8**	13,2	8220,00	2112,00
	50	0,48	4	4	13500,00	8312,00
	75	0,80	2	**1**	15776,00	14260,00

Besides testing algorithms on completely random graphs, we also investigated the performance of these algorithms on more deterministic graphs since deterministic vertex deployment has many advantages. Thus, we have used BHOSLIB [18] and DIMACS instances [19]. BHOSLIB instances are designed for the MIS and the minimum vertex cover problem. DIMACS instances are designed for the maximum clique problem, so we have used the complement of the graph because of the relation between MIS and minimum CDS. In Table I, the results of comparison on the random unit disk graphs are given. The results are the average value of 5 instances for each n and r. In Table II, the results on BHOSLIB and DIMACS instances are given. The results are the average value of all instances with the same number of vertices in a graph family. In Table I, the average computational times are given in milliseconds, whereas in Table II, they are given in seconds.

TABLE II. RESULTS OF COMPUTATIONAL EXPERIMENTS ON BHOSLIB AND DIMACS INSTANCES

Graph Name	*n*	*d*	*BCOP*	*2LBG*	T_{BCOP}	T_{2LBG}
brock200	200	0,37	7,75	**7,00**	0,07	0,05
brock400	400	0,25	13,50	**11,75**	0,48	0,21
brock800	800	0,35	12,25	**10,25**	4,61	2,50
c-fat200	200	0,78	2,00	2,00	0,08	0,06
c-fat500	500	0,90	2,00	2,00	1,14	0,97
p_hat300	300	0,51	6,67	**5,67**	0,30	0,19
p_hat500	500	0,50	6,67	**6,33**	1,37	0,92
p_hat700	700	0,50	7,33	**7,00**	3,73	2,58
p_hat1000	1000	0,51	7,67	**7,33**	10,91	7,56
p_hat1500	1500	0,50	9,67	**8,33**	36,98	27,91
san200	200	0,18	18,20	**15,20**	0,05	0,02
san400	400	0,22	14,20	**11,60**	0,52	0,23
sanr200	200	0,20	16,00	**14,00**	0,06	0,02
sanr400	400	0,40	9,50	**9,00**	0,63	0,38
frb30-15	450	0,18	24	**17,8**	0,54	0,16
frb35-17	595	0,16	30,8	**20,4**	1,21	0,31
frb40-19	760	0,14	37	**27,4**	2,40	0,58
frb45-21	945	0,13	37,2	**29,6**	4,66	0,97
frb50-23	1150	0,12	43,4	**31**	8,48	1,82
frb53-24	1272	0,11	47,2	**35**	10,53	2,07
frb56-25	1400	0,11	49,8	**34,6**	14,18	2,78
frb59-26	1534	0,11	53,4	**36,6**	18,62	3,61

The simulation shows the performance of 2LBH is better compared to BCOP. In random UDG, 2LBH found 15 better solutions than BCOP in different graphs. However, BCOP found only 3 better solutions than 2LBH. It can be said that BCOP is more competitive with 2LBH on sparse UDG. When d is close to 0.5, both BCOP and 2LBH found similar results on random UDG. However, in dense UDG, 2LBH outperforms BCOP. The results of computation experiments on BHOSLIB and DIMACS instances suggest that 2LBH is superior to BCOP on more deterministic graphs. Although the time complexity of BCOP is smaller than 2LBH, 2LBH can find a solution faster than BCOP. Since the real life instances of WANETs are usually dense and complex structure, it can be said that 2LBH is a better algorithm for the MCDS problem.

IV. CONCLUSIONS

In this paper, a new subtraction based heuristic 2LBH for the MCDS problem is proposed. The efficiency of 2LBH is examined on random unit disk graphs and well-known benchmark instances. The comprehensive experimental results show that 2-Length Betweenness is a better greedy criterion than the degree of vertices for subtraction based heuristics for the MCDS problem. Also, we share our test instances which are used in experiments. Thus, researchers can use the instances to compare the performance of their algorithms.

ACKNOWLEDGMENT

The first and second authors gratefully acknowledge the support of TUBITAK 2211 program. Also, the third author is supported by Ege University Research Fund Project (Project No: 2016\Science\056).

REFERENCES

[1] M.K.A. Salhieh, J. Weinmann, L. Schwiebert., Power efficient topologies for wireless sensor networks., in: Proceedings of International Conference on Parallel Processing, IEEE, 2001, pp. 156–163.

[2] D. Z. Du and P. J. Wan,. Connected dominating set: theory and applications. Springer Science & Business, (2012).

[3] V. Bharghavan and B. Das, "Routing in Ad Hoc Networks Using Minimum Connected Dominating Sets", International Conference on Communications'97, Montreal, Canada. June 1997.

[4] M. R. Garey and D. S. Johnson, Computers and Intractability – A Guide to the Theory of NP-Completeness, W. H. Freeman, San Francisco CA,1979.

[5] J. Blum, M. Ding, A. Thaeler, X. Chen, Connected dominating set in sensor networks and MANETs, Handbook of combinatorial optimization. Springer US, 2004. 329-369.

[6] B. S. Baker, Approximation algorithms for NP-complete problems on planar graphs, Journal of the ACM (JACM), 41(1):153–180, 1994.

[7] Y. Li et al, On greedy construction of connected dominating sets in wireless network,. Wireless Communications and Mobile Computing, 2005, 5(8), 927-932.

[8] M. T. Thai, F. Wang, D. Liu, S. Zhu, and D. Z. Du, Connected dominating sets in wireless networks with different transmission ranges, IEEE Transactions on Mobile Computing, vol. 6, pp. 721-730, July 2007.

[9] B. Das, R. Sivakumar, and V. Bhargavan, Routing in Ad Hoc Networks Using a Spine, Proc. Sixth IEEE Int'l Conf. Computers Comm. and Networks (IC3N '97), pp. 1-20, Sept. 1997.

[10] S. Guha and S. Khuller, Approximation algorithms for connected dominating sets, Algorithmica, 20(4):374–387, 1998.

[11] M. Rai, S. Verma, S. Tapaswi, A power aware minimum connected dominating set for wireless sensor networks, Journal of networks 4.6 (2009): 511-519.

[12] J. Wu and H. Li, On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks, In Proceedings of the 3rd ACM International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, 1999, Pages 7-14.

[13] P.J. Wan, K. M. Alzoubi and O. Frieder, Distributed construction of connected dominating set in wireless networks, in: Proceedings of IEEE Conference on, Computer Communications (INFOCOM), April 2002, pp. 141–149.

[14] S. Butenko, X. Cheng, C.A. Oliveira, P.M. Pardalos, A new heuristic for the minimum connected dominating set problem on ad hoc wireless network,. In: Recent developments in cooperative control and optimization. Springer US, 2004. p. 61-73.

[15] J. Yu, N. Wang, G. Wang, D. Yu, Connected dominating sets in wireless ad hoc and sensor networks–A comprehensive survey, Computer Communications 36.2 (2013): 121-134.

[16] http://fen.ege.edu.tr/~urfat/CDS_instances/index.html

[17] L. C. Freeman, A set of measures of centrality based on betweenness. Sociometry, 1977, 35-41.

[18] http://www.nlsde.buaa.edu.cn/~kexu/benchmarks/graph-benchmarks.htm

[19] https://turing.cs.hbg.psu.edu/txn131/clique.html

Automation of Interaction of Customers and Suppliers in Control System of Providing Electronic Services to The Population within Network of Portals

Volkov V.N., Stichuk A.A., Lunev R.A., Demidov A.V.
Orel State University
Orel, Russia
vadimvolkov@list.ru, stichuck@yandex.ru, rolu@yandex.ru, a.demidov@ostu.ru

***Abstract*—The problems of a current state of the sphere of implementation of the electronic services to the population (ESP) are considered here. Need of creation of specialized organizational and technical support system - a producer of ESP which functions are directed on minimization of time of receiving necessary service by the applicant - is revealed and then proved. It is established that the most effective interaction of customers and suppliers of electronic services is realized by the means of the web services within network of portals including three components: identifying, billing and service.**

***Index Terms*—electronic services to the population, automation, control system, network of portals.**

I. Introduction

The social and economic development of regions which is one of priority problems of the state social policy of the Russian Federation is carried out, as a rule, on the basis of the regional social and economic programs consisting of several projects [1]. One of the priority directions of development of Russia is now creation of information society which major principle is the wide spreading and availability to the population of electronic services [2].

The analysis of a current state of this sphere opens a number of negative factors [3]:

- considerable expenses of temporary and work forces of the population when receiving services;
- absence of sufficient information in electronic form;
- customer's low possibility to control process of execution of service and result of its realization;
- insufficient competence of the service customer of a range of possible ways and ways of its receiving;
- repeated duplication of papers;
- low efficiency of distribution and direction of state intellectual resources;
- essential lag in use of an automation equipment and information technologies at realization of services to the population from the overall level of informatization;
- low degree of information security regarding observance of a trade secret.

The Electronic Services to the Population (ESP) are any kind of services which are rendered by the organizations of all forms of ownership from various areas of a national economy, and also individual entrepreneurs, to physical and juridical persons, individual entrepreneurs, citizens of the Russian Federation, foreign citizens and stateless persons with use of means, tools, and methods of information technologies [4]. However, the state and municipal services rendered in an electronic form when functioning the electronic governments, the electronic cities, regions, etc. are not the most part of a broad spectrum of ESP. Thus, ESP can differ considerably not only by branches of a national economy, but also by types of customers and suppliers, forms of the access organization, extent of intellectualization, automation of processes and participation of the customer at service realization, and also scales of possible realization (width of coverage of various segments of the population and territories), difficulties and degrees of implementation process hierarchy (number and degree of subordination of intermediate suppliers, need of formation and realization of compound services) [5].

The natural form of organization of the processes associated with ESP is a necessary level of security web portal [6]. The range of services that can be expressed in electronic form, is very wide, therefore, special measures are needed to integrate heterogeneous web portals of organizations that provide specific services [6-10].

All this defines need of designing of many ESP ordered by consumers as essentially new, not having approved regulations (unlike the state and municipal services entered into the obligatory list).

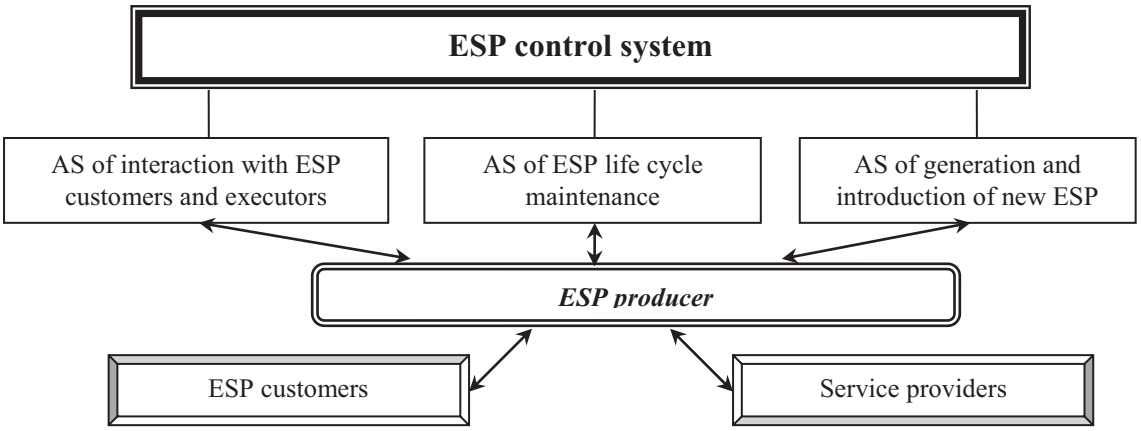


Fig. 1. Scheme of interaction of implementation ESP process participants

II. Control System of the Electronic Services Rendered to the Population

For ensuring effective granting ESP it is necessary to create technologies and workbenches of support of control system (creation, maintenance and realization) ESP, including management of orders, processes of interaction of objects and subjects of ESP granting, coordination of actions of all participants of ESP granting process, management of bases and databanks, monitoring of the service realization course, etc. [11].

Providing the effective organization and high-quality operation of ESP control system and the entered automated systems (AS) of interaction of customers and suppliers, maintenances of life cycle, generation and introduction of new ESP, their rational interaction among themselves, and with consumers and service providers as well, an output of the last to the wide market is connected with need of creation of the specialized organizational and technical support system - ESP producer [12]. In Fig. 1 the integrated scheme of interaction of implementation ESP process participants is presented.

This scheme is the result of research carried out under the numerous scientific grants and the basis for Control system of the electronic services rendered to the population was registered as Software in The Federal Institute of Industrial Property (Russia).

Actions of a producer, each his employee have to be directed on minimization of time which is spent by the applicant on receiving necessary service. It is only required from the recipient to submit the application for service and to receive result in due time, and employees of a producer and the corresponding physical and juridical persons who participate in service implementation process have to carry out all other work.

Basic functions of a producer:

- reception of applicants requests for ESP granting;
- ESP designing and creation of regulations according to the universal mechanism;
- interaction with the organizations and the individual entrepreneurs realizing service;
- informing the population and the organizations, a receiving and delivery of the documents, personal data processing connected with providing the specified services;
- representation of applicants interests at interaction with suppliers of service, with other organizations participating in services rendering;
- representation of suppliers interests; delivery the documents created in ESP implementation process to applicants;
- service implementation process monitoring;
- providing information protection.

III. Applicant and Responsible Executive Interaction

Depending on character of the recipient with a producer and the direct executer of service interaction it is possible to select three essentially different types of ESP: information service; simple service; difficult service.

Information service is in obtaining necessary information that it is possible according to several schemes. So, in case of obtaining information in the online mode on the Internet the official inquiry is not necessary, and thus we defined two alternative options of interactions at service realization:

1. The applicant does not use services of a ESP producer, realizing service independently by means of the Internet. Scheme of interaction in this case: Applicant → Responsible executive (organization or individual entrepreneur) → Applicant.

2. The applicant makes the order to ESP producer that generates and realizes this service. Interaction in this case:

Applicant (order of service) → ESP producer (formation of inquiry) → Responsible executive (service realization) → ESP producer → Applicant (receiving result of service).

When obtaining information with formation of official inquiry there are also two alternative options of interactions at service realization, but interaction schemes can be constructed similar to previous case.

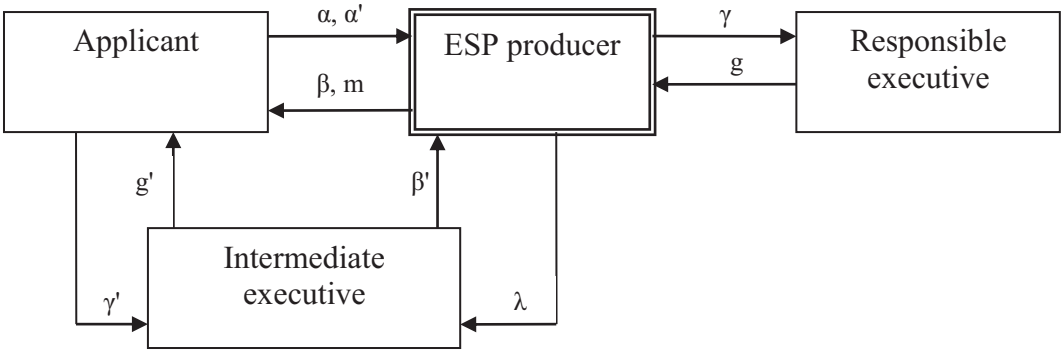


Fig. 2. Applicant-responsible executive interaction scheme

Schemes of interaction at difficult services realization is more difficult. There are three alternative options of interactions at service realization:

1. The ESP producer informs the applicant on forms, templates and samples, of what documents are necessary, and the applicant independently forms the order and sends a document package, conditions and requirements to the executive. Interaction is carried out according to the scheme: The applicant (the order of service) → ESP producer (informing on necessary documents, on forms, templates and samples) → the Applicant (collecting additional documents) → the intermediate supplier (realization of supplementary services) → the Applicant (receiving result of supplementary services, forms a document package) → the Responsible executive (realization of the main service) → the Applicant (receiving result of service).

2. ESP producer on the basis of the documents provided from the applicant (including additionally collected) forms by the applicant proxy the document package, conditions and requirements, interacting with intermediate suppliers (the organizations, individual entrepreneurs, physical persons) and sends it further to the executive. Interaction in this case: The applicant (the order of service) → ESP producer (informing on necessary documents, on forms, templates and samples) → the Applicant (collecting additional documents) → intermediate suppliers (realization of supplementary services) → the Applicant (receiving supplementary services) → ESP producer (formation of a document package for realization of the main service) → the executive (realization of the main service) → ESP producer → the Applicant (receiving result of service).

3. ESP producer by the applicant proxy completely assumes collecting additional documents, forms a document package, conditions and requirements, further interacting with the executive. Interaction: The applicant (the order of service) → ESP producer (organizes the collecting the documents which are not available for the applicant, demanding obligatory participation of the applicant) → the Applicant (participates in the collecting additional documents organized by ESP producer) → intermediate suppliers (realization of supplementary services) → the Applicant (receiving supplementary services) → ESP producer (forms a document package) → the executive (realization of the main service) → ESP producer → the Applicant (receiving result of service) (figure 2).

Designations on the scheme given above:

α - the order (the input document), the documents necessary for realization of service, his condition and the requirement for service which are available for the applicant;

α' - the documents collected by the applicant in addition for realization of the main service;

β - information arriving from ESP producer to the applicant on documents, necessary for service realization: samples, forms, templates, forms;

β' - information necessary for the organization and implementation of collecting the additional documents necessary for service realization;

γ - the full document package, conditions and requirements formed for realization of the main service;

γ' - document packages on realization of supplementary services;

λ - the organization of collecting additional documents;

g - information on result of realization of the main service;

g' - information on result of realization of a supplementary service;

m - output document on result of rendering service.

Similarly, ESP producer will allow to provide effective ESP realization of any kind of complexity.

The scheme presented here is the most common for different application spheres. Use-case diagram of Request for ESP process is presented on figure 3.

IV. Automation of ESP customers and suppliers interaction on the basis of web services

Rendering the presented types of electronic services with use of the offered control system most effectively is realized by means of web services implemented as network of portals. In turn, there are a need of unification of information processes, data structures and interaction interfaces of a producer with the consumer of service [13]. To formalize components of electronic service as web service at first it is necessary to categorize the functions realizing electronic service. Practically for any web service all functions realized for rendering electronic service can conditionally be divided into two categories:

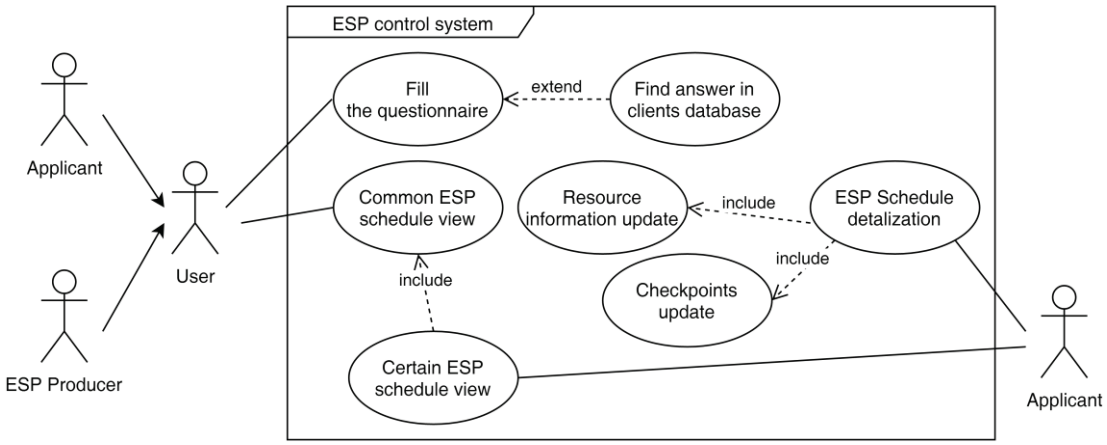


Fig. 3. Use-case diagram of a request for ESP process

• support functions of participants of process of rendering electronic service;

• functions of work with the data necessary for rendering electronic service.

It is possible to present any electronic service in the form of the information process including three basic components: Identifying, Billing and Service.

The identifying component - the firmware subsystem which is defining and unambiguously identifying the consumer of service.

Billing component - the program component of electronic service defining character and an order of financial mutual settlements of the consumer of service with the supplier by means of web service.

One of billing functions is the history maintaining or journalizing of interaction of the user with web service. This function is necessary as from the point of view of ensuring execution of process of rendering service in accuracy according to regulations, and from the point of view of permission of various conflicts and collisions which emergence is possible in the course of rendering service.

Service component - the firmware or procedural part of electronic service providing implementation of the list of actions for rendering the service [14]. In other words, it is realization of interaction of the supplier of service with its consumer for the purpose of rendering the demanded service [15].

The main participants of process of rendering electronic service are: the customer - the party receiving service, the supplier - the party executing some operations directed on service execution, a producer - the party providing organizational and technical support and management of the course of rendering service. Interaction between participants of process of rendering service is regulated by the relevant regulations of various level. The first obstacle in a way of realization of electronic services is the absence of accurate regulations of all its operations components [16]. The formalized regulation allows to solve the following problems peculiar to the sphere of rendering services:

- deadlocks – the situation which is not described by regulations;
- contradictions – possibility of double interpretation of regulations provisions; generally, it is connected with features of submission of the service description in natural language;
- irresponsibility zones – impossibility of establishment responsible for execution (non- suppliers) of concrete operations.

Besides, regulations formalization provides possibility of control realization of rendering service course and its results as well.

The graphic form of representation of process is deprived of these shortcomings. According to the set requirements the LOGI language was developed and intended for the formal description of regulations of rendering service. Rather small amount of language elements allows to describe completely any regulations with the necessary detail level. The description of regulations of service represents the graph which knots are actions and services which need to be executed in the course of electronic service providing.

V. Conclusion

Under the research there were developed the structure of control system of the electronic services rendered to the population and the common scheme of ESP process participants interaction. Each of the presented components of electronic service is obligatory for realization. At the same time functional requirements to each component, program and technical and hardware requirements can be various and depend on many parameters including determined at a stage of creation of electronic service by service provider. Thus, it is possible to certainly say that any electronic service can be realized through web service, and described as the information process including the identifying, billing and service components, and also including a complex of software and hardware tools of realization of work functions with the data which are necessary for rendering electronic service to the population.

The integrated approach to interaction automation of customers and suppliers in control system rendering electronic services to the population and its realization in the form of a set of web services implemented as network of portals will allow to achieve the following significant results:

- decrease in corruption when granting to the population, including small business, various types of service;
- involvement of an inactive intellectual resource of the country in development of real economy with creation of professional collectives;
- intensification of intellectual work;
- release of intellectual and physical work forces of consumers for an effective solution of tasks in their direct professional activity;
- development of informatization in the social and economic sphere;
- providing the required level of confidentiality;
- essential reduction of time and labor expenditure on services realization.

ACKNOWLEDGMENT

This article was prepared with support of grant of President of Russian Federation № MK-5958.2016.9.

REFERENCES

[1] Klochikhin E.A., "Russia's innovation policy: Stubborn path-dependencies and new approaches", Research Policy, 41.9 (2012), pp. 1620-1630.

[2] Daekook Kang, Yongtae Park, "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach", Expert Systems with Applications, 41.4 (2014), pp. 1041-1050.

[3] V.A. Zelepukhina, T.S. Danilova, A.S. Burmistrov, Yu.Yu. Tarasevich, "Particular experience in design and implementation of a Current Research Information System in Russia: national specificity", Procedia Computer Science, 33 (2014), pp. 168-173.

[4] G. Eason, B. Noble, I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions", Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[5] I. Lindgren, G. Jansson, "Electronic services in the public sector: A conceptual framework", Government Information Quarterly, 30.2 (2013), pp. 163-172.

[6] Hamed Taherdoost, Shamsul Sahibuddin, Neda Jalaliyoon, "A Review Paper on e-service; Technology Concepts", Procedia Technology, 19 (2015), pp. 1067-1074.

[7] Lazarev S.A., Konstantinov I.S., Mihalev O.V., Demidov A.V., Shateev R.V., "The development of infrastructure security for distributed information computer environment based on secured portal network", International Journal of Applied Engineering Research, 10 (17), 2015, pp. 38116-38120.

[8] I.S. Konstantinov, Chashin, J.G., Lazarev, S.A., "Simulation of the software-defined network for a high-performance computing cluster", Research Journal of Applied Sciences, 9 (10), 2014, pp. 704-706.

[9] S.A. Lazarev, O.A. Ivashchuk, I.S. Konstantinov, K.A. Rubcov, "Mechanism of Information Exchange Management within Portal Network of Environmental Monitoring Subjects", International Journal of Applied Engineering Research, 9 (22), 2014, pp. 16789-16794.

[10] S.A. Lazarev, I.S. Konstantinov, O.V. Mihalev, V.L. Kurbatov, "Analysis of the single session access model in the distributed portal network of the interacting parties of the informational space", Research Journal of Applied Sciences, 9 (11), 2014, pp. 771-773.

[11] S.A. Lazarev, Konstantinov I.S., Mihalev O.V., Kiselev V.E., "Implementation of unified session access model in a closed virtual environment of distributed information-computational resource system as a secured portal network", Research Journal of Applied Science, 10 (10), 2015, pp. 629-632.

[12] Rudolf R. Sinkovics, Daekwan Kim, Ruey-Jer Jean, "The impact of technological, organizational and environmental characteristics on electronic collaboration and relationship performance in international customer–supplier relationships", Information & Management, 51.7 (2014), pp. 854-864.

[13] Neil F. Doherty, Mahani Shakur, Fiona Ellis-Chadwick, "The role of e-service quality management in the delivery business value", Journal of Retailing and Consumer Services, 27 (2015), pp. 52-62.

[14] Ivan Breskovic, Jörn Altmann, Ivona Brandic, "Creating standardized products for electronic markets", Future generation computer systems, 29.4 (2013), pp. 1000-1011.

[15] Markus Blut, Nivriti Chowdhry, Vikas Mittal, Christian Brock, "E-Service Quality: A Meta-Analytic Review", Journal of Retailing, 2015.

[16] Ramzi El-Haddadeh, Vishanth Weerakkody, Shafi Al-Shafi, "The complexities of electronic services implementation and institutionalisation in the public sector", Information & Management, 50.4 (2013, pp. 135-143.

[17] Jesper Holgersson, Fredrik Karlsson, "Public e-service development: Understanding citizens' conditions for participation", Government Information Quarterly, 31.3 (2014), pp. 396-410.

The crosstalk compensator for fiber-optic data transmission systems with division multiplexing

Dmitry Kiesewetter, Victor Malyugin
Peter the Great Saint Petersburg Polytechnic University
Saint Petersburg, Russia
dmitrykiesewetter@gmail.com, vim@imop.spbstu.ru

***Abstract*—The possibility of compensation crosstalk for different types of division multiplexing in fiber optical transmission lines is discussed. The principle of operation of the crosstalk compensator is described. An example of creating of fiber optic mode division multiplexing transmission line and electronic compensation of crosstalk is analyzed.**

***Index Terms*—Fiber-optics networks, crosstalk, division multiplexing.**

I. Introduction

Various technical solutions for multiplexing the transmitted signals are widely used in fiber-optic communication lines to increase their throughput. The most common in multiplexed fiber-optic data transmission systems is the wavelength division multiplexing (WDM). However, in addition to traditional multiplexing data transmitted on one channel, there are other types of multiplexing, in particular, using special types of encoding with the transmission of different bits at different wavelengths. Several decades ago, the possibility of using different waveguide modes of multimode optical fiber (OF) as independent data transmission channels – mode group division multiplexing (MGDM) have been discussed. The main problem hindering the implementation of systems with modal group multiplexing (MDM) was a high level of crosstalk in data transmission channels caused by binding of waveguide modes. To reduce the crosstalks the crosstalk compensator (cancellers), as [1], were created. Modern possibilities of improving the quality of signals in MDM systems are considered, in particular, in [2] using the adaptive frequency-domain equalization.

Currently, for technical and economic reasons the system with mode division multiplexing is not yet widely available for data transmission. Although descriptions of laboratory models of such lines quite a lot. The basic principles of these lines are summarized in [3]. A detailed list of references devoted to the issues of mode multiplexing is given in [2-5]. Single cases of the use of such systems for multiplexing the analog signals generated by sources of collimated optical radiation which input at different angles relative to the axis of the fiber with step index profile are known also.

However, the crosstalk-canceller or equalizeres similar to [1, 2], can be used for systems of fiber-optic data transmission with WDM, in which crosstalk arises due to nonlinear effects, as well as in lines on the basis of electric cables, for example unshielded twisted pairs (UTP).

It is useful to consider the example of a laboratory prototype of a system with modal division multiplexing to understand the principle of operation of the crosstalk compensator. From a technical point of view, this problem is interesting to determine the characteristics of the compensator at very high values of the coefficients of cross-talking. For practical use the interest to such systems is reborned [2-11] for the following reasons.

- According to the literature, the magnitude of the crosstalk can be relatively small. That allows you to create a transmission line with a length of 1 km and more [6-7].
- Selective excitation of separate waveguide modes in graded-index optical fibers with a length of up to several hundred meters is possible to achieve, for example, using holographic filters. It is possible to achieve of single-mode propagation in an optical fiber with a relatively large core diameter (in comparison with single-mode OF).
- Mode division multiplexing can be used in conjunction with other types of multiplexing, such as WDM [3, 4] or frequency division multiplexing [8].
- Special passive components for MDM systems are created, as for example, the direct-detection mode-group-division multiplexing using fused fiber couplers [9], wavelength-welective switches [10]. There are ready-to-use systems with MDM, as [11].
- A large number of scientific papers in the field of mode multiplexing are published and presented.

We can assume that the mode multiplexing has the perspective of further use in systems of collecting data from fiber-optic sensors and local communication lines.

Therefore, the use of compensator of crosstalk in fiber-optic systems with different type of multiplexing channels is an important task at the present time.

II. Theory

The mutual exchange of energy between the waveguide modes in optical fibers occurs in the distribution process along the entire length of the fiber. That is, such a process can be seen as a distributed effect on the waveguide modes.

Transformation of the propagating spectrum in WDM transmission systems occurs due to nonlinear effects upon distribution also. However, if the duration of the transmitted pulses is substantially longer than the length change of the group delay at energy transfer to other transmission channels, the result of a distributed process of dissipation of energy between the waveguide modes, (or between radiation with different wavelengths) can be described by a summary transformation. Let $P_1^{(in)}$, $P_2^{(in)}$, $P_3^{(in)}$, …, $P_N^{(in)}$ are the power of signals at the input 1, 2, 3, …, N of data transmission channel. Then the signal on the line output ($P_1^{(out)}$, $P_2^{(out)}$, $P_3^{(out)}$, …, $P_N^{(out)}$) can be described by the following system of equations:

$$\begin{aligned} P_1^{(out)}&=\alpha_{11}P_1^{(in)}+\alpha_{12}P_2^{(in)}+\alpha_{13}P_3^{(in)}+\ldots+\alpha_{1N}P_N^{(in)}\\ P_2^{(out)}&=\alpha_{21}P_1^{(in)}+\alpha_{22}P_2^{(in)}+\alpha_{23}P_3^{(in)}+\ldots+\alpha_{2N}P_N^{(in)}\\ &\ldots\\ P_N^{(out)}&=\alpha_{N1}P_1^{(in)}+\alpha_{N2}P_2^{(in)}+\alpha_{N3}P_3^{(in)}+\ldots+\alpha_{NN}P_N^{(in)} \end{aligned} \quad (1)$$

where α_{JK} are the dimensionless coefficients describing the attenuation and the transfer of energy from channel K to the channel J. The diagonal coefficientsα_{JJ} with a small amount of crosstalk characterize the attenuation of the signal in the channel J. In this case we can put:

$$\alpha_{JJ}=\exp(-\alpha_J L) \quad (2)$$

where α_J is the ratio of the specific (linear) attenuation in the J channel, dependent on wavelength and numbers of the waveguide group, L is the length of the line. In contrast to the coefficients α_{JK} coefficients α_J have dimension $м^{-1}$.

For large values of crosstalk coefficients α_{JJ} take into account the return of energy in the channel J through other channels in the distribution process. Therefore, if the attenuation coefficient of one of the channels significantly more than that in the other channels, the value of the coefficient α_{JJ} in the output lines will also be more than the calculated value (2) due to the redistribution of energy between modes. This effect is negligible for fiber-optic WDM systems due to the relatively low power of radiation resulting from nonlinear effects in OF.

The problem of compensating crosstalk for the above approximation reduces to solving the system of the equations (1). It is necessary to find such coefficients β_{JK} which, using the value of the output power for each channel, give ability to determine the power of the input signals, i.e. signals without crosstalk:

$$\begin{aligned} P_1^{(in)}&=\beta_{11}P_1^{(out)}+\beta_{12}P_2^{(out)}+\beta_{13}P_3^{(out)}+\ldots+\beta_{1N}P_N^{(out)}\\ P_2^{(in)}&=\beta_{21}P_1^{(out)}+\beta_{22}P_2^{(out)}+\beta_{23}P_3^{(out)}+\ldots+\beta_{2N}P_N^{(out)}\\ &\ldots\\ P_N^{(in)}&=\beta_{N1}P_1^{(out)}+\beta_{N2}P_2^{(out)}+\beta_{N3}P_3^{(out)}+\ldots+\beta_{NN}P_N^{(out)} \end{aligned} \quad (3)$$

However, the practical implementation of this method for optical fiber communication lines is complicated due to two reasons. First, at high data rates and large numbers of the channels transmitted numerical solution of the system of equations in real time is impossible, so you must use analog signal processing, i.e. in fact, a special analog processor. Second, the coefficients α_{JK}, respectively, and the coefficients β_{JK} for fiber optic lines is unknown at the time of designing the line. For systems with modal multiplexing the coefficients can be affected by bending of the fiber, and in lines with the wavelength multiplexing – are depend on power of radiation sources. Therefore we also need a method of adjustment and the corresponding circuit design of the analog processor.

Due to the fact that the power of the useful signal refers to the coefficients α_{JJ} and the coefficients α_{JK} characterize power of crosstalk, it is useful to consider the relative value of signals in relation to the useful signal. In addition, in fiber optic devices, voltage (U) and current (I) at the output of the photodetector is proportional to the power of the optical signal. Therefore, the system of equations (1) can be represented as follows:

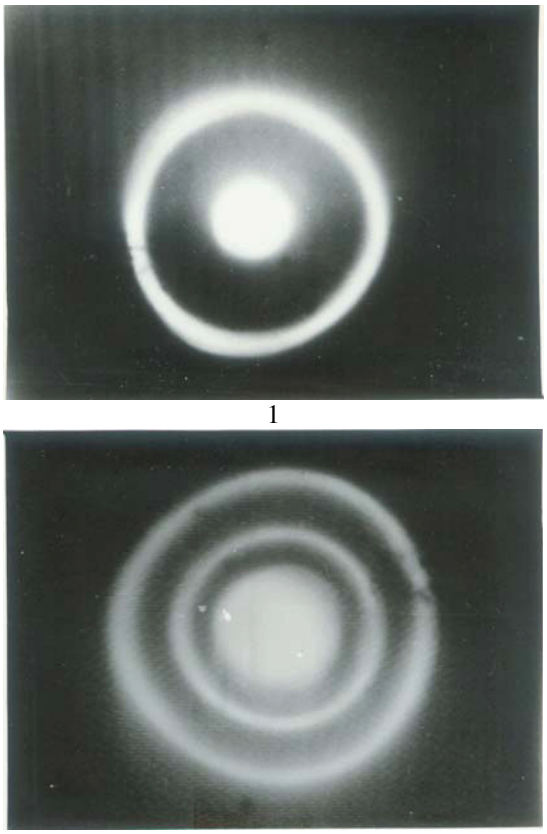
$$\begin{aligned} U_1^{(out)}/(\alpha_{11}U_1^{(in)})&=1+c_{12}u_2^{(in)}+c_{13}u_3^{(in)}+\ldots+c_{1N}u_N^{(in)}\\ U_2^{(out)}/(\alpha_{22}U_2^{(in)})&=c_{21}u_1^{(in)}+1+c_{23}u_3^{(in)}+\ldots+c_{2N}u_N^{(in)}\\ &\ldots\\ U_N^{(out)}/(\alpha_{NN}U_N^{(in)})&=c_{N1}u_1^{(in)}+c_{N2}u_2^{(in)}+c_{N3}u_3^{(in)}+1 \end{aligned} \quad (4)$$

where $c_{JK}=\alpha_{JK}/\alpha_{JJ}$ are is the coefficients of the crosstalk, $u_{JK}=U_{JK}/U_J^{(in)}$.

III. Experimental setup

The simplest version of the mode group multiplexing is implemented using step index optical fiber. In this case, the focused radiation from different sources is introduced at various angles relative to the axis of the fiber. Intensity distribution of exiting radiation from the optical fiber is a ring with the output angle approximately equal to the angle of input relative to the axis or a circle whose center coincides with the axis of the OF. Therefore, it is advisable to use a photodetector with circular form photosensitive elements for effective detecting of the output radiation.

In this work, as the radiation source used semiconductor laser transmitting modules of the type MPD-1, operating up to speeds of 32 Mb/s, the optical fibre with the length of 25 m with the quartz core, with the diameter of 200 microns and with the polymeric cladding. The numerical aperture was approximately equal to 0.3. Collimated radiation from one laser was introduced parallel to the axis of the optical fiber (1 channel). The emission of other lasers was introduced at angles from 10^0 to the value of aperture angle. An example of intensity distribution of outgoing radiation for two and three multiplexing channels is shown in Fig. 1.



1

2

Fig. 1. Intensity distribution of output radiation from the optical fiber with step refractive index profile for: 1 - two channels, 2 - three channels.

For these studies the special silicon photodiode, the design of which is shown in Fig. 2 was made. For this reason, we used the laser transmitting modules with a wavelength of 850 nm are not currently used in communication lines. The photocurrent was amplified by a broadband amplifier. Elements of the photodetector, in which the magnitude of the signal of the respective channel were maximal, were chosen to measure crosstalk and to study working the crosstalk compensator. The exception was channel 1 for which diameter of photosensitive element was chosen approximately half the diameter of the spot formed by low order modal group of OF. This is due to the significant increase power of interference in the channel 1 from the adjacent channels at increasing the ring diameter of the sensor of channel 1 and due to decreasing of level of signal when reducing diameter of the photosensitive element. The use of the central photosensitive element for construction of the photodetector shown in Fig. 2 was possible also.

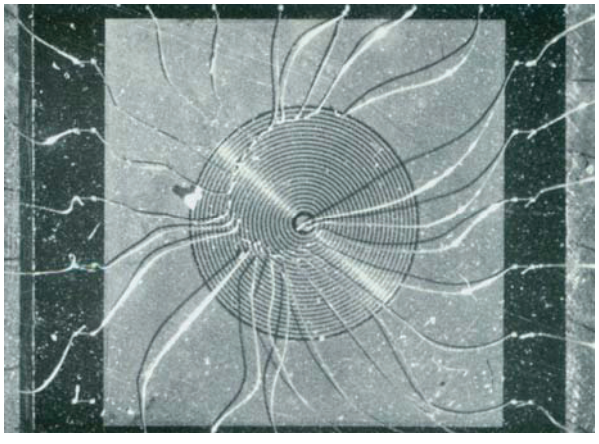
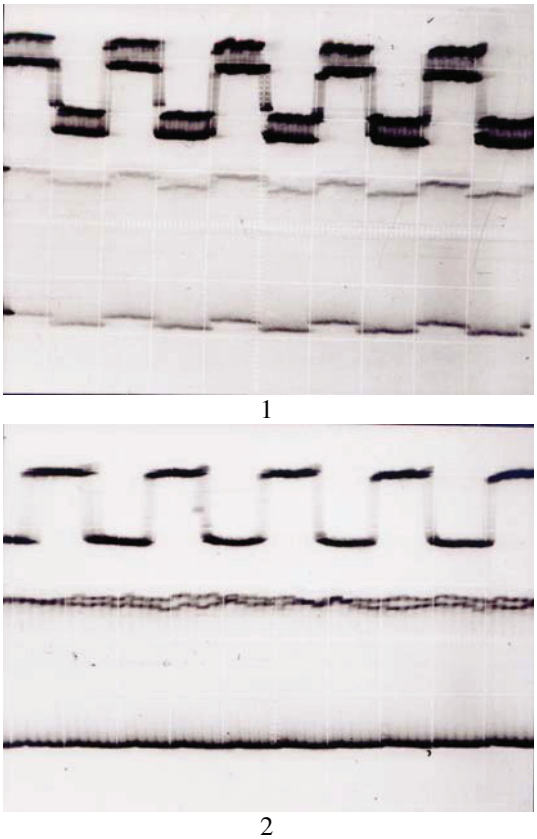


Fig. 2. Appearance of the 24-element photo detector to study of crosstalk.

IV. Obtained results

Examples of waveforms for the case of two multiplexing channels with angles of radiation entering 0^0 and 18^0, obtained on the described experimental setup, are shown in Fig. 3. The pulse duration in first channel is 1 μs, in the second channel – 0.2 μs. In the first case (without the use of a crosstalk canceller, Fig. 3, item 1), the mutual penetration of the transmitted signals is clearly seen in the waveform. Estimated values of the coefficients of crosstalk without the use of a compensator are: c_{12} – 8 dB (15%), c_{21} – 10 dB (10%), when using a compensator – c_{12} – 15 dB (3%), c_{21} – 17 dB (2%). The average decrease of crosstalk when using the compensator was more than 5 times (Fig. 3, item 2). The values of the coefficients of the crosstalk for different cases of launching of modal groups are shown in Table 1 and Table 2.



1

2

Fig. 3. The waveforms for two mode multiplexing channels

TABLE I. THE COEFFICIENTS OF CROSSTALK WITHOUT THE USE OF THE COMPENSATOR FOR THE OPTICAL FIBER WITH LENGTH OF 4.5 M.

K	**The sequence number of the channel that subjected to the influence (*J*)**		
	1	**2**	**3**
1	1	0.8	0.3
2	0.8	1	0.4
3	0.05	0.2	1

TABLE II. THE COEFFICIENTS OF CROSSTALK WITHOUT THE USE OF THE COMPENSATOR FOR THE OPTICAL FIBER WITH LENGTH OF 1.5 M.

K	**The sequence number of the channel that subjected to the influence (*J*)**		
	1	**2**	**3**
1	1	0.5	0.2
2	0.7	1	0.3
3	0.05	0.2	1

The data presented in Table 1 correspond to the optical fiber with the length of 4.5 m quartz core with the diameter of 200 μm and polymeric cladding, with the aperture angle θ_c is approximately equal to 15^0. The input angles (at which the radiation entered into the optical fibre) φ_K (relative to the axis) and output angles (at which were filtering output radiation) γ_K were selected as follows: $\varphi 1=0^0$, $\varphi_2=\gamma_2=0.4\theta_c$, $\varphi_3=\gamma_3=0.67\theta_c$ (channels 1, 2 and 3 representative). The data presented in the Table 2 correspond to the optical fiber length 1.5 m with the quartz core diameter of 50 μm and quartz cladding with the step index refractive profile, with the aperture angle θ_c approximatly equal to 20^0. Angles φ_K and γ_K was following: $\varphi 1=0^0$, $\varphi_2=\gamma_2=0.7\theta_c$, $\varphi_3=\gamma_3=0.85\theta_c$.

The data obtained indicate that first, the coefficients of the crosstalk in optical fibers with step index refractive profile are very high, and secondly, and the use of the crosstalk compensator helps to reduce those values to values acceptable for lines with digital transmission of data.

Pilot study of the possibility of using modal multiplexing in optical fibers with parabolic index profile with a core diameter of 50 μm was carried out. The radiation from the transmitting modules focused in the beam of 5 – 10 μm on the input face of the optical fiber, the center of one of the beams coincide with the axis of the fiber. The center of the second beam was shifted by 10 μm relative to the axis at case of two transmitted channels and by 7.5 μm and 15 μm at case of three channels. The output radiation of the waveguide modes were separated by spatial filters. In this case, the coefficients of the crosstalk were 20 - 50 times less than for the optical fiber with step refractive index profile.

V. FEATURES OF CONSTRUCTION OF THE COMPENSATOR FOR HIGH SPEED DATA TRANSMISSION

The simplest device consisting of inverters and adders, do not to allow to take into consideration the different group delays and dispersion in different channels. The use of the equalizers, similar to [2] in some cases, may be excessive.

VI. CONCLUSION

If chromatic dispersion is dominated in the MDM system, the additional device for it, in comparison with the simplest device, should be a dispersion compensator. If groups delay is dominant then the additional element must be a delay line. Thus, for many applications, the design of the equalizer can be simplified to the above-described crosstalk compensator.

REFERENCES

[1] D. V. Kiesewetter, V. I. Malyugin, V. L. Sukhanov, "The fiber-optic receiving module for mode group division multiplexing system with N transmitted channels", Pat. 1538159, USSR, Bul. Iz., 1990, Issue 3, p. 224 (in Russsian).

[2] S. O. Arik, D. Askarov, J. M. Kahn, "Adaptive Frequency-Domain Equalization in Mode-Division Multiplexing Systems", J. of Lightwave technology, Vol. 32, No. 10, May 15, 2014, p. 1841-1852.

[3] Sercan O. Arik, Joseph M. Kahn, and Keang-Po Ho, "MIMO Signal Processing for Mode-Division Multiplexing. An overview of channel models and signal processing architectures", IEEE Signal Processing Magazine, 2014, March No. 3, p. 25-34.

[4] A. Amphawan, Y. Fazea, H. Ibrahim, "Investigation of Channel Spacing for Hermite-Gaussian Mode Division Multiplexing in Multimode Fiber", 2015 IEEE 11th International Colloquium on Signal Processing and its Applications (CSPA2015), 6-8 Mac. 2015, Kuala Lumpur, Malaysia, p. 34-39.

[5] Keang-Po Ho, J. M. Kahn, "Linear Propagation Effects in Mode-Division Multiplexing Systems", J. of Lightwave technology, Vol. 32, No. 4, Feb. 15, 2014, p. 614-628.

[6] R. Llorente, M. Morant, A. Macho, D. Garcia-Rodriguez, Juan Luis Corral, "Demonstration of a spatially multiplexed multicore fibre-based next-generation radio-access cellular network", 2015 17th International Conference on Transparent Optical Networks (ICTON), 2015, p. 1-4.

[7] C. P. Tsekrekos, A. M. J. Koonen, "Mitigation of Impairments in MGDM Transmission With Mode-Selective Spatial Filtering", IEEE Photonics Technology Letters, Vol. 20, No. 13, July 1, 2008, p. 1112-1114.

[8] S. O. Arik, D. Askarov, J. M. Kahn, "Adaptive Frequency-Domain Equalization in Mode-Division Multiplexing Systems", J. of Lightwave technology, Vol. 32, No. 10, May 15, 2014, p. 1841-1852.

[9] I. Gasulla, J. M. Kahn, "Performance of Direct-Detection Mode-Group-Division Multiplexing Using Fused Fiber Couplers", J. of Lightwave technology, Vol. 33, No. 9, May 1, 2015, p. 1748-1760.

[10] Keang-Po Ho, J. M. Kahn, J. P. Wilde, "Wavelength-Selective Switches for Mode-Division Multiplexing: Scaling and Performance Analysis", J. of Lightwave Technology, 2014, Vol. 32, Issue 22, p. 4326 - 4337.

[11] Mode Division Multiplexor / Demultiplexor, KS Photonics MDM-manual, 2014, 2 p.

A Distributed Particle Filter for Acustic Target Tracking in Wireless Sensor Networks

Faroogh Ashkooti
Department of computer, Mahabad branch,
Islamic Azad University, Mahabad, Iran
Ashkooti_f@yahoo.com

Rahim Rashidy
Department of computer, Boukan branch,
Islamic Azad University, Boukan, Iran
Rahim.rashidy@gmail.com

***Abstract*— One of the applications of wireless sensor networks is target tracking. There are several methods to target tracking and among these methods, particle filter has high capability in solving nonlinear/non-Gaussian systems. Particle filter is one of the methods for Bayesian recursive estimation for position estimation in wireless sensor networks. Clustered management of dense networks is a famous known method and has many benefits in comparison with other management schemes of sensor nodes. In this paper we designed a new distributed version of particle filtering which is more compatible with dynamic clustering of nodes. Our method decrease the processing overhead of sensor nodes and output results are more accurate than current distributed versions of particle filtering in target tracking applications.**

***Keywords*- Wireless sensor networks, target tracking, posterior probability density, sequential Monte Carlo, particle filter.**

I. INTRODUCTION

Recent development in designing RF modules, micro-controllers, sensors and MEMS technology and reducing the cost of production of wireless sensor nodes caused wireless sensor networks (WSNs) to be used more and more in real world applications. Many different and new applications introduced for WSNs. One of these applications that recently gained much attention is tracking of mobile targets. Target tracking can be used in military applications for tracking of enemy's equipment and troops in battlefield and also protection of a banned area for monitoring. There are many challenges ahead to fully accomplish the target tracking in real word applications [1].

One of major issues that we are facing in WSNs is minimizing energy consumption. Since the energy of sensor nodes were supplied by battery, and when battery ran out, the substitution of energy supply is impossible. Another problem is that most of the algorithms that we use for target tracking in WSNs have high communication cost and most of these algorithms uses one fusion center for process of data acquired from network. These algorithms centrally carried out the process. Centralized algorithms reduce the reliability and fault-tolerance of systems and consume more energy in fusion center. Neighboring nodes of fusion center should route data toward this center and this cause unbalanced consumption of energy in networks. Therefore the energy of fusion center and its neighbor nodes consume rapidly and cause the life time of network to be decreased. Fortunately distributed algorithm try to overcome this limitations. Our main goal in distributed algorithms is to evade overusing of one node with distribution of process from the center node among other nodes of network. By doing this we can balance the energy consumption in network and finally it increase the life time of network [2].

We can divide the methods mobile target tracking into three categories as follows: 1) kalman filtering, 2) Markov methods, and 3) Monte Carlo methods. All above mentioned methods can be classified in larger group named Bayesian Estimation methods [7]. The first and second type of these methods has many problems. Study on Bayesian filters lead to creation of Monte Carlo methods. For describing the tracking uncertainty in particle filtering (PF), instead of using direct description of probability density function (pdf) a set of sample particles randomly produce based on probability function that is used. In each step of target tracking we update the particles and by this means the new uncertainty is calculated. Monte Carlo method in comparison with kalman filtering and Markov methods have some advantages that are as follows: 1) This methods can display multi modal distributions, 2) Computational load decreased in comparison to Markov methods, 3) Gives more accurate result in Comparison to Markov method with constant cell size 4) There is no need to access prior knowledge about the uncertainty function, 5) Implementation of this method is easy [8].

The remaining of this paper is organized as follows. Section II describes the formulation for target tracking in framework of recursive Bayesian estimation. Sections III discuss the target tracking by means of particle filtering. Section IV introduces our approach to target tracking and sections V and VI introduce evaluation and ends up the paper with conclusion respectively.

II. FORMULATION OF TARGET TRACKING IN RECURSIVE BAYESIAN ESTIMATION FRAMEWORK

Since the target state estimation is a random problem, it belongs to probability framework. Two general ways for

solving random probability equations exist. These methods are Bayesian and Fisherian methods. In Bayesian method we denote the state of target with x and environmental measurements with y and they are considered as random variables. But in Fisherian method we won't treat state of target as a random variable and state of target is considered as constant and unknown quantities and only measurements are treated as random variable. Since in WSNs state of target and environmental measurement are random, we use Bayesian theory to state estimate of target. In Bayesian method the state and measurement are correlated together by likelihood function which is denoted with $p(y|x)$. For defining the equation of target tracking we should models the state of target and environmental measurements. For modeling this random variable we use stochastic differential-algebraic equations (SDAEs). We use discrete time model and in each step forward the time one unit. For modeling the equation, at first we define the vector x as $x=[a,b,c]$. This vector shows the target position in 3-dimentional Cartesian coordinates system. Evolution of target position equation $\{x_t , t \in \aleph\}$ is computed based on Markov process with information of prior state $x_0 \sim p(x_0)$, system model $p(x_{t+1}| x_t)$ and formulized as follows[3,4]:

$$x_t = f(x_{t-1}, w_t) \quad (1)$$

In Eq. (1), t is the time. We referred to this kind of equation as target model or system model. Function f is nonlinear function of x_{t-1} and w_t is an independent identically distributed (i.i.d) noise. This noise is memory less and is independent from past noises. In other words the noise is white and we consider its distribution as known and we showed it as $p(w_t)$. We formulize the measurement model or sensory system that is a likelihood function of $p(y_t|x_t)$ as follows:

$$y_t = h(x_t, e_t) \quad (2)$$

In Eq. (2), h_t is a nonlinear function of x_t and e_t is also i.i.d noise of measurement. This noise is white and we make it known, we showed it as $p(e_t)$. Target tracking is a computation of recursive estimation of x_t. We can calculate x_t if the prior density $p(x_0)$, system model, and likelihood function be known. Of course all above items should be applied along with Bayesian theorem[3,4].

We denote all measurements up to time t, i.e., $\{y_i\}_{i=0}^{t}$ with Y_t. Therefore we want to design a recursive scheme for how to update the a posteriori density recursively when new measurements is done. Bayesian filter is a well known method for performing this recursive updates. Bayesian filter estimation performs in two phases as follows:1) prediction phase and 2) update phase

Prediction phase: In this phase we use system model to predict the current state of the target objects. We want to compute probability distribution function $p(x_t|Y_{t-1})$. Also the system model has defined as conditional density distribution $p(x_t|x_{t-1})$, i.e., based on Markov property, we can say that current state of target is dependent to previous state of target. Therefore predicted density distribution of x_t with a integral can be calculated as follows:

$$p(x_t | Y_{t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | Y_{t-1}) dx_{t-1} \quad (3)$$

Using system model an through chapman-kolmogrov method in this phase we can calculate the target state at time t [4,7].

Updating Phase: In this phase we assume that measurements of time t is done and we can calculate posterior probability function $p(x_t|Y_t)$. The measurement y_t is independent from Y_{t-1} (all previous measurements). Therefore this density could update by means of Bayes theorem as follows:

$$p(x_t | Y_t) = \frac{p(y_t | x_t) p(x_t | Y_{t-1})}{p(y_t | Y_{t-1})} \quad (4)$$

The denominator can be calculated as follow:

$$p(y_t | Y_{t-1}) = \int p(y_t | x_t) p(x_t | Y_{t-1}) dx_t \quad (5)$$

In the next time step, predicting and updating phase will repeat in recursive mode so that in discrete time steps they could estimate posterior density function and by this means they could track the target. This schema is known as recursive Bayesian filtering. This schema is shown in figure 1.

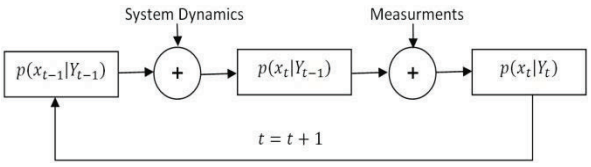


Figure 1. Basic schema of Bayesian filter.

The main problem in Bayesian methods is that solving this equation through analytical methods is difficult. The multidimensional integrals involved only permit an analytical solution in a few special cases. The most important special case is when the dynamic model is linear/Gaussian. In this case kalman filter can be used for solving these equations. But in many applications the model is nonlinear/non Gaussian and we should offer suitable method for solving these equations. One solution is using numerical methods. The main method which we used in this paper is using of sequential Monte Carlo method as a numerical methods. In the nest section we discuss about this algorithm [4,8].

III. SEQUENTIAL MONTE CARLO METHODS

According to the Bayesian method, $p(x_t|Y_t)$ contains all statistical information available about the state variable x_t based on the information in the measurements Y_t. Sequential Monte Carlo methods or particle methods, deal with the problem of recursively estimating the probability density

function $p(x_t|Y_t)$. Sequential Monte Carlo Methods belongs to the classification of simulative-based methods and The main idea underlying the sequential Monte Carlo methods is to represent the probability density function by a set of samples and associated weights .These samples known as particles. This method is a general framework for estimating the nonlinear/non Gaussian systems and has high flexibility for modeling of sophisticated systems. This method should evaluate thousands of particles until reaches satisfactory estimation in each time step. In practice the only disadvantage of this method is high processing overhead and numerous calculations of particles and this causes the infeasibility of using this method in real-time applications. In order to display the steps of particle filter algorithm we introduce some notations. We consider that M is number of particles and show the state vector of particles and its weights as $\{x_t^i, q_t^i\}_{i=1}^{M}$. In this notation i denote the number of particle and t is the time step which particle is produced. These particles were produced and updated in a recursive fashion and they accomplish the estimation of posterior density function $p(x_t|Y_t)$. By considering these particles the posterior density function will be estimated as follows [4,5]:

$$p(x_t \mid Y_t) \approx \sum_{i=1}^{M} q_t^i \delta(x_t - x_t^i) \qquad (6)$$

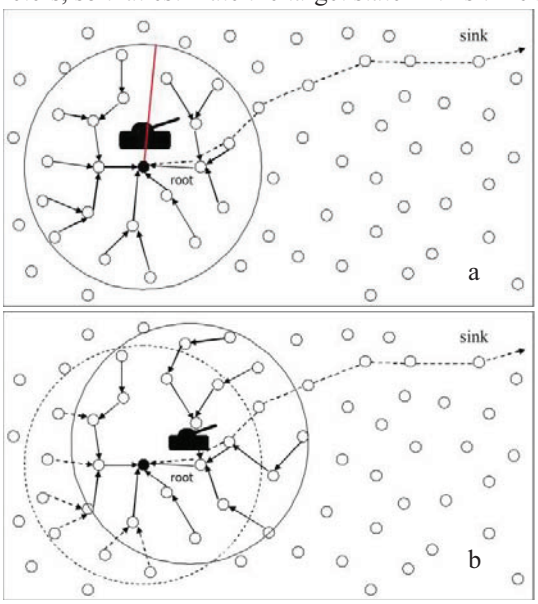
In this equation δ(0) is Delta Dirac function. When the number of particles are large in number (M→∞) we got an estimation that approximately is equal to real value. In other word we achieve an approximation answer which its accuracy directly relates to the number of used particles. In real application we are unable to choose a large number of particles because this causes that the computation cost of algorithm arises and consume more energy of sensor nodes. A version of particle filter algorithm that known as Sampling-Importance-Resampling (SIR) algorithm is presented below [6,5]:

1. Initialize the particles, $\{x_0^i\}_{i=1}^{M} \sim p(x_0)$ and set t= 0.
2. Measurement update: calculate importance weights $\{q^i\}_{i=1}^{M}$ according to $q_t^{(i)} = p(y_t \mid x_t^{(i)}) \quad i=1,2,...,M$ and normalize them $\tilde{q}_t^i = \frac{q_t^i}{\sum_{j=1}^{M} q_t^j}$
3.Calculate the estimation as follows: $\hat{x}_t = \sum_{i=1}^{M} \tilde{q}_t^{(i)} x_t^{(i)}$
4.Resampling, draw M particles, with replacement, according to $\Pr(x_t^{(i)} = x_t^{(j)}) = \tilde{q}_t^{(j)}$ i=1,…,M
5.Time update: predict new particles according to $x_{t+1\|t}^{(i)} \sim p(x_{t+1\|t} \mid x_{t\|t}^{(i)})$ i=1,…,M
6. Set t= t + 1 and iterate from step 2.

In the above mentioned algorithm, we draw samples according to prior density function $p(x_0)$ at first. New measurement is used for each particle to assign a probability represented by the normalized importance weight in the measurement update. This probability is calculated using the likelihood function which describes how likely it was to obtain the measurement given the information available in the particles. The normalized importance weights and the corresponding particles constitute an approximation of the filtering density. The re-sampling step will then return particles which are equally probable. The time update is just a matter of predicting new particles according to the system model. These predicted particles form the starting point for another iteration of the algorithm.

IV. DESINED SOLUTION FOR TARGET TRACKING

This scenario will be as follows. When a mobile target enters the monitored environment by wireless sensor network, at any time step dynamic clustering over the nodes will be carried out base on estimation of previous time step. Therefore in all time steps the target will be in one active cluster. We assume some services such as routing and dynamic clustering are available. Then base on an algorithm one of nodes inside the cluster will be chosen as leader [12]. The active nodes inside the cluster will measure the environment and based on distributed particle filter algorithm, they will do operation over this observation and send their result to leader. The leader takes all data and performs the particle filter algorithm with adjusted parameters, so that estimate the target state in this time step.



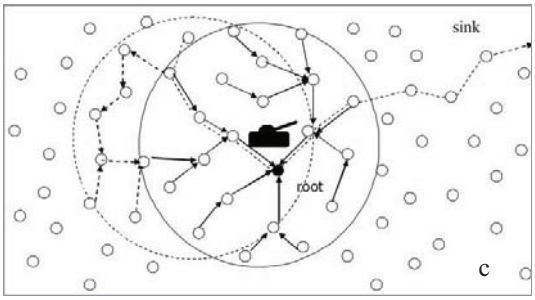


Figure 2. Target tracking scenarios.

After this operation, leader sends the estimation to sink using a used routing algorithm. After completion of operation in this time step we move to next time step and the operation will be continued as we described. Figure 2 shows this scenario.

In our proposed algorithm, those newly entered nodes takes X_{t-1}^s from nearest neighbor which participated in tracking in previous time step. Then new nodes produced their particles based on these particles. Some of notations that is used in representing the algorithm is as follows [11]:

$\{\tilde{x}_t^i\}_{i=1}^M = \tilde{X}_t$: set of particles before resmpling.

$\{\tilde{q}_t^i\}_{i=1}^M = \tilde{Q}_t$: set of set of un-normalized weights.

$\{x_t^i\}_{i=1}^M = X_t$: set of particles after re-sampling.

$\{q_t^i\}_{i=1}^M = Q_t$: set of normalized weights.

In the following, a superscript s indicates node s ∈ {1, ..., S}, and a superscript i indicates the ith particle. So based on proposed method, we redesign our algorithm as follow:

1. The Leader partition set of particles X_{t-1} to subsets of X_{t-1}^s,(s=1,…,S). it means that $X_{t-1}^s = \{x_{t-1}^{s,i}\}_{i=1}^{M_s}$ that $X_{t-1} = \bigcup_{s=1}^S X_{t-1}^s$
2. Leader send $\{X_{t-1}^s\}$ to nodes (s=1,…, S).
3. The active nodes inside the cluster do these operation parallel:
   a. They do environmental observations.
   b. If this node was inside the cluster in previous time step, produce $\tilde{X}_t^s$ based on X_{t-1}^s.
   c. If this node was not inside the cluster in previous time step, It takes $V^s = \{x_{t-1}^{s,i_S}, q_{t-1}^{s,i_S}\}_{i_S=1}^k$ from the nearest neighbor which was inside the cluster in previous time step.

      $Chosen\ neighbor = \min_{s \in S}((x_s - x)^2 + (y_s - y)^2)$

      This node after receiving V^s, produce $\tilde{X}_t^s$ particles based on V^s.
   d. Calculate the weights of $\tilde{Q}_t^s$.
   e. Evaluates the sum of local weights, $\tilde{q}_t^s$ as follows:

$$\tilde{q}_t^s = \sum_{i_S=1}^{M_S} \tilde{q}_t^{s,i_S}$$

   f. Normalizes locally $\tilde{Q}_t^s$ to obtain Q_t^s.
   g. Computes the local estimate, $\hat{x}_t^s$ as follows:

$$\hat{x}_t^s = \sum_{i_S=1}^{M_S} q_t^{s,i_S} \tilde{x}_t^{s,i_S}$$

   h. Re-sampling on $\tilde{X}_t^s$ to produce X_t^s.
   i. Sends $(\hat{x}_t^s, \tilde{q}_t^s)$ to leader.
4. The leader perform this operation after receiving the response of all nodes
   a. Compute $\tilde{q}_t = \sum_{s=1}^S \tilde{q}_t^s$
   b. Normalization globally weights as follow,

$$q_t^s \triangleq \frac{\tilde{q}_t^s}{\tilde{q}_t}$$

   c. Compute the estimation, $\hat{x}_t$ as follows,

$$\hat{x}_t = \sum_{s=1}^S q_t^s \hat{x}_t^s$$

   d. Send $\hat{x}_t$ to sink and set time as *t=t+1*.
5. For operations of next time step do clustering algorithm and leader selection algorithm, then go to stage 3.

As you see, all nodes produce samples after performing environmental observations. But other nodes which are remained in the cluster from previous time step produce particles based on the particles of their previous time step and continue operations. But newly-entered nodes into cluster start operation based on received particles from nearest neighbor nodes. In our simulation we compare our algorithm with local distributed particle filter [11].

In this paper, we assume that the distance sensor nodes to an audio source that is our target object is at least 5 times the size of the target object. Furthermore, we assume that the target object is a point of audio source. Target object located in uniformly smooth surface and sound waves propagate in two-dimensional space and sound reflection effect has not been considered in collision with terrain and obstacles. Sound intensity at a distance r from the sound source is calculated according to the following equation.

$$I = \frac{w}{4\pi r^2} \quad (7)$$

In this formula, W is power of source audio. Sound intensity is measured by watts per square meter. Human's power hearing of is 10^{-12} w/m^2. It is hard to measure sound intensity and sound pressure instead measured by the microphone. P is sound pressure. Unit of measurement of sound pressure base on reference pressure $p_{ref} = 20\mu Pa = 20 \times 10^{-6}(\frac{N}{m^2})$ is micro-pascal, and base on the decibel scale is expressed by the following equation [13].

$$SPL = 10Log_{10}(\frac{p^2}{pref}) \quad (8)$$

In this study, the audio source is considered a heavy tank in which has sound pressure of 100 dB that equal with 2 Pascal. We consider the state vector as $\{x, \dot{x}, y, \dot{y}\}$ that represents the position and velocity of target. Base on equation (1) we assume system has Gaussian noise and is as follow:

$$x_t = f(x_{t-1}) + G_t w_t \quad (9)$$

$$f(x) = \begin{bmatrix} 1 & \frac{\sin(\omega T)}{\omega} & 0 & -\frac{1-\cos(\omega T)}{\omega} & 0 \\ 0 & \cos(\omega T) & 0 & -\sin(\omega T) & 0 \\ 0 & \frac{1-\cos(\omega T)}{\omega} & 1 & \frac{\sin(\omega T)}{\omega} & 0 \\ 0 & \sin(\omega T) & 0 & \cos(\omega T) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} x$$

$$G_t = \begin{bmatrix} \frac{1}{2}T^2 & 0 & 0 \\ T & 0 & 0 \\ 0 & \frac{1}{2}T^2 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}, \quad Q = \begin{bmatrix} 0.5^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 0.02^2 \end{bmatrix}$$

Base on equation (2) we have assumed sensing model as follows:

$$y_t = h(x_t) + e_t$$

$$h(x_t) = \sqrt{x^2 + y^2}$$

V. EVALUATION

For simulations we consider 30 sensor nodes with acoustic sensing and communicating range of 80 meters. These sensors were dispersed on a square area with 300*300 meters dimensions. We assumed that time synchronization accomplished with maximum error 10^{-4}. The simulation process has carried out in a discrete time and for 200 time steps and with 15 Monte Carlo iterations. We carried out simulation by using number of 500, 1000 and 3000 particles. Based on this simulation the average error in each time step has been computed.

Figure 3 shows the average error of Local Distributed Particle Filter (LDPF) method and our proposed Particle Filter (PF) method with 15 Monte Carlo iterations with *M=500* particles. Our proposed algorithm had $\overline{avg} = 0.75$ average error and locally distributed algorithm had $\overline{avg} = 0.87$ average error.

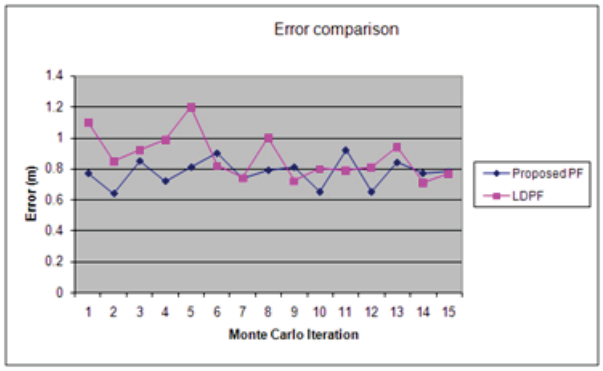


Figure 3. Comparisons of average error using M=500.

Figure 4 shows the average error of LDPF method and our proposed PF method with 15 Monte Carlo iterations with *M=1000* particles. Our proposed algorithm had average error $\overline{avg} = 0.59$ and locally distributed algorithm had $\overline{avg} = 0.69$ average error.

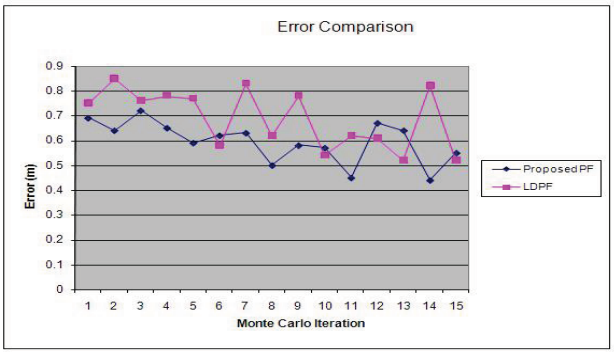


Figure 4. Comparisons of average errors using M=1000.

Figure 5 shows the average error of LDPF method and our proposed PF method with 15 Monte Carlo iterations with *M=3000* particles. Our proposed algorithm had $\overline{avg} = 0.37$ average error and locally distributed algorithm had $\overline{avg} = 0.49$ average error.

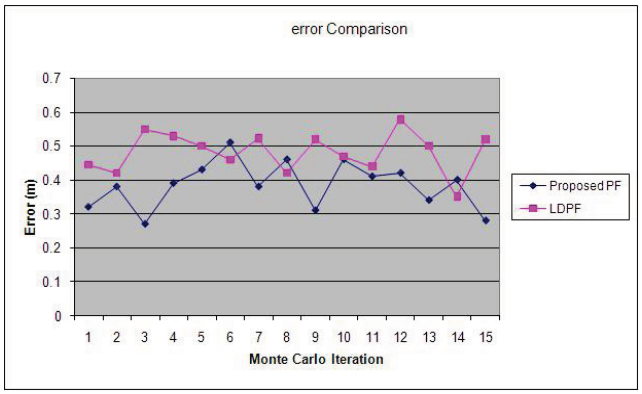


Figure 5. Comparisons of average errors using M=3000.

In the below table you'll see results.

Table 1-comprasion of error

# of particles	Error	
	Proposed PF	LDPF
500	0.75	0.87
1000	0.59	0.69
3000	0.37	0.49

VI. CONCLUSION

Based on findings from simulations results we conclude that our proposed algorithm has lover average error in comparison with local distributed particle filtering method. We calculated average error using 500, 1000 and 3000 particles and we found that for each of these conditions we have 0.12, 0.10 and 0.12 fewer error compared to local distributed particle filtering method. Beyond these, our proposed method is more compatible with clustered management of WSNs. This method decreases the communication and processing overhead of sensor nodes in doing target tracking by using information of neighbor nodes for assigning initial values for some of the particles.

REFERENCES

[1] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam and Erdal Cayiric, “A Survey on Sensor Networks”, IEEE Comunication Magzine, August 2005.

[2] S. Pashazadeh and M. Sharifi, "Simulative Study of Target Tracking Accuracy Based on Time Synchronization Error in Wireless Sensor Networks", IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems (VECIMS 2008), Istanbul, Turkey, 14-16 July 2008, pp. 68-73

[3] Haug, A.J. "A Tutorial on Bayesian Estimation and Tracking Techniques Applicable to Nonlinear and Non-Gaussian Processes". The MITRE Corporation, USA, Tech. Rep, Feb 2005.

[4] D M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", IEEE Transactions on Signal Processing, 50(2):174-188, February 2002.

[5] Doucet, N. de Frietas, and N. Gordon, editors. "Sequential Monte Carlo Methods in Practice. Statistics for Engineering and Information Science." Springer-Verlag, New York, 2003

[6] Ristic, B.; Arulampalam, S.; Gordon, N. "Beyond the Kalman Filter: Particle Filters for Tracking Applications",Artech House(2004).

[7] L.Moreno, S.Garrido, and M.L.Munoz, "Evolutionary filter for robust mobile robot localization." Robotics and Autonomous Systems 54 (2006) pp 590–600.

[8] F.Dellaert, D. Fox, W. Burgard, and S. Thrun, “Monte carlo localization for mobile robots,” Proceedings of the IEEE International Conference on Robotics and Automation (ICRA-99), pp. 1322–1328, 1999.

[9] Garrick Ing, Mark J. Coates," Parallel particle filters for tracking in wireless sensor networks", Signal Processing Advances in Wireless Communications, 2005 IEEE 6th Workshop onVolume , Issue , 5-8 June 2005 Page(s): 935 – 939.

[10] M.J. Coates, “Distributed particle filtering for sensor networks,” in Proc. IEEE/ACM Int. Symp. IPSN, Berkeley, CA, Apr. 2004.

[11] Anwer S. Bashi, Vesselin P. Jilkov, X. Rong Li, Huimin Chen. " Distributed Implementations of Particle Filters", isif 2003.

[12] YAN Zhen-ya, ZHENG Bao-yu, CUI Jing-wu. " Distributed Particle filter for target tracking in wireless sensor network", This work is supported by Natural Science Foundation of China under Grant 60372107, 2005.

[13] F. A. Everest, and K. C. Pohlmann, "*Master Handbook of Acoustics*", 5th Edition, McGraw-Hill, NewYork, NY, USA, 2009, pp: 19-38.

Performance Improvement of 40 Gb/s WDM Systems by Optimization of Dispersion Map

Muhammad Yousaf Hamza
Pakistan Institute of Engineering and Applied Sciences
Islamabad, Pakistan
myhamza@pieas.edu.pk

Imran Hayee
University of Minnesota, Duluth, Minnesota, USA
ihayee@d.umn.edu

***Abstract*— We report the results of 5 channels, 40 Gb/s/channel WDM system with significant launch power in the presence of fiber nonlinearities, by propagating intensity modulated CSRZ pulse train through standard SMF. These results show that traditional approach of complete dispersion compensation using DCF just after each SMF span is not necessarily the optimum choice for better system performance. Rather, the performance of optical links highly depends on the percent of dispersion compensation at different location of DCFs. These investigations demonstrate that the performance of practical WDM systems can be enhanced significantly by using optimum dispersion map for WDM communication systems.**

***Keywords*—WDM transmission, dispersion management, fiber nonlinearities, optimum dispersion map.**

I. INTRODUCTION

The data volume is increasing at very fast rate due to the video-centric demand in the present world of communication. In addition, higher transmission rate with more accuracy is highly desired across the globe. Optical fibers, as they have extremely larger bandwidth and many other attractive features, [1-3], are in commercial use to achieve these desires.

The quality of the signal is degraded when it is propagated in optical fiber. It is due to the degradation effects as fiber attenuation, group velocity dispersion (GVD) and nonlinearities including self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) [4-7] etc. The minimization of these degradation effects is always desired in optical communication systems to maximize the transmission distance, the date rate and the accuracy of optical fiber links.

Optical amplifiers [4] and dispersion compensating fibers (DCFs) [8-10] are used to compensate fiber loss and GVD effects, respectively. If fiber nonlinearities can be ignored, the GVD effect can be compensated completely using any dispersion map, and hence the input signal can be recovered at the receiver side in its original form. However, the periodical use of optical amplifiers in communication links enhances the launch power at each of single mode fiber (SMF) span, thus nonlinearities can't be ignored in the modern WDM communication systems. The efforts are being done to enhance the data rate of WDM optical systems [11-14]. One such effort is dispersion management [7,15-18] which is of prime worth, particularly when optical pulses having smaller width and higher power are used in WDM systems to maximize the transmission distance and the data rate.

For this work, we consider 5 channels, 40 Gb/s/channel, intensity modulated WDM system comprising of a 2-spans×100 km/span link along with the use of optical amplifiers and DCFs, and investigate the results for the optimum dispersion map to enhance the system performance. For this work, numerical split step Fourier model [19-23] is used for solving nonlinear Schrodinger equation for signal propagation considering the effects of fiber attenuation, group velocity dispersion and nonlinearities such as SPM, XPM, and FWM. We explore and present the results of the optimization of dispersion maps, the corresponding eye diagrams and Q-values for various setups of fiber links to highlight the importance of the location of DCFs and their compensation values for the improvement in the performance of WDM systems.

II. THEORY

In single mode fiber, the quality of the transmitted optical signal is degraded mainly due to fiber attenuation, GVD and nonlinearities. The degradation of the signal impose limits on the date rate, the accuracy and the capacity of the transmission systems. Fiber degradation effects are briefly described below.

Fiber loss attenuates the pulse power and hence limits the maximum transmission distance. Erbium-doped fiber amplifiers are used to compensate fiber loss in WDM links. Group velocity dispersion arises because fiber refractive index is wavelength dependent. It broadens the pulse and thus limits the data rate of transmission links. For compensation of GVD, dispersion compensating fibers (DCFs) [4,15-17], have opposite sign of GVD parameter to that of SMF spans, and hence can be used to compensate the effects of GVD in long-haul WDM transmission systems.

Fiber nonlinearities arise because fiber refractive index is also dependent on the intensity of the input signal. It is desired to launch the input signal having higher power to propagate it for longer distance, however the higher power level enhances the nonlinear effects as SPM, XPM and FWM [4,5,7] that are

M. Y Hamza was at University of Minnesota, Duluth, USA when he carried this work. He acknowledges Higher Education Commission (HEC), Pakistan for providing financial support to conduct this work.

the major sources to degraded the performance of optical communication systems.

Nonlinear Schrodinger equation (NLSE) is used to illustrate the effects of fiber attenuation, dispersion and nonlinearities. Split step Fourier model is used in this work to solve NLSE numerically and to propagate the signal through 5 channel, 40 Gb/s/channel WDM system. DCFs are used to compensate the effect of fiber dispersion of SMF spans. The following equation shows the condition for 100% dispersion compensation [4],

$$L_1D_1 + L_2D_2 = 0, \quad (1)$$

where L_1 represents the net length of SMF spans in the link, L_2 represents the net length of the DCF that's needed for 100% compensation. The D_1 and D_2 are dispersion parameters for SMF and DCF, respectively.

In traditional approach, the dispersion of each SMF span is 100% compensated by using a DCF just after the span, as shown in Fig. 1. In this setup, the length of each of the DCF is pre-fixed for 100% compensation.

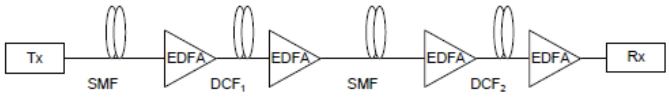


Fig. 1. Traditional setup for dispersion compensation using DCF for 2-spans system. Tx: Transmitter, Rx: Receiver

When the signal power is significant, and optical amplifiers are used periodically in fiber links, then due to the interaction of GVD and fiber nonlinearities, the performance of WDM systems is dependent on the dispersion-map and launch power. Different dispersion maps will yield different Q-values. Hence, the optimization of dispersion map is of prime worth to enhance the performance of optical communication systems. The system and its parameters considered for this work, the results and discussions for the proposed optimization of dispersion map are provided below.

III. Results and Discussions

We analyze 5 channels intensity modulated WDM system, each channel operating at 40 Gb/s with channel spacing of 200 GHz. The fiber link consists of two SMF spans, each of length 100 km. Fiber loss is compensated using Erbium-doped fiber amplifier (EDFA). The noise figure of EDFA is considered as 6 dB.

For compensation of group velocity dispersion in 2-spans system, DCFs can be placed at different location as (i) prior to the fiber span, called pre-compensation technique, (ii) after the fiber span, called post-compensation technique, or (iii) distributing at both ends, called combination of pre- and post-compensations. On these bases, Fig. 2 shows three setups for system under consideration. The parameters for SMF and DCF are provided in Table I.

An independent 64-bit pseudo-random carrier suppressed

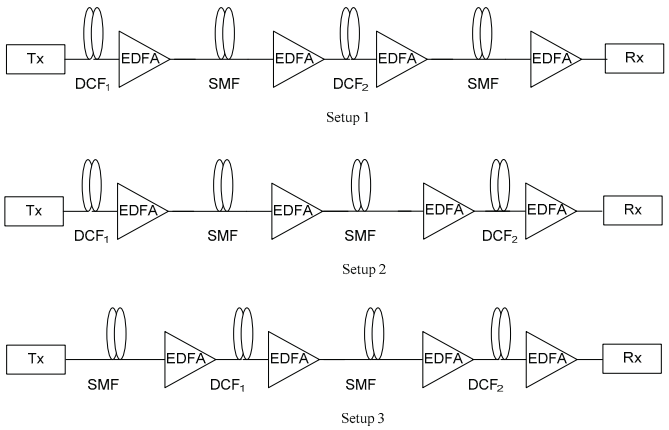


Fig. 2. Three setups for dispersion compensated transmission link using DCFs. Tx: Transmitter, Rx: Receiver

TABLE I
Parameters of SMF and DCF

Fiber Type	SMF	DCF
Loss (dB/km)	0.22	0.5
Dispersion (ps/(nm.km))	17	-85
Effective area $(\mu m)^2$	100	20

return-to-zero (CSRZ) pattern is transmitted in each of the channels. The power 0 dBm is launched at the input of each DCF. However, the launch power into each SMF span was set to much higher level as 14 dBm to have significant fiber nonlinear effects. These required power levels are attained by suitable gains of EDFAs.

The electric field of all the channels is multiplexed at the transmitter side. The multiplexed signal is propagated through fiber link for each of the three setups using SSFM. An optical band-pass filter (BW = R) is used at the receiver to separate each of the channels. Then the middle channel (the testing one) is passed through an electric filter (BW= 0.7R, 20th-order Bessel filter) to obtain an eye diagram. The total accumulated amplifiers noise is added at the receiver. Finally system performance is evaluated by determining the Q-value.

For the system under consideration, the calculations using Eq. (1) show that the total DCF length for 100% compensation is required as 40 km. This length value will be used for computing percent compensation for each of the DCF placed at different locations.

First of all, for a reference and comparison, a traditional approach of post-compensation is used, i.e. each of the SMF spans is compensated completely by its following DCF, the setup already shown in Fig. 1. For this, the net 40 km length of DCF for perfect compensation is distributed equally between DCF_1 and DCF_2, which means the length of each of this DCF is fixed as 20 km. The eye diagram for this setup is shown in Fig. 3.

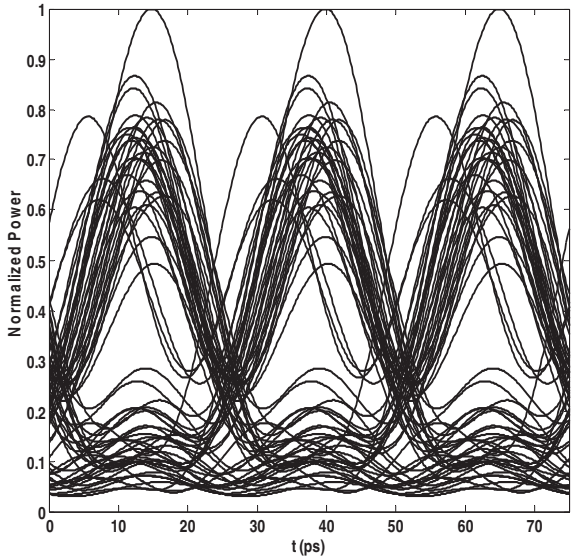


Fig. 3. Eye diagram for net 100% compensation divided equally between DCF_1 and DCF_2 as traditionally post-compensation technique.

The Q-parameter corresponding to this eye diagram is calculated as 22.81 dB. For all of the following results and discussions, this value of Q will be considered as a reference for the comparison of the results with optimum dispersion maps, as discussed below.

To explore the optimization dispersion maps for each of three setups of Fig. 2, the length of none of the DCF is pre-fixed. Rather, the total length of 40 km is varied from 0 to 40 km (in a step of 1 km) for DCF_1. The length of DCF_2 is accordingly set so that the net compensation remains 100%. The Q-value is determined for each of the possible combinations of the length of DCF_1 and DCF_2 and report the best Q-value for each of the setups.

For setup 1, the investigations showed that maximum Q-value as 21.88 dB can be achieved with 80% (32 km length) compensation at DCF_1 and 20% (8 km length) at DCF_2 locations. This Q-value is even slightly lower than the reference value for the traditional approach. The corresponding eye diagram is shown in Fig. 4.

Fig. 5 shows the eye diagram for setup 2, for which it is found that 55% of pre- and 45% of post-compensations is the optimum choice with maximum Q-value as 27.68 dB which is ~ 6 dB higher than the reference value. The improvement in Q-value for this setup is due to the fact that eye for this setup is more opened than for the eye shown in Fig. 3.

Finally, for setup 3, the investigations showed that DCF_1 with 68% and DCF_2 with 32% compensations yield the best result with Q-value of 30.42 dB i.e. about 7.6 dB higher than the reference one. The opening of the corresponding eye is much more as shown in Fig. 6. The results for all these setups are summarized in Table II.

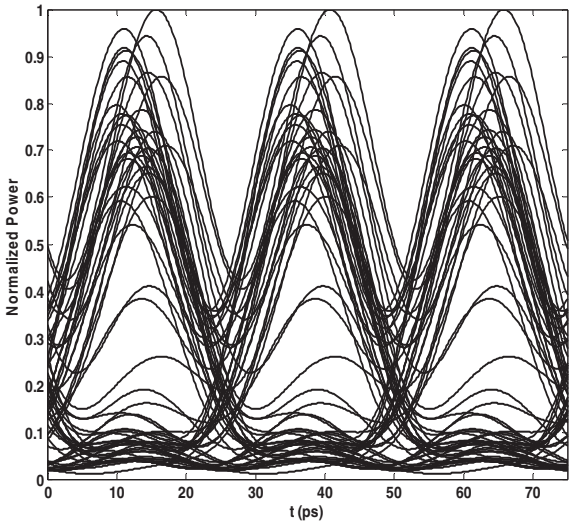


Fig. 4. Eye diagram for setup 1, with optimum distribution of net 100% dispersion as 80% at DCF_1 and 20% at DCF_2 locations.

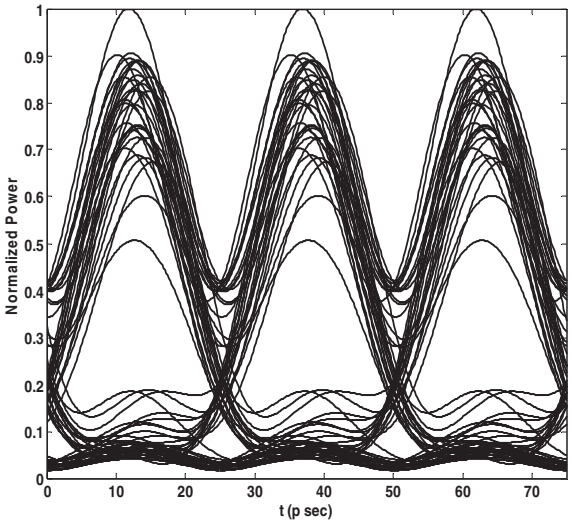


Fig. 5. Eye diagram for setup 2, with optimum distribution of net 100% dispersion as 55% at DCF_1 and 45% at DCF_2 locations.

The dispersion maps for the traditional approach and for all three setups are shown in Fig. 7. In fact, these maps show that how the net dispersion evolves for SMF and DCF combinations for various setups.

It is mentioned that the procedure using exhaustive approach to explore the optimum dispersion map is discussed here to highlight the importance of the optimization of dispersion map. The exhaustive based optimization procedure is not scalable for systems with more than two spans. A suitable optimization technique may be used to explore optimum dispersion map for long haul transmission systems.

These results show that Q-value and hence the performance of WDM communication systems can be enhanced using optimum dispersion map design.

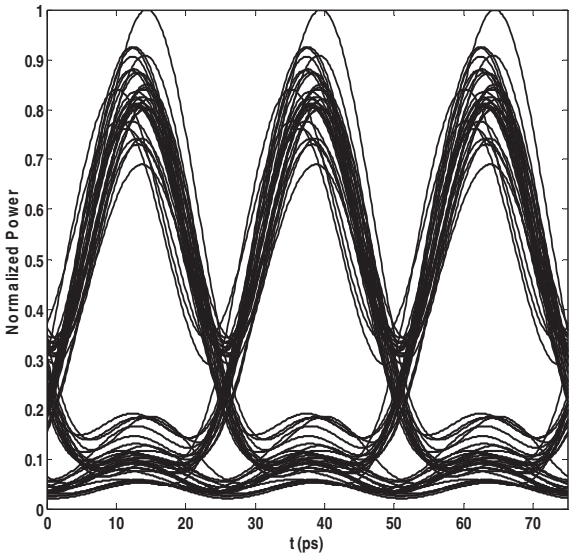


Fig. 6. Eye diagram for setup 3, with optimum distribution of net 100% dispersion as 68% at DCF_1 and 32% at DCF_2 locations.

TABLE II
COMPENSATION VALUES FOR OPTIMUM DISPERSION MAP
(Net compensation is 100%)

Setup	Percent of net dispersion compensation		Q (dB)
	DCF_1	DCF_2	
Traditional Approach	50	50	22.81
Setup 1	80	20	21.88
Setup 2	55	45	27.68
Setup 3	68	32	30.42

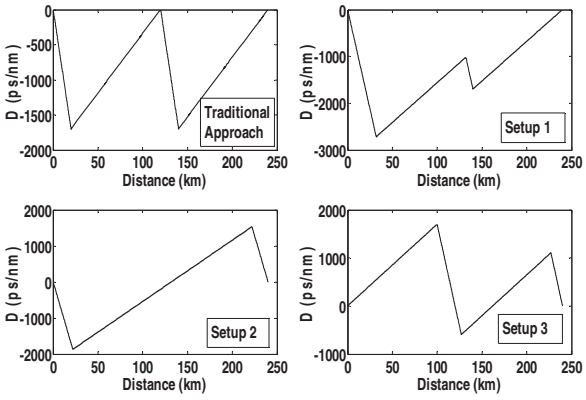


Fig. 7. Dispersion maps for traditional approach and three setups of Fig. 2.

IV. CONCLUSION

The investigations and results of dispersion map optimization have been presented for 5-channels, 40 Gb/s/channel intensity modulated WDM system by propagating CSRZ pulse train through single mode fiber in the presence of fiber attenuation, dispersion and nonlinearities. Split-step Fourier model of optical communications is used for these investigations. These results clearly indicate that optimum selection of the dispersion map leads to achieve the higher Q-value and hence better system performance, even at higher launch power. The work shows the significance of optimization of dispersion map to enhance the performance of practical WDM systems at higher data rate.

REFERENCES

[1] G. Keiser, Optical Fiber Communications, 5th ed. McGraw Hill, 2013.
[2] J. A. Buck, Fundamentals of Optical Fibers, Canada: John Wiley & Sons, 2004.
[3] J. M. Senior, Optical Fiber Communications: Principles and Practices, Boston: Prentice-Hall, 2009.
[4] G. P. Agrawal, Nonlinear Fiber Optics, 5th ed., Academic Press, New York, 2012.
[5] T. Murakawa, and T. Konishi, "Pulse-by-pulse near 800-nm band power stabilization using all-optical limiter based on self-phase modulation", Optics Communications, vol. 344, pp. 134-138, June 2015.
[6] K. A. Taher, S. P. Majumder, and B. M. Azizur Rahman, "Performance of multilevel modulation formats in 92 Gb/s systems in the presence of PMD and nonlinear effects", *International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, Dhaka, Bangladesh, 2014.
[7] S. Cui, C. Qiu, et. al., "Simultaneous chromatic dispersion monitoring and optical modulation format identification utilizing four wave mixing", Optics Communications, vol. 354, pp. 59-65, November 2015.
[8] M. I. Hasan, S. M. A. Razzak, and M. S. Habib, "Design and characterization of highly birefringent residual dispersion compensating photonic crystal fiber", Journal of Lightwave Technology, vol. 32, pp. 4578–4584, 2014.
[9] S. T. Ahmad, and K. P. Kumar, "Normalized LMS filter for chromatic dispersion compensation in single polarization 40 Gbps CO-OFDM system in presence of timing and frequency offsets", Optik - International Journal for Light and Electron Optics, vol. 127, pp. 668-672, January 2016.
[10] T. Konishi, T. Murakawa, et. al. , "Flexible OFDM-based access systems with intrinsic function of chromatic dispersion compensation", Optical Fiber Technology, vol. 26, Part A, pp. 94-99, December 2015.
[11] N. Ahmed, M. I. Hayee, and Q. Zhang, "Electronic post-compensation of fiber nonlinearity for 40 Gb/s WDM systems," Joint IEEE and OSA's Journal of Optical Communications and Networking, vol. 2, pp. 456–462, 2010.
[12] M. Rajabvand, and F. Behnia, "In-fiber wavelength-selective modulation of WDM channels with performance analysis", Journal of Lightwave Technology, vol. 28, pp. 141-147, 2010.
[13] Sharma, A. Kakkar, and A. Sachdeva, "Optimized WDM network with consideration of lesser blocking probability & shortest path selection", *IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, India, pp. 1-6, 2012.
[14] D. D. Le, F. Zhou, M. Molnar, "Minimizing blocking probability for the multicast routing and wavelength assignment problem in WDM networks: Exact solutions and heuristic algorithms", IEEE/OSA Journal of Optical Communications and Networking, vol. 7, pp. 36–48, 2015.
[15] A. Yang, X. Li, A. Xu, and D. Wu, "Combined impact of group velocity dispersion, Kerr effect and polarization mode dispersion in optical fiber," Optics Communications, vol. 214, pp. 133-139, Dec. 2002.
[16] M. Nouri, M. G. Shayesteh, and N. Farhangian, "Chromatic dispersion and nonlinear phase noise compensation based on KLMS method", Optics Communications, vol. 351, pp. 149-154, September 2015.

[17] M. Y. Hamza, N. Akhtar, N. Sarwar, and S. Yang, "Investigations of chirped tan-hyperbolic pulse propagation behavior in the presence of fiber Loss, dispersion and self-phase modulation of single mode fiber", *Proceedings of 34th International Conference on Telecommunications* and Signal Processing (TSP), Budapest, Hungary, August 18-20, 2011.

[18] M. Y. Hamza, S. Saeed, N. Sarwar, and S. Yang, "Investigations for the evolution behavior of cos-Gauss pulse in dispersion dominant regime of single mode optical fiber", Engineering Letters, vol. 22, issue 2, pp. 83-86, Hong Kong, June 2014.

[19] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," Journal of Lightwave Technology, vol. 21, pp. 61-68, January 2003.

[20] M. Y. Hamza, N. Akhtar, N. Sarwar, and S. Yang, "Evolution behavior of chirped tan-hyperbolic pulse through single mode fiber in the simultaneous presence of fiber loss, dispersion and self-phase modulation", Telecommun. Syst., vol. 55, pp. 451-459, Springer Science + Business Media, New York, USA, April 2014.

[21] M. Secondini, D. Marsella, and E. Forestieri "Enhanced split-step Fourier method for digital back propagation", *European Conference on Optical Communication (ECOC)*, France, Sept. 2014.

[22] J. Shao, X. Liang, and S. Kumar, "Comparison of split-step Fourier schemes for simulating fiber optic communication systems", IEEE Photonics Journal, vol. 6, Article #: 7200515, August 2014.

[23] S. Balac, and F. Mahé, "An Embedded Split-Step method for solving the nonlinear Schrödinger equation in optics", Journal of Computational Physics, vol. 280, pp. 295-305, 2015.

eU-OFDM Based Multiple Access for Visible Light Communication Networks

Gerard Djengomemgoto[1], Omer Narmanlioglu[1], Refik Caglar Kizilirmak[2] and Murat Uysal[1]

[1]Department of Electrical and Electronics Engineering, Ozyegin University, Istanbul, Turkey, 34794
E-mail: omer.narmanlioglu@ozyegin.edu.tr
[2]Department of Electrical and Electronics Engineering, Nazarbayev University, Astana, Kazakhstan, 010000
E-mail: refik.kizilirmak@nu.edu.kz

***Abstract*—In this paper, we propose a novel protocol for cooperative multiple access in visible light communication (VLC) networks. The proposed protocol exploits the features of enhanced unipolar orthogonal frequency division multiplexing (eU-OFDM) scheme in order to enable user terminals to sequentially access the multi-hop network. The signal-to-noise ratios (SNRs) for user terminals within the network are derived and the individual bit-error-rate (BER) expressions are presented. We evaluate the BER performances of each user in the network through Monte Carlo simulation. The theoretical results validates the simulation results.**

***Index Terms*—eU-OFDM, multiple access, multi-hop relaying, visible light communication.**

I. Introduction

The recent advances and breakthroughs in solid-state lighting have led to the development and deployment of energy efficient light emitting diodes (LEDs) with longer luminary lifespan compared to the conventional incandescent light bulbs. LEDs have revolutionized the field of illumination technology and they have also sparked tremendous research effort in the area of visible light communication (VLC) [1]. VLC is a recent low-cost, secure and energy efficient technology with unregulated spectrum that uses light for high speed data transmission without interfering with existing radio frequency (RF) based systems. Due to spectrum scarcity of the conventional RF together with the higher demand for wireless services, VLC with its distinct features and great prospects, has been proven to be one of the vital and ultimate complementary solutions to the existing RF technology for ubiquitous wireless connectivity [2].

In a typical VLC system, intensity modulation and direct detection (IM/DD) schemes are used where information is encoded with the light intensity while photodetectors (PDs) handle the signal reception. Under this scheme, the information waveform needs to be real and positive valued. These constraints limit the usage of conventional complex or bipolar modulation schemes used in RF-based systems. Initial works in VLC considered simple pulse modulation techniques such as on-off keying (OOK) or pulse position modulation (PPM) [3]. Recent studies [4]–[7] have adopted orthogonal frequency division multiplexing (OFDM), which has been widely deployed in RF-based wireless communication systems due its salient advantages such as higher spectral efficiency and stronger immunity to multipath fading. Conventional OFDM structure, however, cannot be directly applied to VLC due to the IM/DD constraints. Therefore, different optical OFDM systems such as asymmetrically clipped optical OFDM (ACO-OFDM) [4], Flip-OFDM [5], DC biased optical OFDM (DCO-OFDM) [6] and enhanced unipolar OFDM (eU-OFDM) [7] have been proposed. It is important to note that, all these optical OFDM techniques generate real-value signals by imposing the Hermitian symmetry property on the complex symbol vector at the input of $N-$point inverse discrete Fourier transform ($N-$IDFT).

In DCO-OFDM [6], frame structure is formed by rearranging the $N/2-1$ data symbols to satisfy Hermitian symmetry. In this optical OFDM scheme, a DC-bias is added to the output of $N-$point IDFT to get unipolar signal at the input of LEDs. The additional DC-bias makes this scheme inefficient in terms of power consumption. In ACO-OFDM [4], the complex data symbols are carried only on the odd indexed sub-carriers. Based on this structure, for a signal vector to the input of $N-$point IDFT, only $N/4$ data symbols are encoded. The negative part of the output of IDFT is then clipped without affecting the data, and the clipped signal is then used for transmission through the optical system. Similarly, the frame construction for Flip-OFDM [5] is formed by encoding $N/2-1$ data symbols. After $N-$point IDFT process, the resulting data vector undergoes two processes in order to produce two positive data sub-frames. The first sub-frame is obtained by setting all the negative components of IDFT output vector to zeros. The second sub-frame is formed by flipping (inverting the polarity) of IDFT output vector prior to clipping the signal. The two sub-frames are consecutively transmitted through the IM/DD system. Since the encoded $N/2-1$ data symbols in the original signal is transmitted in two different sub-frames, Flip-OFDM presents almost similar spectral efficiency (SE) and the same bit error rate (BER) performance as ACO-OFDM scheme. DCO-OFDM, on the other hand, provides the highest SE as compared to Flip- and ACO-OFDM.

In order to reach high SE levels as DCO-OFDM without the need of additional DC power, enhanced unipolar OFDM (eU-OFDM) has been proposed [7]. The eU-OFDM enables different depths (layers) that allows conveying independent

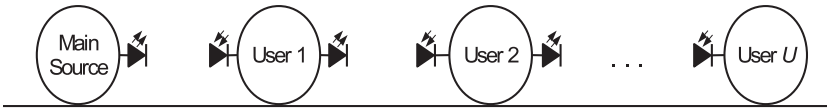


Figure 1: Considered network topology.

D1:	Neg. x_{14}	CP	Pos. x_{14}	CP	Neg. x_{13}	CP	Pos. x_{13}	CP	Neg. x_{12}	CP	Pos. x_{12}	CP	Neg. x_{11}	CP	Pos. x_{11}	CP
D2:	Neg. x_{22}	CP	Neg. x_{22}	CP	Pos. x_{22}	CP	Pos. x_{22}	CP	Neg. x_{21}	CP	Neg. x_{21}	CP	Pos. x_{21}	CP	Pos. x_{21}	CP
D3:	Neg. x_{31}	CP	Neg. x_{31}	CP	Neg. x_{31}	CP	Neg. x_{31}	CP	Pos. x_{31}	CP	Pos. x_{31}	CP	Pos. x_{31}	CP	Pos. x_{31}	CP

Figure 2: eU-OFDM frame structure.

streams of data in a single frame.

Furthermore, relaying techniques have also been proven to improve the performance of VLC systems [8]–[11]. In this paper, we present a new multiple access solution based on eU-OFDM scheme for the networks with line topology. In the considered multiple access scenario, multiple users are located sequentially and the data is forwarded from one user to another with the multi-hop amplify and forward (AF) relaying protocol in which each independent stream of eU-OFDM is allocated to each user. In respect to this, signal-to-noise-ratio (SNR) and BER derivations are provided for each user. We further verify our analytical results with Monte Carlo simulations.

The remainder of the paper is structured as follows. In Section II, the adopted VLC channel model is briefly described. Section III and IV present the system model for eU-OFDM based multiple access VLC networks and the BER analysis, respectively. The numerical results are presented in Section V. Section VI concludes the paper.

Notation: $(.)^*$ and $[.]^{\mathrm{T}}$ denote complex conjugate and transpose, respectively. $\mathbf{E}\{.\}$ denotes the statistical averaging and $Q(.)$ is the tail probability of standard normal distribution.

II. Channel Model

The multiple access network topology used in this study is depicted in Fig. 1. It is assumed that the main source node and U users are located sequentially and the transmission is based on line-of-sight (LoS) link. The flat response of a VLC link is given by [12]

$$h = \left\{ \begin{array}{cc} \frac{a+1}{2\pi d^2}\cos^a(\phi)\cos(\psi) & 0 \leq \psi \leq \Psi_{\frac{1}{2}} \\ 0 & \psi \geq \Psi_{\frac{1}{2}} \end{array} \right\} \quad (1)$$

where ϕ denotes the irradiance angle with respect to the LED axis, ψ is the angle of incidence with respect to PD axis, d is the distance between LED and PD, $a = -\ln(2) \Big/ \ln\left(\cos\left(\Phi_{\frac{1}{2}}\right)\right)$, the LED semi-angle $\Phi_{\frac{1}{2}}$ is set to 60°, the field-of-view (FOV) semi-angle of the photodetector $\Psi_{\frac{1}{2}}$ is assumed to be 60°.

III. System Model

We consider eU-OFDM for the proposed system structure. Let $s_1\ s_2\ ...\ s_{N/2-1}$ with the average energy of E, be the $M-$ary quadrature amplitude modulation (QAM) or phase shift keying (PSK) complex data symbols to be transmitted through the VLC system. The size of the modulated data symbols is similar to that of U-OFDM $(N/2-1)$. Thus, prior to IDFT of size N, the stream of modulated complex data symbols is subjected to Hermitian symmetry in order to obtain real-value signals. The rearranged data structure is written as

$$\mathbf{X} = \left[0\ s_1\ s_2\ s_3\ ...\ s_{N/2-1}\ 0\ s^*_{N/2-1}\ ...\ s^*_2\ s^*_1\right]^{\mathrm{T}}. \quad (2)$$

The real-valued signal resulting from the $N-$point IDFT sequence is defined as

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi nk}{N}}. \quad (3)$$

The U-OFDM data sequences are then used to build the eU-OFDM frame prior to prepending the data with cyclic prefix (CP) whose length is denoted by N_{CP}. For a given maximum depth of U (with $d = 1, 2, ..., U$), the eU-OFDM frame is formed by summing the data arrays from U layers. The first layer of data (first depth) follows the structure of U-OFDM. For the subsequent depths $(d > 1)$, each U-OFDM data sequence at depth d is duplicated 2^{d-1} times and scaled by a factor of $1/\sqrt{2^{d-1}}$ in order to keep the energy per symbol constant. The data sequences from all depths undergo a data array summation and the resulting data vector is the eU-OFDM symbol. This data symbol goes through parallel to serial conversion and digital-to-analog conversion prior to transmission via the optical link. As an illustration, the frame structure of eU-OFDM for a maximum depth of three is shown in Fig 2. By setting the spectral efficiency (SE) of the first to 1 bit/s/Hz, then the SEs of second and third depths become 0.5 and 0.25 bit/s/Hz, respectively.

For an eU-OFDM of a maximum depth of U, let $x(t)$ denote the resulting time domain signal after summation of all depths. E_T denotes its energy and is given by (4) (see the top of the next page). where Θ is the probability density function of standard normal distribution and $x_d(t)$ is the transmitted signal in the d^{th} depth [7].

In order to serve U users with proposed system architecture, an eU-OFDM frame of maximum depth U have to be considered. The encoded data from the first depth are intended for the first user terminal, data from second depth is to be transmitted

$$E_T = \mathbf{E}\left\{\left(\sum_{d=1}^{U} x_d(t)\right)^2\right\} = \frac{E(N-2)}{2N}\left(2 - \frac{1}{2^{U-1}} + 4\Theta^2(0)\sum_{d_1=1}^{U}\sum_{\substack{d_2=1 \\ d_1\neq d_2}}^{U}\frac{1}{\sqrt{2^{d_1+d_2}}}\right) \quad (4)$$

to second user and so on, and the last user receives data from U^{th} depth. At the first terminal, the received signal can be written as

$$y_1(t) = Rh_1x(t) + n_1(t) \quad (5)$$

where R is the responsivity of photodetector (A/W), $n_1(t)$ is additive white Gaussian noise (AWGN) with zero mean and σ_N^2 variance at first user node and h_1 is the VLC flat channel response from main source to the first user node. The detection of data intended for this node is straightforward. The decision process for each OFDM frame starts right after getting both the positive and negative sequences. The combining process, which is done through the subtraction of the negative sequence from the positive parts, consequently leads to noise enhancement by a factor of two, however, the effects of higher depths are cancelled out. After serial-to-parallel conversion, cyclic prefix removed and DFT process, the signal can be decided optimally with *Maximum Likelihood* (ML). Under the assumption of proper and successful detection of symbols, its interfering contribution in the received data $y_1(t)$ is removed after passing the estimated data through the estimated channel. The data to be amplified and forwarded is given by

$$\hat{y}_1(t) = y_1(t) - Rh_1\hat{x}_1(t). \quad (6)$$

The positive parts and negative parts carrying the same information of the second depth are summed and moved to first depth, the third depth data form the second depth in the new frame and so on. A magnitude scaling of $\sqrt{0.5}$ is applied for the conservation of the energy during this transformation in eU-OFDM frame structure and the signal is amplified with a factor of G_1 in order to set the power level to E_T. At the second user terminal, the received signal is given by

$$y_2(t) = RG_1h_2\hat{y}_1(t) + n_2(t) \quad (7)$$

where $n_2(t)$ is additive white Gaussian noise (AWGN) with zero mean and variance of σ_N^2 in the second node and h_2 is VLC flat channel response between first and second users. The similar demodulation and detection processes used at first user terminal are applied for the detection and estimation of data for this second user. After subtraction of estimated symbols for user from $y_2(t)$, the resulting new data is then amplified by a factor of G_2 and forwarded to next user terminal. This data undergoes also the eU-OFDM frame reconstruction similar to that of the previous terminal. This process continues until all the users terminals receives successfully their respective data. In general, received signal at i^{th} terminal can be written as

$$y_i(t) = RG_{i-1}h_i\hat{y}_{i-1}(t) + n_i(t) \quad (8)$$

where $\hat{y}_{i-1}(t) = y_{i-1}(t) - Rh_{i-1}\hat{x}_{i-1}(t)$ and G_{i-1} denotes the amplification factor to scale the signal energy to E_T at $i-1^{th}$ user node and $n_i(t)$ denotes the noise term with the same statistics at i^{th} PD.

IV. SNR and BER Analyses

Under the assumption that the symbols on each depth are perfectly estimated, the generalized SNR for i^{th} user node can be written as

$$\mathrm{SNR}_i = \frac{R^{2i}E\prod_{k=1}^{i}h_k^2G_{k-1}^2}{2\sigma_N^2(1+\beta_i)} \quad (9)$$

where β_i is equal to $\alpha_i + \alpha_i\beta_{i-1}$ in which $\alpha_i = R^2G_{i-1}^2h_i^2$ for $i \in \{2,...,U\}$ and 0 for $i \in \{0,1\}$ with initial definitions of $\beta_0 = 0$ and $G_0 = 1$. Relatively, BER [13] can be written as (10). As the SNR increases, the BERs on previous nodes decrease and the performances of all depths converge to analytical results [7]. Therefore, (9) and (10) can be used to estimate BER of eU-OFDM system at high SNR values.

V. Numerical Results

In this Section, we present numerical results to confirm the analytical SNR and BER expressions. In simulations, σ_N^2, R, N and N_{CP} are set to -35 dBm, 0.54 A/W, 64 and 4, respectively. In all the simulations, we assume that perfect channel state information is available at all nodes. We consider a scenario including one main source and three users, and the distances between them are set at 2 m.

In Fig. 3, we present BER results obtained by both theoretical and Monte Carlo simulations. 16-QAM is selected as modulation scheme for each depth. Therefore, the throughput gradually decreases for each node. It can be seen that since interference from higher depths are perfectly cancelled out and BER results are matched with analytical expression for the first user. In order to get 10^{-3} BER, 13.12 dBm eU-OFDM signal power is required. This level increases to 15.45 dBm for the second user and 16.32 dBm for the third user. The increment of the required power level is due to both noise amplification in each node and channel attenuation. However, the differences between the levels gradually decrease because each node extracts its own information and the remaining data is transmitted with higher energy with respect to previous one. Additionally, it should be noted that simulation and theoretical results for the second and third users converge at high SNR since incorrect decision in previous users leads imperfect interference cancellation for the following nodes.

$$\text{BER} \approx \left\{ \begin{array}{cc} Q\left(\sqrt{2\text{SNR}}\right) & , \quad 2-\text{PSK} \\ \frac{2\left(\sqrt{M}-1\right)}{\sqrt{M}\log_2\sqrt{M}} Q\left(\sqrt{\frac{3\text{SNR}}{(M-1)}}\right) & , \quad \text{square}-M-\text{QAM} \\ \frac{2}{\log_2(U\text{x}J)}\left[\frac{U-1}{U} Q\left(\sqrt{\frac{6\text{SNR}}{U^2+J^2-2}}\right) + \frac{J-1}{J} Q\left(\sqrt{\frac{6\text{SNR}}{U^2+J^2-2}}\right)\right] & , \quad \text{rectangular}-M=U\text{x}J-\text{QAM} \end{array} \right\} \quad (10)$$

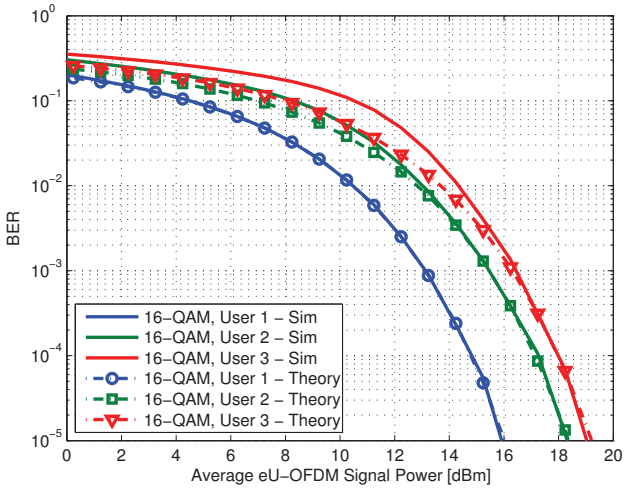


Figure 3: BER performances of three users when 16-QAM is deployed in each stream.

In Fig. 4, we present BER results for a scenario in which all users have equal throughput. For this reason, we set the modulation schemes for the first, second and third users to 2-PSK, 4-QAM and 16-QAM, respectively. At the target BER of 10^{-3}, the required signal power for the first user is 3.4 dBm. This value increases to 8.7 dBm and 16.3 dBm for second and third users because of deployment of higher modulation orders.

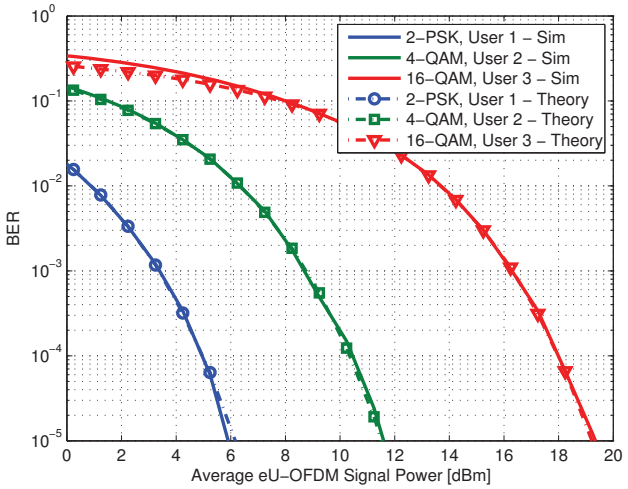


Figure 4: BER performances when 2-PSK, 4-QAM and 16-QAM are used for the first, second and third users.

VI. Conclusion

A multiple access protocol for eU-OFDM based VLC systems has been proposed. For the considered network topology, we have derived the SNR and BER for each node in the network. For a given number of users, we have demonstrated that the subsequent nodes present BER performance with respect to their preceding user terminal when a 16-QAM is employed for all users. Consequently, the throughput is reduced by half for each subsequent user terminal. In order to maintain the throughput at the same level for all users, 2-PSK, 4-QAM and 16-QAM can be used respectively for the first, second and third users, however, it leads the increment of the difference between required signal energy level for each user to achieve the same BER target.

References

[1] S. Arnon, J. Barry, G. Karagiannidis, R. Schober, and M. Uysal, *Advanced optical wireless communication systems*. Cambridge university press, 2012.

[2] S. Wu, H. Wang, and C. H. Youn, "Visible light communications for 5G wireless networking systems: from fixed to mobile communications," *IEEE Network*, vol. 28, pp. 41–45, Nov 2014.

[3] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 100–107, Feb 2004.

[4] J. Armstrong and A. J. Lowery, "Power efficient optical OFDM," *Electronics Letters*, vol. 42, pp. 370–372, March 2006.

[5] N. Fernando, Y. Hong, and E. Viterbo, "Flip-OFDM for Unipolar Communication Systems," *IEEE Transactions on Communications*, vol. 60, pp. 3726–3733, December 2012.

[6] J. Armstrong, "OFDM for Optical Communications," *Journal of Lightwave Technology*, vol. 27, pp. 189–204, Feb 2009.

[7] D. Tsonev, S. Videv, and H. Haas, "Unlocking Spectral Efficiency in Intensity Modulation and Direct Detection Systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1758–1770, Sept 2015.

[8] R. C. Kizilirmak and M. Uysal, "Relay-assisted OFDM transmission for indoor visible light communication," in *Communications and Networking (BlackSeaCom), 2014 IEEE International Black Sea Conference on*, pp. 11–15, May 2014.

[9] O. Narmanlioglu, R. C. Kizilirmak, and M. Uysal, "Relay-assisted OFDM-based visible light communications over multipath channels," in *2015 17th International Conference on Transparent Optical Networks (ICTON)*, pp. 1–4, July 2015.

[10] R. C. Kizilirmak, O. Narmanlioglu, and M. Uysal, "Relay-Assisted OFDM-Based Visible Light Communications," *IEEE Transactions on Communications*, vol. 63, pp. 3765–3778, Oct 2015.

[11] O. Narmanlioglu, R. C. Kizilirmak, F. Miramirkhani, and M. Uysal, "Cooperative Visible Light Communications," in *Optical Wireless Communication An Emerging Technology*, ch. 15, Springer, 2016.

[12] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, 1997.

[13] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, pp. 1074–1080, Jul 2002.

Analysis of Applicability of Open Single Sign-On Protocols in Distributed Information-Computing Environment

Lazarev S.A.
Belgorod State University
Belgorod, Russia
lazarev_s@bsu.edu.ru

Demidov A.V., Volkov V.N., Stychuk A.A., Polovinkin D.A.
Orel State University
Orel, Russia
a.demidov@ostu.ru, vadimvolkov@list.ru, stichuk@yandex.ru, dolovinkin@gmail.com

***Abstract*—This article contains analysis and comparison of different standards that implement Single Sign-On technology in context of their applicability to constructing security infrastructure of a distributed information-computing environment. Various non-proprietary protocols responsible for different aspects of secure information exchange were examined for creating a single authentication and authorization system in heterogeneous environment presented by a distributed network of portals.**

***Index Terms*—distributed information-computing environment, information access control management, authentication, authorization, network of portals, single sign-on.**

I. Introduction

The concept of building a security infrastructure of a distributed information-computing environment (DICE) presented as a network of enterprise portals is to analyze and design a solution based on open technologies and protocols that allow for secure authentication and authorization, while ensuring the required level of security of information exchange through the Internet.

A network of portals as a mechanism having a single point of entry and providing a unified access control policies for users and administrators is described in [1-2] and represents a set of access control units combined into a single network with a portal network management center (PNMC). As part of the PNMC a unified policy for controlling information exchange has been implemented, including the ability to secure authorized access to information and computational resources throughout the network, and a single user session management mechanism that provides the necessary level of security. That, in turn, leads to the solution of providing safe, reliable and efficient interaction of distributed system of information flow between users and server components. For this purpose it is necessary to analyze existing authentication protocols and to ensure safety of information exchange.

Nowadays many authentication and authorization protocols have been developed and are actively used, but not all of them are non-proprietary. Main objective of this article is to analyze authentication and authorization protocols, that is OpenID, OAuth, OpenID Connect, Single Sign-On based on Security Assertion Markup Language (SAML) [3-14] and their applicability and ease of use in a single authentication and authorization system for a DICE.

II. Authentication and authorization protocols

A. Single Sign-On

Single Sign-On is a technology that provides a single entry point to the corporate network, while giving user the ability to move from one portal to another without re-authentication and using portals as the preferred mechanism of interaction with the end user [4]. The main purpose of this concept is to minimize the number of passwords that the user has to remember and the number of authentication acts performed by them for a specific time period. This is a centralized technology, because one authentication server is used by all the portals within a corporate network [5, 6].

There are two technically possible variants of implementing Single Sign-On in a heterogeneous software environment:

- Using specialized software that intercepts authentication requests sent by the various services of web portals to user and automatically submits their credentials, such as account name and password. Since the authentication data are stored in the Single Sign-On database, the interception is not noticeable to the end user;
- Integration with authentication process of other software.

The main advantages of this solution are:

- Reduced amount of authentication data required to work with corporate network services;
- Reduced time needed for authentication on multiple portals of the corporate network.

The main disadvantage of this concept is the risk of the single password being acquired by a third party, which could lead to unauthorized access to the entire working environment.

All the following authentication and authorization protocols described in this article, OpenID, OAuth, SSO based on

SAML, are basically different approaches to the implementation of this technology.

B. OpenID decentralized authentication system

OpenID is an open decentralized system which allows the user to use a single account for authentication on multiple portals that are not connected with each other.

The main purpose of using OpenID [9] is to simplify the user's registration, achieved by a decentralized single sign-on system, that allows one to use a single set of authentication credentials on many web portals, as well as provide the necessary level of protection against unauthorized access.

OpenID authentication flow is relatively simple. A user opens a desired network resource (Relying Party, RP) using a web-browser with the intention to perform a specific action. If OpenID is not supported by the RP the user is prompted to register an account, that implies creating a new set of authentication credentials. Otherwise, OpenID authentication is offered as an alternative to registration (OpenID support is commonly indicated by the OpenID logo on login page).

On the next step, the user enters their OpenID - address, the address of the user's personal webpage on the server of their OpenID provider (OP). For example, if the user has received their OpenID on www.example.com, it will look like "username.example.domain". Further authentication process is shown in Fig. 1.

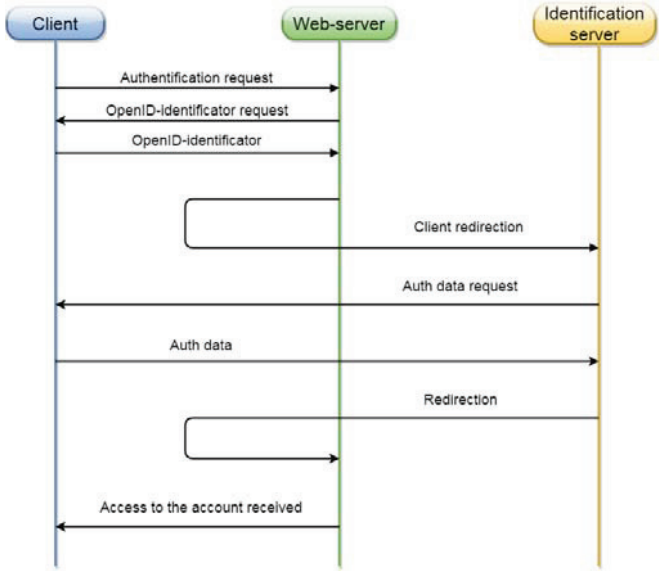


Fig. 1. OpenID authentication process

The main benefits of the OpenID authentication protocol are:

- OpenID is a convenient way to authenticate on web portals that support it.
- A user that has OpenID does not need to remember multiple passwords for different web portals and provide them any additional data, such as an email-address, the only set of credentials they need to know are their OpenID address and their password.
- After authenticating on OP website once, re-entering your OpenID password is not required if the session is kept active.
- Modules for integrating OpenID authentication into your software are freely available for most major programming languages, which makes it easier to start using OpenID.

OpenID has gained major popularity during the second half of the 00s. However, now it is not used as widely.

C. OAuth authorization protocol

OAuth authorization protocol allows third parties to provide limited access to protected resources without the need for the user to provide them (third party) authentication credentials. OAuth 2.0 is an authorization protocol that allows one to give a service a right to access the resources on another service. This fact makes OAuth suitable to be used in heterogeneous web services environment [11]. The protocol eliminates the need to trust an application login and password, and also allows you to issue a limited set of rights, rather than all at once. It is very important to note that OAuth is not an authentication protocol, however it is possible to use it for pseudo-authentication by using some API that the service has gained access to after authorization to verify the user's identity.

OAuth 2.0 involves 3 parties: a user (resource owner), who intends to carry out some action on their resources, uploaded to a third party web portal (server) using the specific web service (client). OAuth authorization flow is the following:

1. The client sends the request to the server that contains the client ID, timestamp and a callback address to which the access token should be returned.

2. The server acknowledges the request and responds to the client by sending a request token and a part of the divided secret.

3. The client uses the request token to redirect the user to the server.

4. After receiving the token, the server prompts the user to confirm the client's access to resources (authorization), then the user is redirected back from the server to the client with an authorization code. Usually, before prompting, the server authenticates the user.

5. The client sends the authorization code to the server requesting access token.

6. The server acknowledges the request and responds to the client a new access token.

7. Using the access token the client accesses the server's resources.

8. The server acknowledges the request and provides the resources.

The flow is shown on Fig. 2.

Despite the fact that the protocol can be used for the pseudo-authentication, it is very important to understand that authentication does not occur and that the access token does not contain any information about its who it was issued to (or its audience) which means that it can be reused by another client. If the user authenticates on a fraudulent server, the latter will be able to re-use the token on another server, posing

as the user. This is a serious security issue; however, it is well-known. For example, Facebook uses a proprietary modified version of the protocol - Facebook Connect, in which all requests are signed by a special way, so that it is impossible to exploit this vulnerability.

It is worth mentioning that OAuth can be used in Enterprise-level cloud services, an example of such an implementation is given in [12, 13].

To sum up, OAuth is a simple authentication standard based on the basic principles of the Internet, which makes it

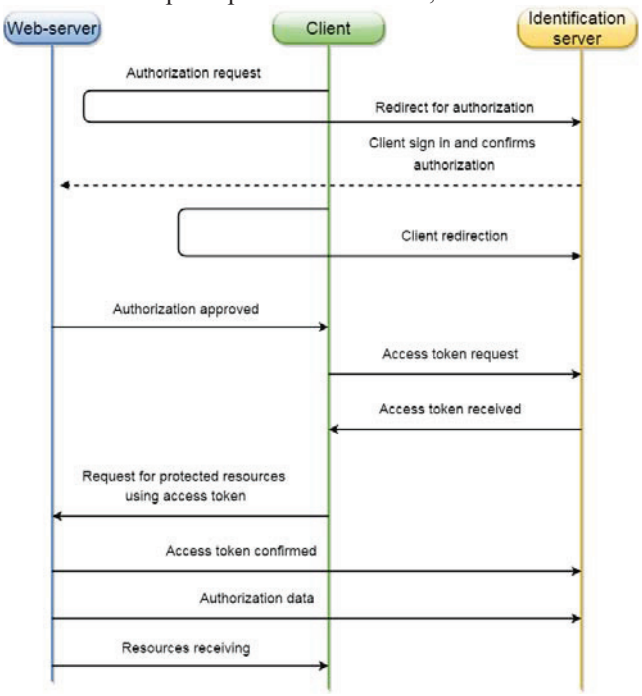


Fig. 2. OAuth authorization flow

possible to use authorization on almost any platform. It serves its purpose well, but there are serious flaws to be considered while using it as authentication protocol. The standard has gained wide popularity, many companies are using it in their services, including VK, LinkedIn, Facebook, and others.

D. SAML-based Single Sign-On

SAML (Security Assertion Markup Language) is a language based on XML that is used for security assertions. Open protocol based on the SAML was developed for authentication and authorization data exchange between secure domains, in particular between the identity provider (IdP) and the service provider (SP), solving the problem of providing pass-through authentication when using a web-browser [14].

SAML specification defines three roles: the main role - the user, IdP and SP. During SAML-authentication the service provider requests verification service from the Identity Provider. SP makes a decision about providing a service to the user based on an assertion made by identity provider.

SAML is mainly used for Web Browser Single Sign-On (SSO). The user requests a web resource that is protected by SAML Service Provider. Service Provider creates an authentication request to the identity provider and sends it via the user. The whole authentication flow is in Fig. 3.

SAML protocol uses security mechanisms such as TLS 1.0 or higher to ensure safety on the transport layer and XML-Signature and XML Encryption to ensure security at the message level.

However, this approach also has its flaws: a malicious service provider may forward the authentication request to another service provider. This flaw is a potential launching pad for cross-site scripting and cross-site request forgery attacks [6].

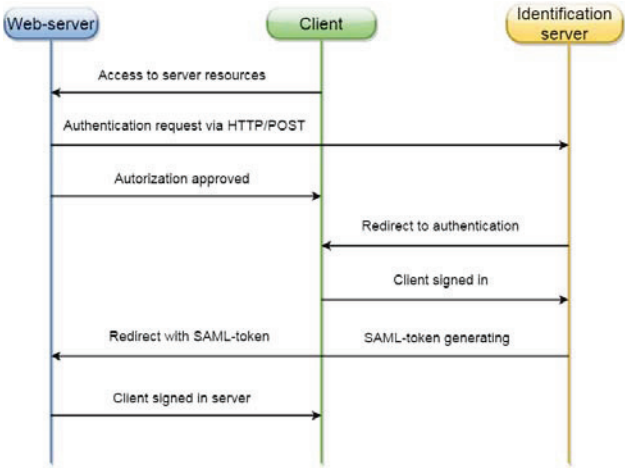


Fig. 3. Figure 3 - SAML authentication flow

SAML-based Single Sign On is widely used by many companies, including Google, which uses this protocol for a single authentication system for their services.

E. Authentication layer for OAuth - OpenID Connect

OpenID Connect was intended to address the underlying problems of OAuth protocol when it is used for pseudo-authentication - the lack of actual identity verification for user, since as noted earlier, the presence of the access token does not allow us to assert that the user is authentic. During the protocol's development one of the main objectives was to ensure compatibility with OAuth 2.0, so that developers can easily add OpenID Connect to their existing applications. Even though the standard is simple to use, it did not impact negatively on its security capabilities: the protocol provides a number of additional features that make it able to meet the high demands of enterprise-level systems.

The key difference between this protocol on OAuth - the presence of identity token that authenticates the user. After authorization the client application receives the identity token in the form of JSON web token (JWT). JWT is an open, industry standard method for representing claims securely between two parties. It supports a large number of signature algorithms and data encryption. Features of the identification token:

- Asserts the user's identity;
- Specifies issuing authority;
- Is generated for a particular audience.

Since this protocol is based on OAuth, the authentication flow is almost the same as OAuth authorization flow (Fig. 4).

Despite the novelty of this protocol, many companies have been certified for compliance with this standard, including Google, Microsoft, Ping Identity and PayPal.

OpenID Connect is already used by Google for the feature that allows user to sign in with their Google+ account. In addition to Google, OpenID Connect is also used by Gakunin (network of Japanese universities), Microsoft, Ping Identity and others.

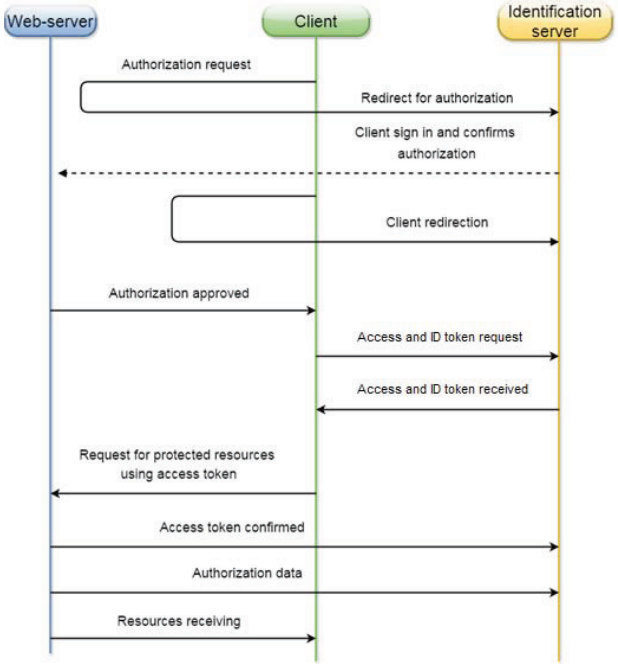


Fig. 4. OpenID Connect authentication flow

III. PROTOCOLS COMPARISON

The main results of this paper are recommendations about using protocols to ensure safety and reliability of a DICE.

Table 1 provides an analysis of the authentication and authorization protocols: OpenID, OAuth, SAML, OpenID Connect.

We should take a closer look at each criterion:

1) Main purpose

In general, all of these technologies can be used to solve and reach said goal. However, OAuth is primarily an authorization protocol. As mentioned earlier, when it is used for authentication, it allows attackers easily to compromise the process. An effective solution to this problem is OpenID Connect.

2) Interoperability

SAML has limited support for mobile applications, in addition, in cases of multiple identity providers it requires a complex broker service to be implemented. The rest of the technologies are open standards, and used quite frequently, which testifies to their high interoperability.

TABLE I. COMPARISON OF AUTHENTICATION AND AUTHORIZATION PROTOCOLS

Criterion	OpenID	OAuth	SAML	OpenID Connect
Main purpose	SSO	API authorization	SSO	Authentication layer for OAuth
Interoperability	High	High	Medium	High
Implementation complexity	Medium	Medium	Medium	Medium
Integration complexity	Low	Low	Low	Low
Various encryption algorithms support	No	No	Yes	Yes
Stable release	Yes	Yes	Yes	Yes

3) Implementation complexity

There are modules for most of major programming for each of the technologies which makes it easier to use them in various projects.

4) Integration complexity

Since this study was conducted in the framework of the development of an authentication system for DICE, it is reasonable to consider its deployment diagram shown in Fig. 5.

All the technologies except RADIUS would allow us to use only existing components. The use of RADIUS entail adding additional entities: RADIUS server, an additional database for it.

5) Various encryption algorithms support

OpenID and the OAuth, delegate encryption to the transport layer, where TLS can be used. SAML and OpenID Connect allow you to use a variety of encryption algorithms, including encryption in accordance with Russian GOST.

Summing up this comparison, OpenID Connect seems like the best choice for the given task. That is, it looks like the most promising option for use in the creation of a single authentication and authorization system for a DICE.

IV. CONCLUSION

In the process of analysis strengths and weaknesses of non-proprietary protocols were identified in context of their applicability for constructing a single authentication and authorization service for a DICE

In the future, it is suggested to use these results for:

- carrying out the development of a complex of new scientific and technical solutions in the field of DICE infrastructure software security based on non-proprietary protocols that will allow to simplify the use of DICE for end users, as well as the administration of the infrastructure, while providing a decent security level equal to security level of systems built using a public key infrastructure or even exceeding that level;
- carrying out research and development work on DICE security software.

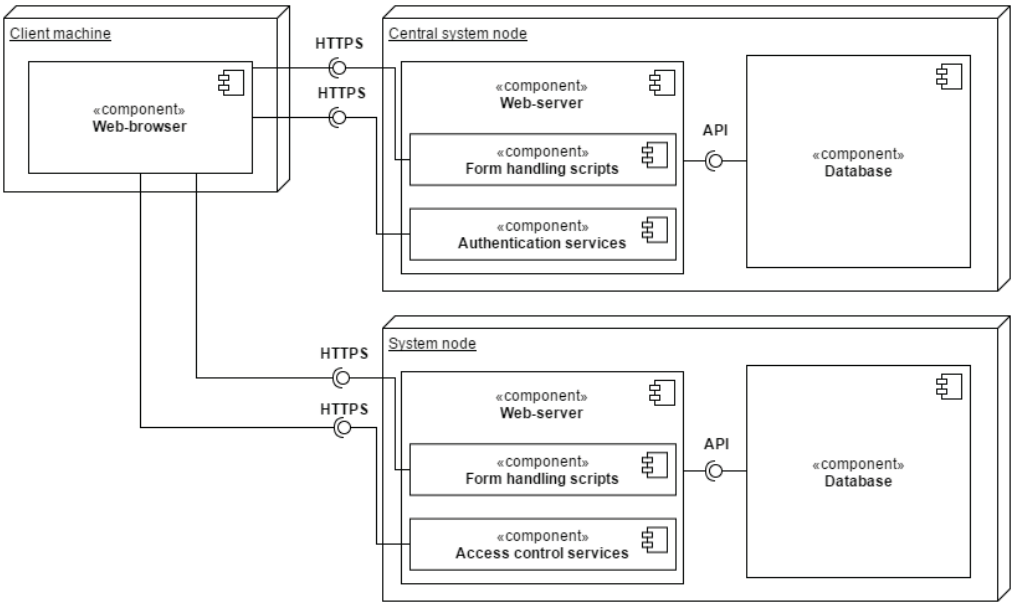


Fig. 5. DICE deployment diagram

ACKNOWLEDGMENT

This article was prepared with support of grant of President of Russian Federation № MK-5958.2016.9.

REFERENCES

[1] Lazarev S.A., Demidov, A.V., "The concept of construction of a control system of an information exchange in the network of corporate portals", Information Systems and Technologies, 2010, 4(60), pp. 123–129.

[2] Konstantinov I.S., Lazarev S.A., Mihalev O.V., Demidov A.V., Shateev R.V., "The development of infrastructure security for distributed information computer environment based on secured portal network", International Journal of Applied Engineering Research, 2015, 10 (17), pp. 38116-38120.

[3] Rigney, C., Rubens, A., Simpson, W., S. Willens, "Remote authentication dial in user service (RADIUS)", April 1997. Retrieved from: www.rfc-editor.org/rfc/rfc2865.txt.

[4] Tsyrklivich E., Tsyrklivich V., "Single sign-on for the internet: a security story", 2007, BlackHat USA, Las Vegas.

[5] Radha V., Hitha Reddy V., "A survey on single sign-on techniques", Procedia Technology, 2012, 2nd International Conference on Computer, Communication, Control and Information Technology (C3IT-2012), pp: 134-139.

[6] Armando A., Carbone R., Compagna L., Cuellar J., Pellegrino G., Sorniotti A., "An authentication flaw in browser-based single sign-on protocols: impact and remediations", 2013, Computers & Security, 33, pp. 41-58.

[7] Lazarev, S.A., Konstantinov I.S., Silaev P.P., "Safety mechanism for multi-factor authentication with digital access key use in closed virtual environment of distributed information resources", 2015, International Journal of Applied Engineering Research, 10 (24), pp. 44927-44923.

[8] Brands S., "The problem(s) with OpenID". Retrieved from: idcorner.org/2007/08/22/the-problems-with-openid/.

[9] Waters J.K., "OpenID updates identity spec", 2008. Retrieved from: web.archive.org/web/20080208155322/http://reddevnews.com/news/devnews/article.aspx?editorialsid=913.

[10] Sun S., Hawkey K., Beznosov K., "Systematically breaking and fixing OpenID security: formal analysis, semi-automated empirical evaluation, and practical countermeasures", Computers & Security, 2012, 31(4), pp. 465-483.

[11] Torroglosa-Garcia, E., Perez-Morales A.D., Martinez-Julia P., Lopez D.R., "Integration of the OAuth and web service family security standards", Computer Networks, 2013, 57(10), pp. 2233-2249.

[12] Noureddine, M., Bashroush M., "An authentication model towards cloud federation in the enterprise", Journal of Systems and Software, 2013, 86(9), pp. 2269-2275.

[13] Zhao, R., Yue C., "Toward a secure and usable cloud-based password manager for web browsers", Computers & Security, 2014, 46, pp. 32-47.

[14] Mishra P., Chopra D., Moreh J., Philpott R., "Differences between OASIS security assertion markup language (SAML) V1.1 and V1.0", 2003. Retrieved from: www.oasis-open.org/committees/download.php/3412.

Multi-objective Optimization Route Selection and Resource Allocation in Wireless Multi-Hop Cooperative Cognitive Radio Sensor Networks based on OFDMA

Ali Shahzadi
Department of Electronic
Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch,
Islamic Azad University
Tehran, Iran
shahzadi@iust.ac.ir

Fereshteh Atri Niasar
Department of Electronic
Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch,
Islamic Azad University
Tehran, Iran
Atri200867@gmail.com

***Abstract*---Following spectrum constraints cognitive radio is introduced, as a technology to eliminate the imbalance between the lack of spectrum and unused spectrum. On the other hand, use of sensors also become more widespread, as in the near future will be an integral part of human life. So network topology in this paper is based on a Wireless Multi-Hop cooperative Cognitive Radio Sensor Networks based on OFDMA.**

in the objective of optimizing different types of end-to-end performance. Aiming to solve it optimally, we first show that this problem of optimal resource allocation and hopes can be formulated as a convex optimization problem and identify its necessary and sufficient conditions. Based on this conclusion, KKT conditions apply to the issue and the optimal solution can be found at any phase. we propose *fmincon* function in MATLAB that uses Lagrangian duality theory. The scheme thus converges to a globally optimal solution.

Resource allocation and route selection is formulated as a multi-objective optimization problem. that one of its objectives is to find the optimal amount of hops and the second is considered to maximize the data transfer rate.also the effect of Interference Temperature constraints is reviewed.

***Index Terms*—Multi-objective optimization, Sensor networks, Cognitive Radio, Route Selection, Resource Allocation.**

I. INTRODUCTION

Following recent progresses in moot point of microelectromechanical systems, their cheap outlining, and up growth, all of them paved the way for more across the board acceptance of sensors and in a very near future, they will be indispensable part of human being life. In another word, wireless sensor networks are a kind of MANET networks holding more nodes and the big-league advantage of these networks is that they do not demand infrastructure. In [1] represents the general specifications of a sensor network together with their existing challenges. As the result of aggrandizement demand for technology, frequency spectrum and chain for the wide spread usage of higher data rate and spectrum, we need new plans for more productivity from spectrum. In [4] touches on this point that the reason for current problem of spectrum inadequacy is the strict

Regulations for access to spectrum; when authorized users are active, just a limited part of spectrum will use by them. For that reason, Radio Recognizer introduced for the first time as a technology to remove imbalance between spectrum shortage and not used spectrum. For the first time on 1999, JOSEPH MITOLA [2] used the concept and idea of Radio Recognizer in his Ph.D. thesis. As he suggested, Radio Recognizer can enhance individual wireless service through RKRL [3]. Among important Radio Recognizer standards, we can mention to IEEE802.22, this standard take advantage from OFDM method in physical layer. [5], [14]

Using multi hops for forwarding, embraces route selection integrationey challenge and resource allocation simultaneously. In [6] and [7] point to route robustness for choosing route in one multi hops cognitive radio network and show the superiority of the route selection suggested in comparison with route selection strategies in accordance with the existing rates, because the mentioned method takes one robustness threshold for route selection. Following that, some of this subset routes will choose and one spectrum will dedicate to each link existing in these routes holding the maximum system operational power.

Paper [6] offers a two-steps solution by going through inhomogeneous specifications for channel and channel dynamic for cognitive radio networks which simultaneously selects route and spectrum based on maximum operational power. Paper [7] proceeds with one algorithm holding several sentences chronological complex and it figures out that the acquired consequences are close to the optimum amount. Paper [9] formulates the major goal function through the definition for throughput, route selection issue and resource allocation.

The rest of paper organized as follows. Section II describes system model of OFDMA_ based multi-hop CR networks. In Section III, the theoretical framework for optimal cross-layer design provided. Then the centralized scheme proposed in Section IV. Following this, the performance of the proposed scheme investigated by numerical results in Section V. Finally, the conclusions summarized in Section VI.

II. System Model and Problem Formulation

Consider multihop adhoc CR network and orthogonal subchannels $\boldsymbol{N} = \{1,2,\dots,\mathrm{N}\}$. where every user equipped with a multi full-duplex radio for the enhancement of system connection capability. Potential routes show with $\boldsymbol{M} = \{1,2,\dots,M\}$,$\forall m \in \boldsymbol{M}$ and routes link with subset of $\forall V \in \boldsymbol{V}$,$\boldsymbol{V} = \{1,2,\dots,\mathrm{V}\}$.[8]

In furtherance of more investigation about Cognitive Radio effect and the first and second users' functions besides each other, we will Study over the first results with and without the companionship of the first user. For each $m \in \boldsymbol{M}$ we will draw the main graph based on network topology in which every network unit show with one node and direct links of Point to Point between them with links. Contention graph comes from the main graph. Every link in the main graph is modeled with one node in CG. We will have one link between nodes of CG if the main graph nodes stay in radio interference range. The Contention graph will define as, $G_{\mathrm{m}} = \mathrm{CG} = (V_{\mathrm{m}}, E_{\mathrm{m}})$ in which depicts the links along the route as vertex set , V_{m} and the contention relationship of the links as edges between the vertices, V_{m}. If there is not link between the two nodes in CG, those two nodes shape independent subset and links can simultaneously perform forwarding in one similar subchannel (IS). Maximal Independent Subset (MIS) is an IS that is not subset of any other IS. S_m denotes the MIS set of route m. Let τ_{ms}^n denote the occupancy of *n*th subchannel by *s*th MIS of route m, which follows:

$$\tau_{ms}^n \in \{0,1\}, \sum_{s\in S_m} \tau_{ms}^n \le 1 \ , \forall n \in N, m \in M \qquad (1)$$

$$w_{mv}^n = \sum_{s\in S_m} b_{ms}^n \tau_{ms}^n \le 1 \qquad (2)$$

w_{mv}^n is a linear transform and shows allocation of *n*th subchannel to link V from route m or in another word, allocation of n *s*th sub channel to link V. also in (2), b_{ms}^n is $\{0,1\}$ and its amount will be 1 when link V includes *s*th MIS of route m. transmission throughput, expressed as spectrum efficiency, R_m^v of link V which follows:

$$R_m^v = \sum_{n\in N} w_{mv}^n \ \log(1 + p_{mv}^n G_v^n) \qquad (3)$$

defines as this $G_{\mathrm{v}}^{\mathrm{n}} = g_{\mathrm{v}}^{\mathrm{n}}/\Gamma$ in which $g_{\mathrm{v}}^{\mathrm{n}}$ is channel-to-noise ratio and Γ denotes the SNR gap ,$\Gamma = (\frac{0.2}{BER} - 1)/1.5$.

$$\sum_{n\in N} w_{mv}^n \ p_{mv}^n \le \overline{p_v} \ , \ \Delta_v^n \le \ p_{mv}^n \ \le \nabla_v^n \qquad (4)$$

Pay attention that these variables will define separately for each route, $R_m^v, s\tau_{ms}^n, p_{mv}^n, w_{mv}^n$.

A. Inter-route System Architecture

To determine the route performance, first, we define problem target function together with its requirements and conditions, and then we can solve it as a convex optimization issue [11]. Target utility function (U_m), aims to maximize throughput rate and optimize hops quantity and it will define as follows:

$$U_m(R_m)_{NEW} = \frac{m}{M} + (R_{min})^k \sum_v (R_{mv}^n)^{-k} /k \qquad (5)$$

In which $\frac{1}{M}$, $(R_{min})^k$ are weight coefficient for target. pay attention that transmission delay in comparison with propagation delay and queuing delay, is superior delay, for that reason[12], [13], we tend to minimize target function with K=1.

B. Secondary Spectrum Sharing

In order to give the once over operation of the primery and secondery users besides each other we use IT model. Although IT model defined by FCC is an old model, however, we still consider IT as an effective approach for determination of interference with the first user. For modeling of IT restriction, the total for the received interference power for each first user receiver $I = \{1,2,\dots,I\}$ for the following channel shall be less than B_i^n. this condition expressed in equation (6). h_{vi}^n, is frequency channel gain of *v*th CR link to *i*th PU receiver, and B_i^n is PU user *i*'s accumulated interference threshold. The amount of B_i^n calculates based on watt and it is a multiple of interference temperature limit by Boltzmanns constant K = 1.38.7 * 10 -23 J/K and bandwidth. v_m^i is the same v_m^i for CR links in route m which interfere with PU.

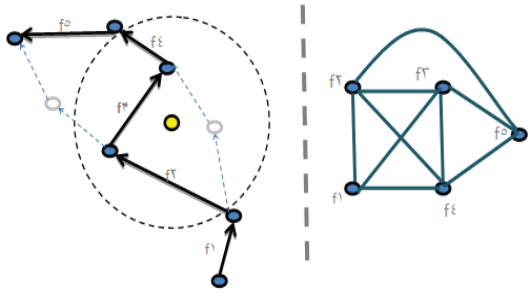


Fig. 1. subset for potential routes and Contention graph

Figure1 is a simple form example of the desired problem and as you see, the potential route for it is $\{f_1, f_2, f_3, f_4, f_5\}$. f_4, f_3, f_2 expose with the interference with the primery user, because the transmitter is close to the PU user and it stays in its interference radios. With regard to the definition of MIS subset and the given elaboration about CG, MIS subset is also $\{\{f_1, f_5\}, f_2, f_3, f_4\}$.

III. Optimal Cross-Layer Optimization

For the selection of the best route among several potential routes, we bring out one route selection factor like ρ_m. The main problem with supposing $M_v \triangleq \{\mathrm{m} \in \mathrm{M} : v \in V_m\}$ and transmission throughput (R_m^v) formulates as follows:

$$\min \sum_{\mathrm{m}\in\mathrm{M}} \rho_{\mathrm{m}} U_m(R_m)$$

s.t (7)

$$\tau_{ms}^n \in \{0,1\}, \ \sum_{s\in S_m} \tau_{ms}^n \le 1$$
$$\sum_{m\in M} \rho_{\mathrm{m}} \sum_{n\in N} w_{mv}^n \ p_{mv}^n \le \overline{p_v}$$
$$\sum_{m\in M} \rho_{\mathrm{m}} \sum_{v\in V_m^i} w_{mv}^n \ p_{mv}^n h_{vi}^n \le B_i^n$$
$$\Delta_v^n \le \ p_{mv}^n \ \le \nabla_v^n$$
$$\rho_m \in \{0,1\}, \sum_{m\in M} \rho_{\mathrm{m}} \le 1$$

To counterbalance of simplify optimization with changing variables of $m = \frac{M\ \tilde{m}}{\rho_m}$, $\tau_{ms}^n = \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}}$, $p_{mv}^n = \frac{\widetilde{Q_{mv}^n}}{w_{mv}^n \ \rho_{\mathrm{m}}}$ mix integer change to relax and equation (7) will rewrite as follows:

$$\min \sum_{\mathrm{m}\in\mathrm{M}} \rho_{\mathrm{m}} U_m(R_m) \qquad (8)$$

s.t

$$\frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} \in (0,1], \ \sum_{s\in S_m} \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} \le 1$$
$$\sum_{m\in Mv} \ \sum_{n\in N} \widetilde{Q_{mv}^n} \le \overline{p_v}$$

$$\sum_{m\in M}\ \sum_{v\in V_m^i} \widetilde{Q_{mv}^n}\, h_{vi}^n \le B_i^n$$

$$\Delta_v^n \le \frac{\widetilde{Q_{mv}^n}}{\rho_{\mathrm{m}}} / \left(\sum_{s\in S_m} b_s^v\, \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}}\right) \le \nabla_v^n$$

$$\rho_m \in (0,1],\ \sum_{m\in M} \rho_{\mathrm{m}} \le 1$$

Now by using Lagrangian method and using its coefficients in equation (8), we have:

$$L = \sum_{\mathrm{m}\in \mathrm{M}} \rho_{\mathrm{m}}\, U_m(R_m) + \sum_{m\in Mv} \alpha_v\ \sum_{n\in N} \widetilde{Q_{mv}^n} + \sum_{m\in M}\ \sum_{n\in N} \beta_{\mathrm{m}}^{\mathrm{n}} \sum_{s\in S_m} \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} + \gamma \sum_{\mathrm{m}\in \mathrm{M}} \rho_{\mathrm{m}} + \sum_{n\in N}\ \sum_{i\in I}\ \psi_{\mathrm{i}}^{\mathrm{n}} \sum_{m\in M}\ \sum_{v\in V_m^i} \widetilde{Q_{mv}^n}\, h_{vi}^n + \sum_{m\in M}\ \sum_{v\in V_m} \sum_{n\in N}\ (\varphi_{\mathrm{mv}}^{\mathrm{n}} - \phi_{\mathrm{mv}}^{\mathrm{n}})\, \frac{\widetilde{Q_{mv}^n}}{\sum_{s\in S_m} b_s^v\, \tilde{\tau}_{ms}^n} \qquad (9)$$

$\beta_{\mathrm{m}}^{\mathrm{n}}$،$\alpha_v$، γ،$\psi_{\mathrm{i}}^{\mathrm{n}}$،$\varphi_{\mathrm{mv}}^{\mathrm{n}}$and $\phi_{\mathrm{mv}}^{\mathrm{n}}$ are Lagrangian multipliers, Then KKT conditions use based on the independent variables of $\widetilde{Q_{mv}^n}$, $\tilde{\tau}_{ms}^n$:

$$\frac{\partial L}{\partial\, \widetilde{Q_{mv}^n}} = \frac{U'_m{}^v(R_m^v)\, G_v^n}{1+p_{mv}^n G_v^n} + \alpha_v + \sum_{i\in I_m^v}\ \psi_{\mathrm{i}}^{\mathrm{n}} h_{vi}^n - \frac{\phi_{\mathrm{mv}}^{\mathrm{n}}}{\rho_{\mathrm{m}}\, w_{mv}^n} + \frac{\varphi_{\mathrm{mv}}^{\mathrm{n}}}{\rho_{\mathrm{m}}\, w_{mv}^n} = 0 \qquad (10)$$

$$\frac{\partial L}{\partial\, \tilde{\tau}_{ms}^n} = \sum_{v\in V_m} \{U'_m{}^v(R_m^v) \left[\log(1+p_{mv}^n G_v^n) - \frac{p_{mv}^n G_v^n}{1+p_{mv}^n G_v^n}\right] + \left[\frac{\Phi\phi_{\mathrm{mv}}^{\mathrm{n}}}{\rho_{\mathrm{m}}\, w_{mv}^n} - \frac{\varphi_{\mathrm{mv}}^{\mathrm{n}}}{\rho_{\mathrm{m}}\, w_{mv}^n}\right] \times p_{mv}^n\} \times b_{ms}^v + \frac{\beta_{\mathrm{m}}^{\mathrm{n}}}{\rho_{\mathrm{m}}} = 0 \qquad (11)$$

$$\frac{\partial L}{\partial\, \tilde{\tau}_{ms}^n} = \lambda_{ms}^n + \frac{\beta_{\mathrm{m}}^{\mathrm{n}}}{\rho_{\mathrm{m}}} \begin{cases} > 0 & if\ \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} \notin (0,1] \\ = 0 & if\ \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} \in (0,1] \end{cases} \qquad (12)$$

In which λ defines as follows:

$$\boldsymbol{\lambda_{ms}^n} = \sum_{v\in V_m} \{\left[U'_m{}^v(R_m^v).\log(1+p_{mv}^n G_v^n)\right] + \left[\alpha_v + \sum_{i\in I_m^i} \psi_{\mathrm{i}}^{\mathrm{n}}\, h_{vi}^n\right] \times p_{mv}^n\} \times b_{ms}^v \qquad (13)$$

It comes at able to rewrite the above equation as follows and for more simplicity:

$$\lambda_{ms}^n = \sum_{v\in V_m}\ \delta_{mv}^n \times b_{ms}^v \qquad (14)$$

In which δ_{mv}^n is the same link weighted v and it defines as follows:

$$\delta_{mv}^n = \sum_{v\in V_m} \{\left[U'_m{}^v(R_m^v).\log(1+p_{mv}^n G_v^n)\right] + \left[\alpha_v + \sum_{i\in I_m^i} \psi_{\mathrm{i}}^{\mathrm{n}}\, h_{vi}^n\right] \times p_{mv}^n\} \qquad (15)$$

Therefore when dedicating the optimum subchannel for an optimizing problem MIS weighted follows this equation:

$$\begin{cases} \tau_{mx}^n = 1\,, & x = \arg\min_{s\in S_m} \sum_{v\in s}\ \delta_{mv}^n \\ \tau_{my}^n = 0, & y \ne x \end{cases} \qquad (16)$$

By using the above equation, we can find argument, which minimizes the phrase. The next condition Karush-Kuhn-Tucker (KKT) is dedicating zero to Lagrangian multipliers in conditions as follows:

$$\Phi_{\mathrm{mv}}^{\mathrm{n}} \left(\frac{\widetilde{Q_{mv}^n}}{\sum_{s\in S_m} b_s^v\, \tilde{\tau}_{ms}^n} - \Delta_v^n\right) = 0\,,\ \phi_{\mathrm{mv}}^{\mathrm{n}} \ge 0 \qquad (17)$$

$$\varphi_{\mathrm{mv}}^{\mathrm{n}} \left(\frac{\widetilde{Q_{mv}^n}}{\sum_{s\in S_m} b_s^v\, \tilde{\tau}_{ms}^n} - \nabla_v^n\right) = 0\,, \varphi_{\mathrm{mv}}^{\mathrm{n}} \qquad (18)$$

Therefore, for dedicating the optimum power we have:

$$p_{m\,v}^n = \left[-\frac{U'_m{}^v(R_m^v)}{\alpha_v + \sum_{\iota\in I_m^v} \psi_\iota^n\, h_{v\iota}^n} - \frac{1}{G_v^n}\right]_{\Delta_v^n}^{\nabla_v^n} \qquad (19)$$

For the selection of the best route, that is calculation of m** with regard to the restriction of ρ_{m}, we will talk about another parameter called M_m :

$$\frac{\partial L}{\partial\, \rho_{\mathrm{m}}} = \mathrm{M}_m + \gamma \begin{cases} > 0, & if\ \rho_{\mathrm{m}} \notin (0,1] \\ = 0 & if\ \rho_{\mathrm{m}} \in (0,1] \end{cases} \qquad (20)$$

$$\frac{\widetilde{Q_{mv}^n}}{\rho_{\mathrm{m}}} = p_{mv}^n \sum_{s\in S_m} \frac{b_{ms}^v\, \tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} \qquad (21)$$

$$\frac{\beta_{\mathrm{m}}^{\mathrm{n}}}{\rho_{\mathrm{m}}} = -\sum_{v\in V_m}\ \sum_{s\in S_m} \frac{b_{ms}^v\, \tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}}\ \lambda_{ms}^n \qquad (22)$$

For the calculation of μ_m we have:

$$\frac{\partial L}{\partial\, \rho_{\mathrm{m}}} = \frac{\partial}{\partial\, \rho_{\mathrm{m}}} \left(\sum_{\mathrm{m}\in \mathrm{M}} \rho_{\mathrm{m}}\, U_m(R_m) + \sum_{m\in M}\ \sum_{n\in N} \beta_{\mathrm{m}}^{\mathrm{n}} \sum_{s\in S_m} \frac{\tilde{\tau}_{ms}^n}{\rho_{\mathrm{m}}} + \gamma \sum_{\mathrm{m}\in \mathrm{M}} \rho_{\mathrm{m}}\right) = U_m(R_m) + \acute{U}_m(R_m) \sum_{n\in N}\ \sum_{s\in S_m}\ b_{ms}^v \tau_{ms}^n\ \log(1 + p_{mv}^n G_v^n) - \frac{1}{\rho_{\mathrm{m}}} \sum_{n\in N} \beta_{\mathrm{m}}^{\mathrm{n}} \sum_{s\in S_m} \tau_{ms}^n\,) + \gamma \qquad (23)$$

$$\boldsymbol{\mu_m} = U_m(R_m) + \acute{U}_m(R_m) U_m(R_m) - \frac{1}{\rho_{\mathrm{m}}} \sum_{n\in N} \beta_{\mathrm{m}}^{\mathrm{n}} \sum_{s\in S_m} \tau_{ms}^n \qquad (24)$$

$\boldsymbol{\mu_m}$ calculated with regard to the above relations and differentiate L based on ρ_{m}. As we expected it became a function of $\mathrm{U_m(R_m)}$, so the result for selection of the optimum route we will have:

$$\begin{cases} \rho_{\mathrm{x}} = 1\,, & \mathrm{x} = \arg\min_{m\in M} \mathrm{M}_m \\ \rho_{\mathrm{y}} = 0\,, & \mathrm{y} \ne \mathrm{x} \end{cases} \qquad (25)$$

IV. Algorithm development

In order to solve optimization problem for the target function, we used interior-point algorithm in fmincon function. Hessian function used as entry in this algorithm. When using hessian matrix, we will acquire a more precise and quicker solution for optimization with constraint.

Algorithm 1 *:Joint Route Selection and Resource Allocation*

$$X = fmincon\,(fun, x_0, A, b, \mathrm{A}_{eq}, \mathrm{B}_{eq}, \Delta_v^n, \nabla_v^n)$$

subject to:

$$\sum_{m\in M} \rho_{\mathrm{m}} \le 1$$

$$\sum_{m\in M} \rho_{\mathrm{m}} \sum_{n\in N} w_{mv}^n\ p_{mv}^n\ \le \overline{p_v}$$

$$\sum_{m\in M} \rho_{\mathrm{m}} \sum_{v\in V_m^i} w_{mv}^n\ p_{mv}^n h_{vi}^n \le B_i^n$$

$$\Delta_v^n \le\ p_{mv}^n\ \le\ \nabla_v^n$$

In which x_0 is the first amount of fun function. A and B are linear unequal. Minimum and maximum for x defines also as (LB, UB) and we will have answers between these two quantities. Aeq and Beq parameters are linear equals.

$$x = [\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, p_1^1, p_1^2, p_2^1, p_2^2, p_2^3, p_2^4, p_3^1, \dots, p_5^{13}, p_5^{14}]$$

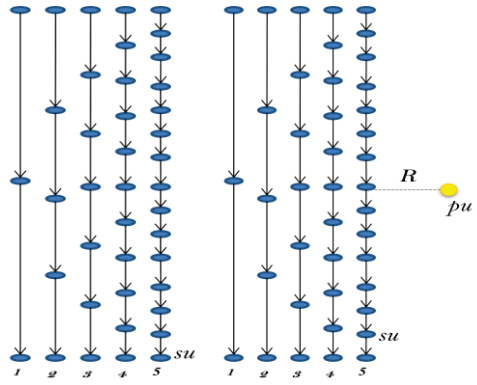
We defined A vector for this equation $\sum_{m\in M}\rho_m = 1$ which its first five elements are one and other elements are zero. Therefore, this equation $A * x = 1$, will provide us with the desired result. In this problem, we have several non-linear non-equations as follows:

$$\sum_{m\in M}\rho_m p_m^v \le \bar{P}_v \qquad (26)$$

For the next nested loops, the main program that controls with the variables of $i1$, $j1$, we will have optimum throughput rate. Then, we will define *RTemp* variable for throughput rate which is first zero and we add the new calculated rate in internal loop, we have:

$$RTemp = RTemp + log(1 + Power(i, sL + j1) * G(k));$$

For the external loop also this total amount will divide in the total amount of route links L(i1) and finally the optimum route will pour in vector and loop will execute again. After leaving loop, the



(a)Rout Index without PU (b)Rout Index with PU

Fig. 2. Rout Index

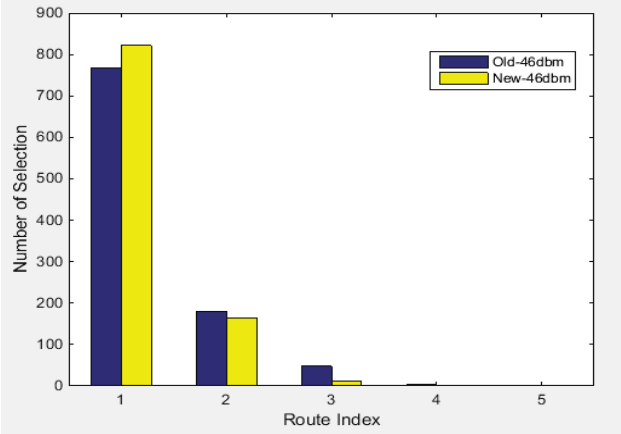
Average rates will calculate for each route (RMean)and we draw histogram for the selected routes together with chart for throughput rates of different routes. (Program execution quantity is 1000 times).

The mentioned algorithm once has been the function for one-purpose target and it just depends on throughput rate that is$U_m(\boldsymbol{R}_m) = \sum_{v\in\mathcal{V}_m}(R_m^v)^{-\kappa}/\kappa$, and another time, algorithm is for when optimization of hops quantity comes together with maximization of information transferring rate, we will solve it and finally make comparison between results

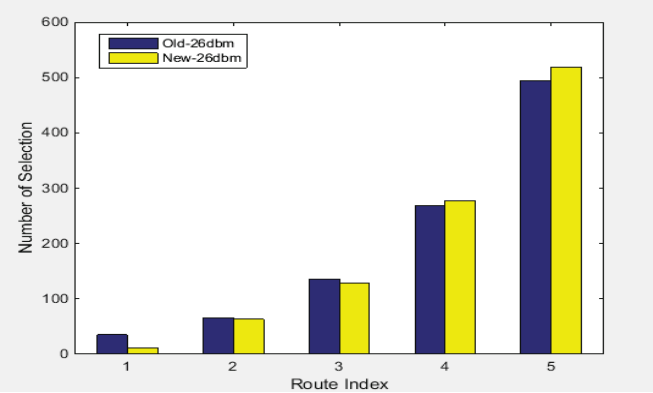
V. Simulations

In this section, we present numerical results to get insight of the optimal cross-layer design in the IEEE 802.22 WRAN system scenario. For the said model, we suppose subcarrier with 3384HZ space and subcarriers in the 6MHz system bandwidth are grouped into 32 subchannels for assignment.

Interference domain between secondary user links is 30km and distance between source, and destination is 60KM. For the first part, supposing the fixed distance between source and destination and the absence of the primery user, we suppose five routes with different hops. We checked for the probability to chose each route for every 1000 repetition. For next part, we consider on the primery user existence and simulation and calculation will bring in to play in accordance with the new situations and throughput average will draw with regard to the transferring power restriction, which is $\bar{P}$ 26 and46 dbm. It is of worth mentioning that the suggested algorithm can follow sudden changes of channels. Variable, $\bar{P}$ which is equal و 26 46dbm for all routes.



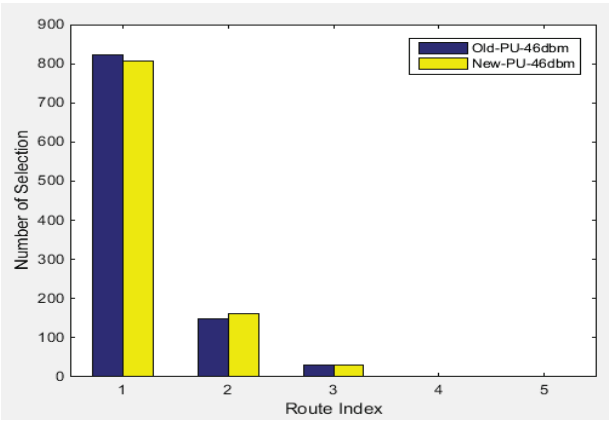
(a)Comparison for routes selection quantity for $\bar{P}$= 46 dbm



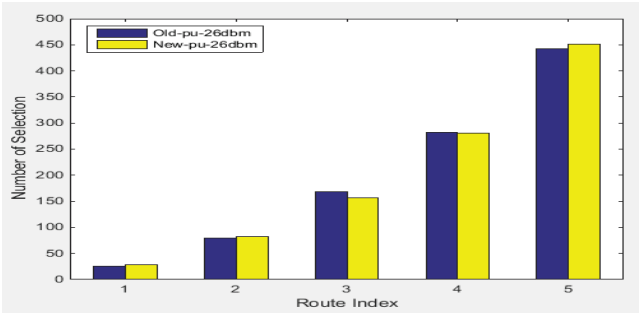
(b)Comparison for routes selection quantity for $\bar{P}$=26

Fig. 3. Relationship between bottleneck throughput and hops, without PU

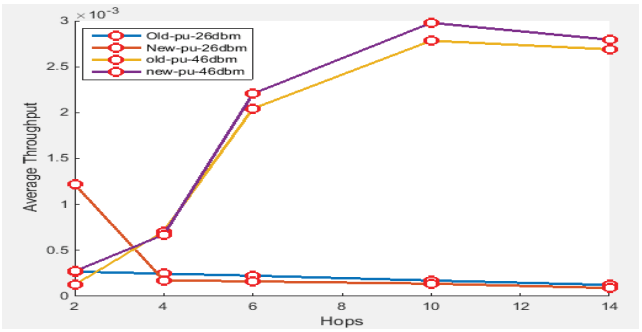
As you see, for $\bar{P}$= 46 dbm, choosing route with less hops quantity and for P = 26 dbm, choosing route with more hops quantity has priority. In fact, by increasing $\bar{P}$, route optimization algorithm chooses less hops quantity.



(a) Comparison for routes selection quantity for $\bar{P}$= 46 dbm



(b)Comparison for routes selection quantity for $\bar{P}$=26



(c)comparison between Average Throughput with powers of 26 and 46 dbm

Fig. 4. Effect of IT constraint

As you saw, for the problem definition we took chance to hops quantity effect directly through delineating new target function. Following that, based on the acquired result we concluded that the priority is for routes with more hops for amounts less than 30 dbm and without the presence of user and the quantity of the new and old methods selection tend toward each other by hops quantity decrease.

This event is true based on vice versa priority for quantities more than 30 dbm, as the selection probabilities for routes 4 and 5 tends toward zero. The probability to choose route1 reaches to its maximum, that means there exist more variance for selections choices; which means our Recognizer selects route1 Greedy and provides less hops routes with less opportunity. In fact, routes with more jump quantity do not help a lot for decreasing target function (and throughput rate increase).

VI. Conclusion

The issue of dedicating resource allocation and route selection formulated as a multi objective optimization subject and through this, we considered hops quantity directly in target function. In accordance with the consequences acquired from new target function, we understand that without primery user presence, we witness for more ravenously action of the new target function and the presence of the first user, leads to the enhancement of the throughput rate. It is also of worth mentioning that the escalation of power narrowness, leads in the enhancement of throughput, because for power restriction higher than 30 dBm, numerical amount (watt) of this parameter is more than1 and it sticks out the throughput to hold more quantities. All the same, when power limitation is less than 30 dBm, numerical amount of this parameter is less than 1 and throughput rate cannot choose greater amounts. In this paper, the nodes ate constant and to reduce complexity, selected single source and destination; so to continue, can work on MIMO network topology and defined or used mobility model. Also it can be improved constraints or the objective function as delay. Finally, the problem can be solved by other methods performance optimization.

References

[1] Ian F. Akyildiz, Mehmet Can Vuran, “Wireless Sensor Networks,” ISBN978-0-470-03601-3 2010.

[2] Joseph Mitola.Cognitive Radio, “An Integrated Agent Architecture for Software Defined Radio,” ISSN 1403-5286 2000.

[3] Anita Garhwal and Partha Pratim Bhattacharya, “A SURVEY ON DYNAMIC SPECTRUM ACCESS TECHNIQUES FOR COGNITIVE RADIO,” in International Journal of Next-Generation Networks (IJNGN), Vol.3, No.4, December 2011.

[4] Federal Communications Commission FCC 03-289,2003

[5] Anna Wisniewska, “Spectrum sharing in Cognitive Radio Networks: A Survey,” in New York: Institute of Electrical and Electron-ics Engineers, 2012.

[6] Chao-Fang Shih and Wanjiun Liao, “Exploiting Route Robustness in Joint Routing and Spectrum Allocation in Multi-Hop Cognitive Radio Networks,” 978-1-4244-6398-5/10, IEEE 2010.

[7] Chao-Fang Shih, Wanjiun Liao, and Hsi-Lu Chao, “Joint Routing and Spectrum Allocation for Multi-Hop Cognitive Radio Networks with Route Robustness Consideration,” in IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 10, NO. 9, SEPTEMBER 2011.

[8] Qing Chen, Qian Zhang and and Zhisheng Niu, “QoS-Aware Cooperative and Opportunistic Scheduling Exploiting Multiuser Diversity for Rate-Adaptive *Ad Hoc* Networks,” in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 57, NO. 2, MARCH 2008.

[9] Qianxi Lu, Tao Peng, Wei Wang, Wenbo Wang and Chao Hu, “Optimal Route Selection and Resource Allocation in Multi-Hop Cognitive Radio Networks,” in IEEE "GLOBECOM" 2009.

[10] Didem Goz upek and Fatih Alagoz, “Throughput and Delay Optimal Scheduling in Cognitive Radio Networks Under Interference Temperature Constraints,” JOURNAL OF COMMUNICATIONS AND NETWORKS, VOL. 11, NO. 2, APRIL 2009.

[11] S. Boyd, “Convex Optimization,” ISBN 978-0-521-83378-3 hardback, Stanford, CA, 2004.

[12] J. Huang, R. A. Berry and M. L. Honig, “Distributed interference compensation for wireless networks,” *IEEE J. Sel. Areas Commun.*, vol.24, pp.1074-1084, May 2006.

[13] D. Bertsekas, and R. Gallager, “*Data Networks. Englewood Cliffs*,” NJ: Prentice-Hall, 1987.

[14] Wonyong Yoon and Wooseong Kim, “Fair rate control for cognitive multi-channel wireless adhoc networks,” EURASIP Journal on Wireless Communications and Networking 2014.

Performance of OFDM-Based Adaptive Visible Light Communications

Omer Narmanlioglu
Dept. of Electrical Electronics Eng.
Ozyegin University
Istanbul, Turkey
E-mail: omer.narmanlioglu@ozyegin.edu.tr

Refik Caglar Kizilirmak
Dept. of Electrical Electronics Eng.
Nazarbayev University
Astana, Kazakhstan
E-mail: refik.kizilirmak@nu.edu.kz

Murat Uysal
Dept. of Electrical Electronics Eng.
Ozyegin University
Istanbul, Turkey
E-mail: murat.uysal@ozyegin.edu.tr

***Abstract*—In this paper, we investigate the performance of a direct current biased optical orthogonal frequency division multiplexing (DCO-OFDM) based adaptive transmission in the context of visible light communication (VLC). We consider the reference channel model (Scenario I) of IEEE 802.15.7r1 where an indoor office room is considered and channel impulse responses are presented for 24 different points. Our results reveal that high data rates such as 100 Mbps can be achieved with the choice of optimal modulation order according to channel conditions, while non-adaptive transmission based on the worst-case design provides only 14 Mbps under the consideration of target bit-error-rate (BER) of 10^{-3}.**

***Index Terms*—Visible light communication, adaptive transmission, OFDM**

I. Introduction

Visible light communication (VLC) is an emerging transmission technology as an alternative short range wireless transmission technique providing low-cost and energy-efficient solutions. VLC uses light emitting diode (LED) [1] based luminaries, that are used for illumination purposes, as a transmitter and photodetector (PD) as a receiver and it is considered as a promising complementary technology [2] to currently used radio frequency (RF) systems, especially for indoor environments.

Intensity modulation and direct detection (IM/DD) schemes are used for VLC due to the non-coherent characteristics of LEDs. The main constraints on IM/DD schemes is that the transmitted signal must be real valued and non-negative. Following this fact, VLC was initially started with simple modulation techniques such as on-off keying (OOK) and pulse place modulation (PPM) [3]. However, recent studies [4]–[8] have adopted orthogonal frequency-division multiplexing (OFDM) which provides reliable transmission and achieves high data rate over wideband channels. Due to the constraints of IM/DD, conventional OFDM structure cannot be directly adopted to VLC. In order to overcome these constraints, different methods such as direct current biased optical-OFDM (DCO-OFDM) [4], asymmetrically clipped O-OFDM (ACO-OFDM) [5], flip (or unipolar) OFDM [6], [7] and enhanced unipolar OFDM [8] have been proposed.

In the literature, a number of studies that investigate the adaptive transmission for VLC exist [9]–[12]. In [9], adaptive modulation has been applied to VLC systems with a single transmitter under frequency flat channel assumption. The selectivity in frequency was associated with the bandwidth of the LEDs. [10] proposes an adaptive control of the OFDM modulation-order to maintain the VLC transmission and to keep the same luminance with respect to the distance and the offset between the LEDs and receiver side. In [11], a bit loading scheme for DCO-OFDM based systems with two luminaries in the room is studied. [12] proposes an adaptive multiple-input and multiple-output (MIMO) OFDM based VLC using a receiver module with angular diversity. With the help of singular value decomposition, bit and power loading schemes are analyzed in two considered points inside an indoor room environment. None of the above works, however, considers the effect of user walking inside a room and realistic frequency-selective characteristics of visible light channel. In this work, we investigate the adaptive transmission in terms of optimal modulation order selection while satisfying a target bit-error-rate (BER). Our analysis includes realistic indoor VLC channel models that have been also accepted as IEEE 802.15.7r1 VLC reference channel model [13].

The remainder of this work is organized as follows. In Section II, we briefly summarize the IEEE 802.15.7r1 channel model. In Section III, we investigate the transmission model of DCO-OFDM based VLC and adaptive transmission algorithm. BER and data rate performance results are available in Section IV. Finally, we conclude the paper in Section V.

Notation: $||.||^2$, $(.)^*$ and $[.]^{\mathrm{T}}$ denote Euclidean distance, complex conjugate and transpose, respectively. $\otimes$ is convolution operator, $\delta(t)$ is the impulse function, $F\{.\}$ denotes Fourier transform and $Q(.)$ is the tail probability of standard normal distribution.

II. Channel Model

We use the reference channel model (see Fig. 1) adopted by IEEE 802.15.7r1 Task Group "Short Range Optical Wireless Communication" [13] where an office environment with the dimensions of 14 m x 14 m x 3 m is considered. The locations of 32 LEDs, which work on the same power level and emit the same information, are shown in Fig. 1 (a). In the model, 24 test points (see Fig. 1 (b)) are chosen which are categorized into three groups. The groups are: (a) the people who stand with a cell phone in hand at the locations D1-D12 presented

by blue color in the corridors at a height of 1.7 m with 45° rotation, (b) the people with a cell phone in hand with the locations D13-D18 presented by red color on the top of chairs at a height of 0.95 m with 45° rotation and (c) the people who sit with a cell phone in hand to his/her ear with the locations of D19-D24 presented by green color on the top of chairs at a height of 1.1 m with 45° rotation.

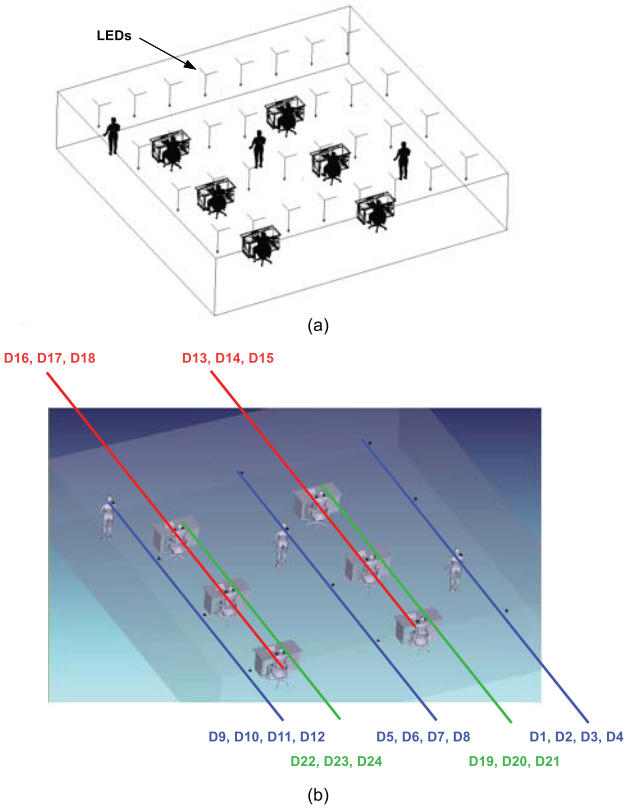


Figure 1: Indoor room environment [13].

In the model, the half viewing angle of LEDs is taken as 40°. The field of view and area of PDs are respectively set to 85° and 1 cm^2. We represent the optical channel impulse response (CIR) for each PD by $c_i(t)$ where $i \in \{1, ..., 24\}$. The CIR for each PD is obtained by ray tracing simulations using Zemax®. The band-limited electrical CIR is denoted by $h_i(t) = g_T(t) \otimes c_i(t) \otimes g_R(t)$ where $g_T(t)$ and $g_R(t)$ denote electrical transmit and receive matched filter responses, respectively. In our work, we use square root raised cosine (SRRC) filter with excess bandwidth parameter (α) of 0.5. Fig. 2 presents the obtained channel frequency responses. There are large variations in channel gains with respect to the locations such as PD4 has the best channel conditions as PD7 has the worst channel conditions. This situation necessitates the use of adaptive transmission.

III. System Model

In DCO-OFDM, binary information is first mapped to complex symbols s_1 s_2 ... $s_{N/2-1}$ based on the deployed constellation scheme such as $M-$order phase shift keying

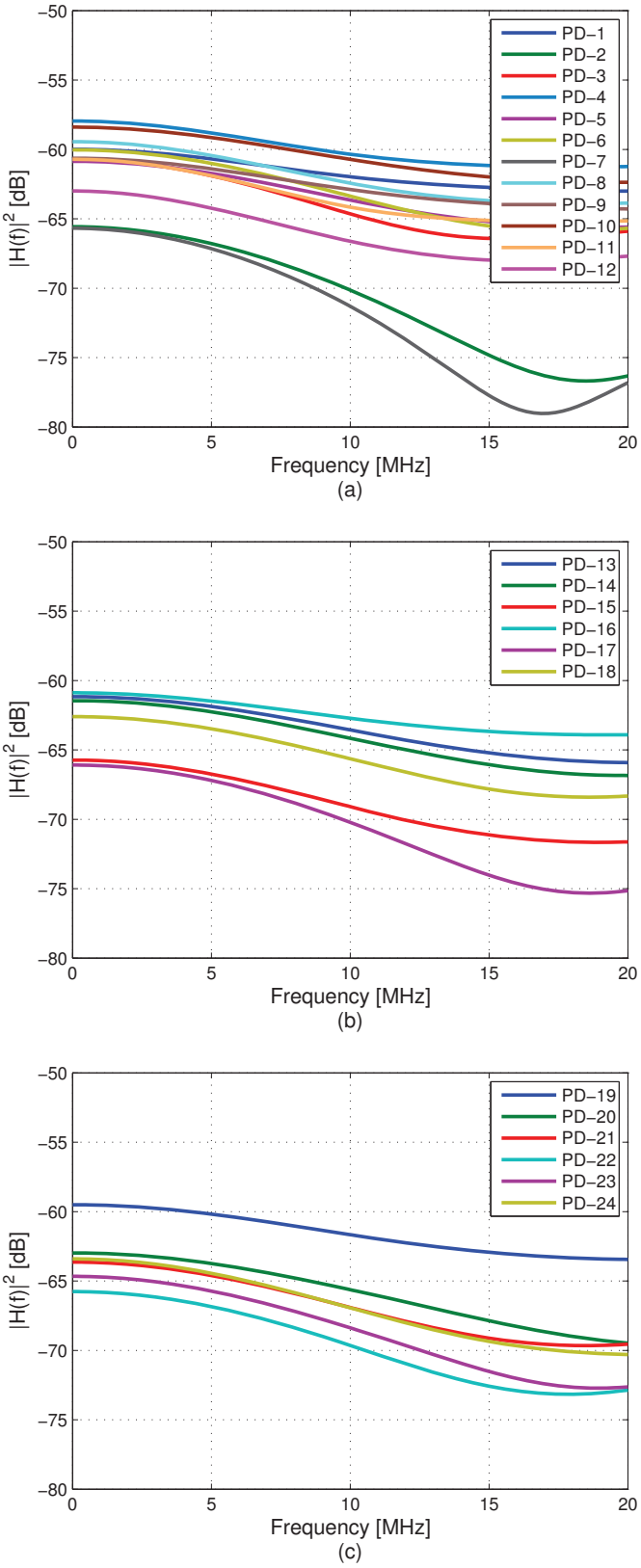


Figure 2: Channel frequency responses for 24 different test points including (a) the people who stand with a cell phone in hand, (b) the people with a cell phone in hand and (c) the people who sit with a cell phone in hand.

(PSK) or quadrature amplitude modulation (QAM) where N is the number of subcarriers. To ensure that the output of inverse fast Fourier transform (IFFT) is real valued, Hermitian symmetry is imposed on the transmitted signal which yields

$$\mathbf{X_S} = \left[0\ s_1\ s_2\ s_3\ ...\ s_{N/2-1}\ 0\ s^*_{N/2-1}\ ...\ s^*_3\ s^*_2\ s^*_1\right]^{\mathrm{T}}. \quad (1)$$

The output of $N-$ IFFT can be written as

$$x_S[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_S[k] e^{j\frac{2\pi nk}{N}}. \quad (2)$$

A cyclic prefix whose duration is greater than or equal to the delay spread of the channel is appended to beginning of each OFDM frame, hence, linear convolution can be modelled as circular convolution. After parallel to serial conversion and addition of DC bias (B_{DC}), the information signal $x_S(t)$ is obtained. The amplitude levels of $x_S(t)$ beyond the dynamic range will be clipped, however, we assume non-clipping case without loss of generality. The time domain continuous waveform that drives LEDs can be written as

$$x_S(t) = \sum_{n=0}^{N+N_{CP}-1} x_S[n]\delta(t - nT_S) + B_{DC} \quad (3)$$

where N_{CP} is the length of appended cyclic prefix (CP) and T_S is the sampling interval. The time domain received signal by PD can be written as

$$y(t) = Rx_S(t) \otimes h(t) + v(t) \quad (4)$$

where R is the responsivity of PD (A/W) and $v(t)$ is additive white Gaussian noise (AWGN) with zero mean and σ^2_N variance calculated by N_0W in which N_0 is noise power spectral density and W denotes the system bandwidth.

In receiver side, after applying matched filter, the signal is converted to parallel streams and cyclic prefix is removed. The signal on the k^{th} subcarrier at the output of $N-$ FFT can be written as

$$Y[k] = RX_S[k]H[k] + V[k] \quad (5)$$

where $Y[k] = F\{y(t)\}$, $X_S[k] = F\{x_S(t)\}$, $H[k] = F\{h(t)\}$ and $V[k] = F\{v(t)\}$. The output of *Maximum Likelihood* (ML) decision rule is then written as

$$\hat{X}[k] = \underset{x\in\Phi}{\arg\min} \left[\|Y[k] - RH[k]x\|^2\right] \quad (6)$$

where Φ denotes the symbols in deployed constellation scheme with respect to the average transmitted power of P_X. The subcarrier-based SNR can be written as

$$\mathrm{SNR}[k] = \frac{R^2 P_X \left|H[k]\right|^2}{\sigma^2_N}. \quad (7)$$

The adaptive transmission method aims to find the optimal modulation order that has the highest data rate while satisfying a target BER for a given SNR values. This process can be run with the help of look-up table (LUT) containing SNR boundaries that are required to satisfy target BER for different modulation orders. The data rate can be calculated as

$$\text{Data Rate} = \frac{1}{T_S} \frac{N_D}{N + N_{CP}} \log_2 M \text{ bit/sec (bps)} \quad (8)$$

where N_D is the number of subcarrier that are used for data transmission. Overall BER can be calculated by averaging the subcarrier-based BER values such as

$$\mathrm{BER} = \frac{1}{N_D} \sum_k \mathrm{BER_{SC}}[k]. \quad (9)$$

In (9), $\mathrm{BER_{SC}}$ can be calculated by using (10) [14] (see the top of the next page).

IV. Numerical Results

In this section, we demonstrate the performance improvement with the use of adaptive transmission method. Simulation parameters are defined in Table I. In all simulations, we assume that perfect channel state information is available at both transmitter and receiver terminals.

TABLE I: Simulation Parameters.

Sampling interval (T_S)	25 nsec
Average transmitted power (P_X)	-30 dBm
Power spectral density of noise (N_0)	10^{-22} W/Hz
Responsivity of PD (R)	1.0 A/W
Number of subcarrier (N)	256
Number of data carrier (N_D)	96
Cyclic prefix length (N_{CP})	12
Target BER	10^{-3}

With the OFDM parameters given in Table I, the data rates for different modulation orders are calculated and presented in Table II. The LUT that is generated based on (10), is given in Table III.

TABLE II: Data Rates Of Different Modulation Orders.

Modulation Order	**Data Rate** (Mbps)
$2-$PSK	14.33
$4-$QAM	28.66
$8-$QAM	42.99
$16-$QAM	57.31
$32-$QAM	71.64
$64-$QAM	85.97
$128-$QAM	100.30
$256-$QAM	114.63

TABLE III: Required SNR Levels To Satisfy Target BER Of 10^{-3}.

Modulation	**Required SNR** [dB]
$2-$PSK	6.79
$4-$QAM	9.80
$8-$QAM	14.42
$16-$QAM	16.54
$32-$QAM	20.57
$64-$QAM	22.55
$128-$QAM	26.46
$256-$QAM	28.42

In Fig. 3, we present the BER values of different modulation orders for each PD location. Relatively, the highest modulation order which satisfies the target BER is selected. For example at D4, $128-$QAM can be deployed since it satisfies the target BER. At D7, however, $2-$PSK should be deployed for the

$$\mathrm{BER}_{\mathrm{SC}}[k] \approx \begin{cases} Q\left(\sqrt{2\mathrm{SNR}[k]}\right) & , \quad 2-\mathrm{PSK} \\ \frac{2\left(\sqrt{M}-1\right)}{\sqrt{M}\log_2\sqrt{M}} Q\left(\sqrt{\frac{3\mathrm{SNR}[k]}{(M-1)}}\right) & , \quad \mathrm{square}-M-\mathrm{QAM} \\ \frac{2}{\log_2(U\mathrm{x}J)}\left[\frac{U-1}{U} Q\left(\sqrt{\frac{6\mathrm{SNR}[k]}{U^2+J^2-2}}\right) + \frac{J-1}{J} Q\left(\sqrt{\frac{6\mathrm{SNR}[k]}{U^2+J^2-2}}\right)\right] & , \quad \mathrm{rectangular}-M=U\mathrm{x}J-\mathrm{QAM} \end{cases} \quad (10)$$

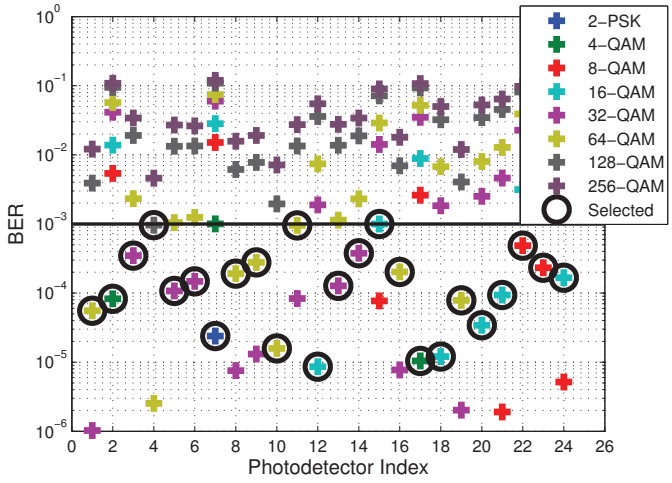


Figure 3: BER results with respect to different modulation order at 24 different locations.

same target BER.

With the use of adaptive modulation, the data rates of a user walking through these points are shown in Fig. 4. Significant improvements are achieved through adaptive transmission such as data rate reaches to 100.30 Mbps while non-adaptive transmission designed with $2-\mathrm{PSK}$ provides only 14.33 Mbps.

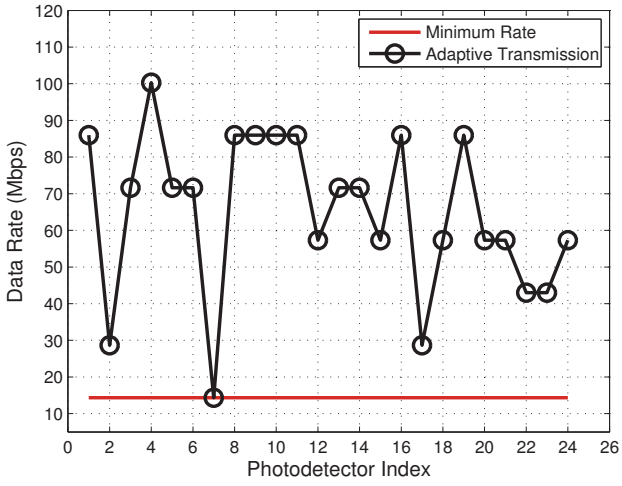


Figure 4: Adaptive transmission data rates at 24 different locations.

V. Conclusion

In this paper, we have investigated the performance of DCO-OFDM based adaptive VLC system under a realistic indoor environment. We have demonstrated superior performance with adaptive transmission for 24 different PD locations. With the choice of optimal modulation order according to channel conditions, the data rate performance is significantly improved (i.e., up to 100 Mbps) while non-adaptive transmission provides only 14.33 Mbps.

References

[1] S. Arnon, J. Barry, G. Karagiannidis, R. Schober, and M. Uysal, *Advanced optical wireless communication systems*. Cambridge university press, 2012.

[2] S. Wu, H. Wang, and C. H. Youn, "Visible light communications for 5G wireless networking systems: from fixed to mobile communications," *IEEE Network*, vol. 28, pp. 41–45, Nov 2014.

[3] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 100–107, Feb 2004.

[4] J. Armstrong, "OFDM for Optical Communications," *Journal of Lightwave Technology*, vol. 27, pp. 189–204, Feb 2009.

[5] J. Armstrong and A. J. Lowery, "Power efficient optical OFDM," *Electronics Letters*, vol. 42, pp. 370–372, March 2006.

[6] N. Fernando, Y. Hong, and E. Viterbo, "Flip-OFDM for Unipolar Communication Systems," *IEEE Transactions on Communications*, vol. 60, pp. 3726–3733, December 2012.

[7] D. Tsonev, S. Sinanovic, and H. Haas, "Novel Unipolar Orthogonal Frequency Division Multiplexing (U-OFDM) for Optical Wireless," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1–5, May 2012.

[8] D. Tsonev, S. Videv, and H. Haas, "Unlocking Spectral Efficiency in Intensity Modulation and Direct Detection Systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1758–1770, Sept 2015.

[9] L. Wu, Z. Zhang, J. Dang, and H. Liu, "Adaptive Modulation Schemes for Visible Light Communications," *Journal of Lightwave Technology*, vol. 33, pp. 117–125, Jan 2015.

[10] Y. F. Liu, C. H. Yeh, C. W. Chow, and Y. Liu, "Adaptive control of OFDM data rate for LED visible light communications in different free space transmission distances and offsets," in *Communications and Photonics Conference (ACP), 2012 Asia*, pp. 1–3, Nov 2012.

[11] N. Kalikulov and R. C. Kizilirmak, "Adaptive OFDM for VLC Systems with multiple transmitters," in *International Wireless Symposium (IWS) International*, pp. 544–549, IEEE, 2016.

[12] Y. Hong, T. Wu, and L. K. Chen, "On the Performance of Adaptive MIMO-OFDM Indoor Visible Light Communications," *IEEE Photonics Technology Letters*, vol. 28, pp. 907–910, April 2016.

[13] M. Uysal, T. Baykas, F. Miramirkhani, N. Serafimovski, and V. Jungnickel, *TG7r1 channel model document for high-rate PD communications*. doc: IEEE 802.15-15/0685r0, September 2015. Available: https://mentor.ieee.org/802.15/dcn/15/15-15-0746-01-007a-tg7r1-channel-model-document-for-high-rate-pd-communications.pdf.

[14] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, pp. 1074–1080, Jul 2002.

Effective Multithreshold Decoding Algorithms for Wireless Communication Channels

Zolotarev V., Ovechkin G.
Department of Computational and Applied Mathematics
Ryazan state radio engineering university
Ryazan, Russian Federation
g_ovechkin@mail.ru

Issainova A., Satybaldina D., Tashatov N.
Department of Computer Engineering
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
satybaldina_dzh@enu.kz

***Abstract*— Multithreshold decoding algorithms for self-orthogonal error-correction codes have been considered. Analytical estimations of their efficiency in the uncorrelated Rayleigh and Rician channels have been represented. Results of the computer simulation have been obtained. Possibilities to apply multithreshold decoders in combination with such technologies of the efficiency increase of data transmission systems in wireless fading channels as OFDM, MIMO, precoding are considered. The paper has shown that application of precoding under such conditions ensures a significant improvement of coding gain of the data transmission system.**

***Index Terms*— error-correcting coding, self-orthogonal code, multithreshold decoder, coding gain, fading channel**

I. Introduction

Error-correction coding are applied in communication systems widely. At present there are a lot of codes and algorithms of their decoding, for example, Bose-Chaudhuri-Hocquenghem block codes, Reed-Solomon codes, convolutional codes, Low-Density Parity-Check (LDPC) codes and turbo codes. New and very effective solution of the low complexity decoding under simultaneous high coding gain on the base of multithreshold decoders (MTD) for self-orthogonal codes (SOC) has been suggested by Russian specialists [1].

At present MTD characteristics are well investigated for channels with independent errors where such method ensure near to optimal decoding (OD) even for very long codes only with linear implementation complexity [1, 2, 3]. For such channels a range of methods to improve MTD efficiency is known [1, 2, 4]. Also interesting results have been obtained for symbol MTD [4, 5]. Results of theoretical and experimental researches show MTD decoders require hundreds time less additive equivalent operation in comparison with known decoders for turbo and LDPC codes at similar coding gain. So MTD is very interesting for implementation in high rate communication and data storage systems [3, 4, 6, 7]. At the same time according to the coding theory, MTD efficiency in Gaussian channels can be slightly improved. Besides, future communication systems operate under more sophisticated conditions occurred due to the multipath signal propagation, Doppler shift and other reasons. As a result, errors occurred in channels are grouped into packages. Under such conditions, effect of the coding application is higher than in channels with independent errors since here in some cases it is impossible to decrease a bit error probability only by increase of the transmitter power [8]. Consequently, development of algorithms for increase of the MTD efficiency in channels with grouped errors (***that is the aim of the present paper***) will allow increasing an coding gain which can be used for improvement of technical characteristics of data transmission systems and significantly expand an application area of such method. Besides, it is very important to keep and only insignificantly increase a sophistication of the initial multithreshold decoder realization because only the simplest methods of error correction can ensure presently required decoding rate about tens Gb/s.

II. Multithreshold decoders

Multithreshold decoders are used for decoding of block or convolutional self-orthogonal codes. Main principles of MTD operation used for decoding of block SOC with code rate 1/2 and length 26 bits set by the generator polynomial $g(x) = 1 + x^1 + x^4 + x^6$ are demonstrated by the scheme shown in Fig. 1 [1]. It should be noted that MTD composition includes registers, half-adders and a threshold element summing its inputs and comparing an obtained amount with the threshold. It makes MTD to be the simplest device for realization which can ensure maximum possible decoding rate.

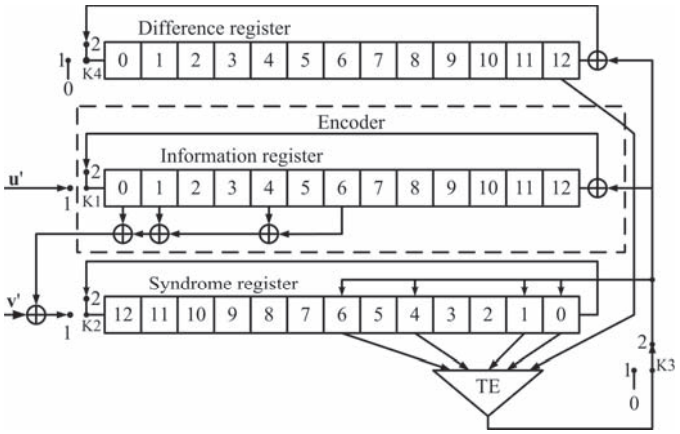


Fig. 1. MTD circuit of the block SOC

Under MTD operation in the binary symmetric channel, function L_j is calculated for the current decoding symbol u_j at the threshold element. This function depends on syndrome elements s_{jk} and corresponding element of the difference register d_j:

$$L_j = \sum_{s_{jk} \in \{S_j\}} s_{jk} + d_j \,, \tag{1}$$

where $\{S_j\}$ – a set of checks (syndrome elements) in relation to the error e_j in the decoding symbol u_j. When the function exceeds some threshold value, change of the decoding symbol, corresponding element of the difference register and syndrome elements participating in calculation is executed.

Under MTD operation in the Gaussian channel, function L_j is also calculated for the current symbol u_j. And syndrome elements and element of the difference register are summed with some coefficients reflecting their reliability:

$$L_j = \sum_{s_{jk} \in \{S_j\}} (2s_{j_k} - 1)w_{j_k} + (2d_j - 1)w_j, \tag{2}$$

where w_{j_k} – a reliability of the check s_{j_k}; w_j – a reliability of the accepted symbol u_j. For example, an absolute value of likelihood ratio logarithm can be used as estimations of reliability of symbols accepted from the channel.

Papers [1, 2] show that decision of MTD converges to the OD decision under each change of the decoding symbol. This is because total weight of the syndrome and difference registers decreases under change of the symbol. Thus transfer to the maximum likelihood codeword may be done. However, due to occurrence of the error propagation effect [1, 2], process of transfer from one codeword to another that is more likely, can be stopped before MTD reaches a solution of the OD. Such effect is that a large number of errors made by the decoder enter the syndrome through branches of feedback after the decoder error occurrence. These errors prevent from correct decoding of informational symbols at following iterations. This leads to the fact that probability of the second error increases and error burst appears. Consequently, it is necessary to use codes maximally resistant to error propagation [1, 2] in order to approximate a probability of MTD error correction to the OD one. The error propagation estimation described in [1] can be used under code construction.

III. Analytical estimation of the MTD error probability in Rayleigh and Rician channels

The simplest models of fading channels are models of channels with Rayleigh and Rician fading which occur under presence of a multipath signal propagation. At the same time there is no line of sight between a transmitter and receiver in the Rayleigh channel, but it is present in the Rician channel.

Let's obtain a lower bound of the MTD error probability. For simplicity, we assume that BPSK or QPSK modulation and hard decision demodulator are used. Let's note that for MTD the lower bound of error probability in the binary symmetric channel is determined by error probability of the OD calculated according to expression [1, 2]:

$$P_b = \sum_{i=(d+1)/2}^{d} C_d^i \, p^i \left(1-p\right)^{d-i}, \tag{3}$$

where d – a code distance of the used SOC (assumed as odd); p – probability of an error in the channel.

For the noncorrelated Rayleigh channel, expression of the error probability is known [9]

$$p = \frac{1}{2}\left(1 - \sqrt{\frac{E_S / N_0}{1 + E_S / N_0}}\right), \tag{4}$$

where E_s/N_0 – a signal to noise ratio. Such estimation is correct for BPSK or QPSK modulation and Doppler shift $F_d = 0$. In the case of the channel with Rician fading error probability according to [9] is determined as

$$p = 2Q\left(\sqrt{\frac{2kE_S / N_0}{k + E_S / N_0}}\right), \tag{5}$$

where k – a Rician coefficient; $Q(x)$ is determined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt \,. \tag{6}$$

Taking into account that MTD can operate almost as OD, we put (4) into (3) and obtain the lower bound of the bit error probability for MTD in the Rayleigh channel:

$$P_b = \sum_{i=(d+1)/2}^{d} \left(C_d^i \frac{1}{2}\left(1 - \sqrt{\frac{E_S / N_0}{1 + E_S / N_0}}\right)^i \cdot \left(1 - \frac{1}{2}\left(1 - \sqrt{\frac{E_S / N_0}{1 + E_S / N_0}}\right)\right)^{d-i} \right). \tag{7}$$

Similarly, bit error probability in the Rician channel is determined as

$$P_b = \sum_{i=(d+1)/2}^{d} \left(C_d^i Q\left[\sqrt{\frac{2kE_S / N_0}{k + E_S / N_0}}\right]^i \cdot \left(1 - Q\left[\sqrt{\frac{2kE_S / N_0}{k + E_S / N_0}}\right]\right)^{d-i} \right). \tag{8}$$

Let's note that expressions in such form allow obtaining an estimation of the MTD error probability under usage of SOC with the odd code distance.

In Fig. 2 curves *5* and *6* show estimation of the error probability in the noncorrelated Rayleigh channel and error probability after MTD decoder for SOC with a minimum code distance $d = 9$. Let's note that losses about 1,5 dB are observed in comparison with the Gaussian channel. In the same figure curves *1* and *2* show estimation of the channel and MTD error probability obtained by means of computer simulation using a block SOC with length 20000 bits, code rate $R = 1/2$ and code distance $d = 9$ and MTD with 15 decoding iterations. Let's note that estimation of the channel error probability well conforms to the analytical one that allows using the latest one in obtained formulas. Besides, we can see that obtained analytical estimation of the decoder error probability under usage of MTD in the noncorrelated Rayleigh channel is enough correct for the area of MTD effective operation. Similar results for the Rician channel under $k = 5$ are shown in Fig. 2 by curves *7, 8* for analytical estimations and curves *3*, *4* for results of the simulation. And in this case obtained estimation is good for the area of MTD effective operation.

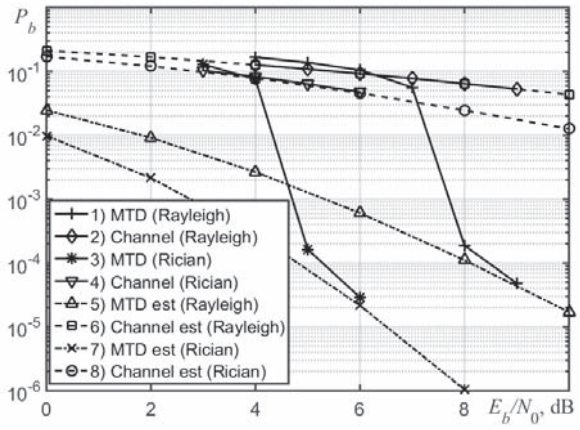


Fig. 2. Analytical estimations of the error probability and results of the simulation for MTD in Rayleigh and Rician channels

Let's note that it is not possible to obtain analytical estimations with accepted accuracy under present of correlated fading or usage of the modulation of a higher order for larger noise. So, under such conditions computer simulation should be used for estimation of the MTD efficiency.

IV. Results of the Simulation of MTD in Rayleigh and Rician Channels

In Fig. 3 curves *2*, *4* and *6* show results of the simulation for Rayleigh and Rician fading under Doppler frequency 100 and 150 Hz and usage of BPSK modulation and hard decision demodulator. Here MTD was used for the code similar to the code presented in the Fig. 2. We note that presence of correlated fading significantly (for Rayleigh fading – up to 6 dB) worsens an operation signal to noise ratio. This is because in such case a code with parallel concatenation was used and data from the coder are delivered into the modulator "by lines" without usage of an interleaver. As a result under occurrence of fading, the whole branches with greater check dimension could be distorted. That led to appearance of blocks with a significant number of errors after decoding. Additional interleaver should be used, that is not always convenient, or internal interlacing should be organized giving data from the coder "by columns" in order to improve the MTD efficiency. In Fig. 3 curves *1*, *3* and *5* show results of the simulation corresponding to the suggested parallel bit transmission. Schedule review has shown that in the case of bit transmission "by columns" in the channel with Rician fading, decrease of the decoding error probability in comparison with the original variant, for example, under signal to noise ratio 7 dB under $k = 5$ and $F_d = 150$, becomes greater two decimal exponents. Gain is also observed under other simulation conditions. This is because under parallel transmission, errors caused by signal fading become diversed by different branches. Besides, branches with greater check dimension are distorted much less and MTD becomes capable to ensure less decoding error probability.

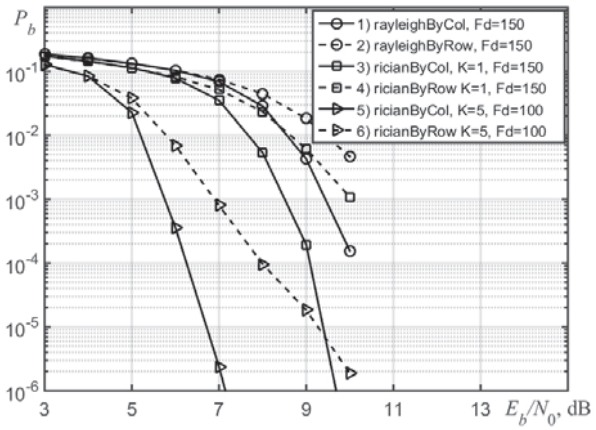


Fig. 3. MTD characteristics in the fading channels under successive and parallel bit transmission

V. MTD Operation in Wireless Communication Systems with Precoding

We should note that mentioned cases of the MTD application allow successfully error correction if there is no any intersymbol interference. If intersymbol interference is present, it is necessary to use additional facilitates as orthogonal frequency division multiplexing (OFDM) or others. Besides, analysis of results presented in [10] has shown that decrease of the error probability is slower in channels with fading under increase of the signal to noise ratio in comparison with the Gaussian channel. This is because under long fading even under high signal to noise ratio, a large error packages appears at the decoder input. The decoder cannot correct them. Technology of the spatial diversity is often used to decrease influence of such effect when several transmitting and receiving antennas are used (Multiple Input Multiple Output – MIMO). Under correct development of such system we can consider that resulting sub channels turn out to be independent. Consequently, probability of the fact that all channels are simultaneously exposed to fading will be much less in comparison with the single-channel system. It allows significantly decreasing a reliability of the data transmission.

We note that under usage of MIMO there is an additional possibility to use of Space-Time Coding (STC) [11] – coding reflecting transmitting symbols of the signal constellation to transmitting antennas which usage allows significantly improving characteristics of the system using MTD [12].

Also we note that previously obtained results supposed that the receiving side exactly knows a channel state, i.e. multiplicative noise component is known for each transmitting channel. In practice pilot symbols added to the transmitting channel are used for estimation of the channel. The demodulator used this exact information under estimation of signals received from the channel. In some cases such information is also known at the transmitting side (for example, in the case when channel is changed enough slowly, transmission of this information can be organized by the reverse channel). The transmitter can use information of the channel state for redistribution of the signal energy by spatial channels by loading "good" channels and emptying "bad" channels. Such actions are fulfilled by so-called precoding of the transmitting signal. Let's consider precoding in details.

In the case of usage of MIMO transmitter architecture when each antenna transmittes its own symbol, the channel model is described by expression

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{7}$$

where $\mathbf{x} \in \mathrm{C}^M$ – a transmitting vector from M complex values; $\mathbf{n} \in \mathrm{C}^N$ – a vector of the uncorrelated complex Gaussian noise; $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_N) \in \mathrm{C}^{M \times N}$ – a complex matrix of the channel; M – a number of transmitting antennas; N – a number of receiving antennas. We note that values $h_{i,j}$ are random, but for the mentioned variant they are known by the transmitting and receiving sides.

For such channel model, vector of informational symbols $\mathbf{s} \in \mathrm{C}^L$ from some alphabet is precoded by the linear operation

$$\mathbf{x} = \mathbf{B}\mathbf{s}, \tag{8}$$

where $\mathbf{B} \in \mathrm{C}^{M \times L}$ – a complex matrix of the precoder; $L \le \mathrm{rank}(\mathbf{H})$ – a number of active channels.

Optimal MMSE receiver for such channel and precoder is represented as a Wiener filter with matrix

$$\mathbf{G} = \mathbf{B}^*\mathbf{H}^*(\mathbf{H}\mathbf{B}\mathbf{B}^*\mathbf{H}^* + \sigma_n^2\mathbf{I}_N)^{-1}, \tag{9}$$

where ()* – a complex conjugation; σ_n^2 – a variance of the Gaussian noise; $\mathbf{I}_N$ – an identity matrix with dimensions NxN.

Such filter provides an optimal estimation of transmitted symbols at the output

$$\hat{\mathbf{s}} = \mathbf{G}\mathbf{y}. \tag{10}$$

There are a great number of various algorithms of precoding including different in complexity of the optimal receiver. Some interesting precoders optimal by some criterion which receiver has a linear complexity are suggested in [13].

The first of these algorithms minimizes a total root-mean-square error at the receiver output. For such precoder matrix of precoding has a form

$$\mathbf{B} = \tilde{\mathbf{V}}\mathbf{\Phi}^{1/2}, \tag{11}$$

where $\tilde{\mathbf{V}} \in \mathrm{C}^{M \times L}$ – a complex matrix made of the first L columns of the unitary matrix contained in the singular decomposition of the symmetric matrix $\mathbf{H}\mathbf{H}^*$ where all eigenvalues λ_{nn} of the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_{11}, \lambda_{22}, ..., \lambda_{NN})$ are negative and ordered by descending; $\mathbf{\Phi} \in \mathrm{R}^{L \times L}$ – a diagonal matrix consisting of L negative numbers

$$\phi_{ii} = \left(\frac{P_0 + \sigma_n^2 \sum_{n=1}^{\bar{L}} \lambda_{nn}^{-1}}{\sum_{n=1}^{\bar{L}} \lambda_{nn}^{-1/2}} \lambda_{ii}^{-1/2} - \frac{\sigma_n^2}{\lambda_{ii}} \right)^+, \; i=1,...,\Lambda, \tag{12}$$

where $(x)^+ = \max(x, 0)$; $\bar{L} \le L$ is chosen as $\phi_{nn} > 0$ for $n \in [1, \bar{L}]$ and $\phi_{nn} = 0$ for all other n; P_0 – a mean power of the transmitting signal.

The second precoder maximizes transmitted information between tranmitting and receiving data. As distinct from the first variant here elements of the matrix $\mathbf{\Phi} \in \mathrm{R}^{L \times L}$ are determined by following

$$\phi_{ii} = \left(\frac{P_0 + \sigma_n^2 \sum_{n=1}^{\bar{L}} \lambda_{nn}^{-1}}{\bar{L}} \lambda_{ii}^{-1} - \frac{\sigma_n^2}{\lambda_{ii}} \right)^+. \; i=1,...,\Lambda. \tag{11}$$

Then we consider an efficiency of the application of these algorithms together with MTD. Under simulation, MTD is used with 30 decoding iterations for the designed SOC with code rate $R = 8/16$, code distance 17 and length 43200 bits. OFDM with 1024 subcarriers is applied together with MTD. Guard interval was 1/16 of the OFDM symbol length. Ordinary QPSK was used as modulation. Spatial Channel Model of the type Urban micro was used under acquisition of results. Maximum Doppler frequency F_d was equal to 0. Figure 4 shows MTD characteristics under above mentioned conditions under usage of a various number of transmitting and receiving antennas and various precoding algorithms.

In this figure, a demodulator formed only hard decisions under acquisition of curves *1* and *3*, and it estimates reliability of decisions under acquisition of other curves. Let's note that using of soft decision demodulator ensures an increase of the coding gain up to 2 dB in comparison with usage of hard decisions. Also we would like to note that 2x2 variant by

energy becomes worse than a single-channel one in 3 dB, but spectral efficiency of the system becomes twice better. Usage of precoding together with two transmitting antennas decreases a loss by energy up to 1 dB. Interesting results are observed for four transmitting antennas. Here under usage of the second precoder results are significantly improved. It becomes possible due to the fact that "bad" spatial channels become disconnected more efficiently in comparison with the first precoder (for example, under signal to noise ratio 5 dB, 2.5 of 4 spatial channels are averagely used, and for the first precoder – 3.5 channels). As a results energy characteristics become better than a single-channel variant more than in 3.5 dB under aproximately 2.5 times better spectral efficiency.

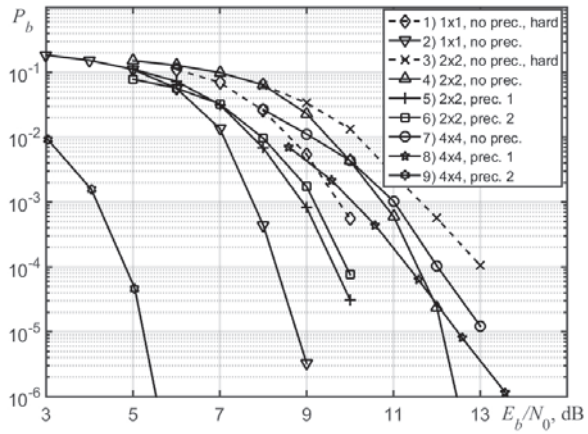


Fig. 4. MTD characteristics for various numbers of receiving and transmitting antennas and various precoding algorithms

So, obtained results allow stating that due to application of MIMO and precoding it is possible to improve significantly MTD efficiency in multipath fading channels in comparison with a single-channel variant. We note that results obtained in this section are acquired within research and development works executed under support of the Russian Scientific Foundation (project №14-19-01263).

VI. Conclusion

The paper has obtained new results of the MTD efficiency research in a range of typical multipath models of communication channels with fading. Recommendations for the better matching of coder and channel, uncorrelated Rayleigh and Rician channels have been given for these channels. The lowest estimations of the multithreshold decoding error probability have been obtained. Research of the precoding application efficiency together with multithreshold decoders and MIMO technology has been executed.

Analysis of the results represented in the paper has shown that MTD providing high efficiency in Gaussian channels become capable to error correction under even more sophisticated conditions of application. Besides, effect of the error-correcting coding application exceeds an effect of the coding application in the Gaussian channel in many times under fading channels.

Acknowledgment

The present work has been executed under financial support of the Russian Foundation of Basic Research and Russian Scientific Foundation in Ryazan state radio engineering university Large volume of additional information on MTD can be found in web-sites www.mtdbest.ru and www.mtdbest.iki.rssi.ru.

References

[1] Zolotarev V.V., Zubarev Y.B., Ovechkin G.V. Optimization Coding Theory and Multithreshold Algorithms. Published in Switzerland by ITU. March 2016. 158 p. URL http://www.itu.int/pub/S-GEN-OCTMA-2015.

[2] Zolotarev V.V., Ovechkin G.V. Increase of reliability of data transmission and data storage using multithreshold methods of decoding of error-correcting codes // Digital signal processing, 2012, No.1, P.16–21.

[3] Zolotarev V, Ovechkin G., Satybaldina D., Tashatov N., Adamova A., Mishin V. Efficiency multithreshold decoders for self-orthogonal block codes for optical channels // International Journal of Circuits, Systems and Signal Processing. ISSN 1998-4464. 2014. Vol.8, P.487–495.

[4] Ovechkin G., Ovechkin P., Satybaldina D., Beisebekova A., Tashatov N. Improving Performance of Non-Binary Multithreshold Decoder's Work Due to Concatenation // Proc. of the 18th Int. Conf. on Communications (part of CSCC '14). Santorini Island, Greece, July 17-21, 2014. pp. 100–104.

[5] Ovechkin G., Zolotarev V., Ovechkin P., Satibaldina D., Tashatov N. The Performance of Concatenated Schemes Based on Non-binary Multithreshold Decoders // Advances in Systems Science. Springer International Publishing, 2014. Vol.240, P.251–259.

[6] M.A. Ullah, K. Okada, H. Ogivara. Multi-Stage Threshold Decoding for Self-Orthogonal Convolutional Codes. IEICE Trans. Fundamentals, Vol.E93-A, Nov. 2010. No.11, P.1932–1941.

[7] M.A. Ullah, R. Omura, T. Sato, H. Ogivara. Multi-Stage Threshold Decoding for High Rate Convolutional Codes for Optical Communications. AICT 2011: The Seventh Advanced Int. Conf. on Telecommunications, 2011. P.87–93.

[8] Zolotarev V.V., Ovechkin G.V., Shevlyakov D.A. The Performance of Multithreshold Decoder over Fading Channels // 2015 International Siberian Conference on Control and Communications (SIBCON). Omsk, 2015.

[9] Viswanathan M. Simulation of Digital Communication Systems Using Matlab [eBook] – Second Edition, 2013.

[10] Ovechkin G.V., Shevlyakov D.A. Efficiency of multithreshold methods of error correction in communication channels with fading // Successes of contemporary radio engineering, M.: Radioengineering, 2014. No.6, P.37–43.

[11] Giosic S. Advanced wireless communications. 4G technologies // S. Giosic Wiley & Sons. 2004. 878 p.

[12] Zolotarev V.V., Ovechkin G.V., Shevlyakov D.A. Research of multithreshold decoder efficiency under mutual usage with space-time coding // "Digital digital processing and its application – DSPA-2015". Moscow, 2015. P.83–87.

[13] Scaglione A., Stoica P., Barbarossa S., Giannakis G.B., Sampath H. Optimal designs for space-time linear precoders and decoders // IEEE Trans. Signal Process. 2002. Vol.50, No.5, P.1051–1064.

Performance of Communication Protocols Under Transfer a Group of Acknowledgment Frames

Kushnazarov Farruh I
Department "Information systems in railway transport"
Tashkent Institute of Railway Transport Engineers
Tashkent, Uzbekistan
k.farruh@bk.ru

Phd, Professor Yakovlev Valentin V.
Department "Information and Computing Systems"
Federal state budgetary educational institution of higher education "St. Petersburg State University of Railways of Emperor Alexander I»
Saint-Petersburg, Russia
jakovlev@pgups.ru

***Abstract*— Increase communication protocols performance continues to be a hot topic, which is reflect in the emergence of new standards and the numerous research projects in this area. This article discusses one option to improve performance of link layer protocol based on the sending receiver group positive acknowledgment to the source of the message. The paper shows the definition of productivity and communication protocols, and the factors affecting their performance. The calculations that determine the number of duplicate acknowledgment receipts under the condition that the sender of the message will be deliver acknowledgment guaranteed. The results of the analytical and simulation in the Matlab environment, confirming the accuracy of the results.**

Index Terms — link-layer protocols, performance, the actual speed, channel utilization, channel noise, computer networks, group of acknowledgment

I. INTRODUCTION

A Nowadays, the correctness of the transfer via communication channels plays an important role. This also reflected in numerous standards and requirements of the system to the data channels [4]. At the same time, analysis of trends in the development of data transmission systems indicates a lack of methods to assess the performance of link-layer protocols, which are mainly aimed at improving the hardware and software components, as well as developing recommendations for the setting of critical parameters of link layer protocols (frame size, timeout and et al.).

Wireless communication channels (local wireless networks, mobile communication links, satellite links, etc..), objectively are the most noisy and therefore unreliable, aided by many factors, such as noisy from microwave source, working almost in the same frequency range [1, 7]. If one bit is equal to the probability of distortion of BER (Bit Error Rate), then the probability that the n-bit frame received correctly, is ~ (1-BER) ^{n} [5]. For Ethernet-frame maximum length at BER=10^{-4} probability of error-free delivery is less than 30%, and at BER = 10^{-5} is distorted already one of 9 frames [11].

Also distortion of frames transmitted over wireless channels is very likely also to their loss.

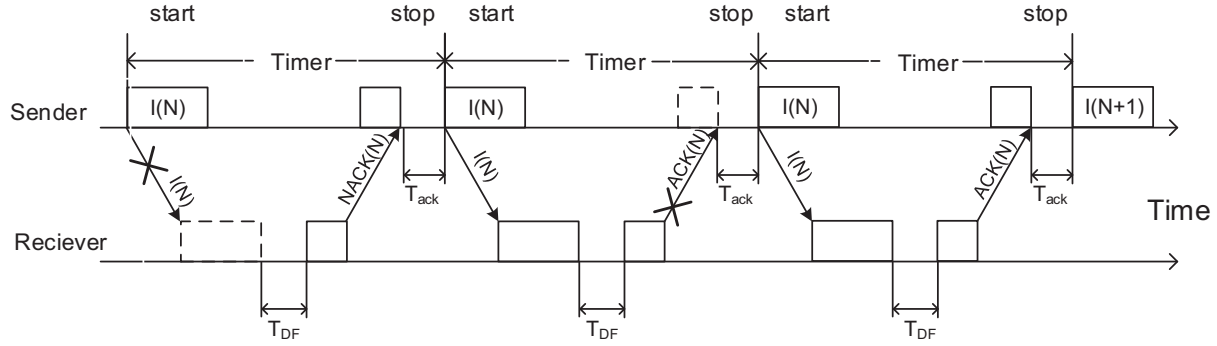


Figure 1. Driving standby in the case of delivery of a distorted picture

In this paper, the process of error is consider based on the model of the binary symmetric channel (BSC). When sending messages by BSC error may occur in communication with each bit BER probability, regardless of any errors in other bits. The error is to replace the character 0 to 1 or 1 to 0. The bit errors are the cause of deterioration of the quality of communication that leads to a distortion in the speech communication channels unreliable transmission of the message, or reduce the nominal data throughput. They are determined by statistical parameters and standards on them and, in turn, determined the corresponding probability of implementation of these standards [1, 2, 4]. These rules listed in the recommendations of standard

ITU-T G.821, G.826, M.2100, M.2110 and M.2120 M.2100 The standard provides a classification of channels on the criterion of error rate. [2]:

- low noise level - BER $<10^{-6}$;
- average noise - $10^{-6} \leq$ BER $<10^{-3}$;
- high level of noise - BER $\geq 10^{-3}$.

Type Style and Fonts

Wherever times specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts. True Type 1 or Open Type fonts are required. Please embed all fonts, in particular symbol fonts, as well, for math, etc.

II. Sending Groups Receipts

For the analysis of the transmission scheme in connection two-point look at the most simple method of transmission – Send and Wait (SW), which may be a special case of the sliding window when the window size is 1. The source that sent the frame, waits for the receipt of the recipient of the message and only then sends the next frame (or It repeats the frame if it has been distorted).

SW provides that upon receipt of each frame sent confirmation to return. If within a certain period of time no acknowledgment received, it assumed that the frame received incorrectly and should be send again.

Consider the algorithm of standby (Figure 1):

Source sends messages I (N)-frame (data frame) and starts a timer.

- The recipient in the preparation of I (N) frames' checks whether the frame is distorted. If the picture distorted, the receiver sends a Negative AcKnowledgment (NAK).
- Source, receiving a NAK, resets the timer and sends again I (N) - the frame and Launches timer.
- The recipient in the preparation of I (N) frames' checks whether the frame is distorted. If the frame not distorted, the receiver sends a positive ACKnowledgment (ACK).
- Source waits for a certain time (timeout). If during this time a positive acknowledgment does not come (ACK - receipt lost from - for interference at channel Communication), then a source reset timer and sends this as I (N) - the frame again. If this situation repeated several time a row, the transmitter suspends transmission, considering that occurred catastrophic renouncement at channel.

The study of previous work to determine the performance of communication protocols, the following conclusions obtained:

1. AT [3, 9] in productivity ionic communication protocols large role assigned hardware component;
2. In the source [6] has been defined "efficiency data about the stakes"; at this We were considered asynchronous protocol AP-70 and a synchronous protocol, and BSC were not considered data transfer modes;
3. In the foreign literature [2] determined notion recycling channel communication is thus not dependent on the size of transmission data level interference at channels communication.

In this regard took definition performance communication protocols (throughput communication protocols) as the rate of successful delivery of the payload of the message source to the message recipient. [5] Under the useful data refers to data that does not include overhead, repeated shots (if error is detected, the message recipient again requests a frame), positive or negative delivery receipt or NDR to the recipient of the message frame, etc.

These parameters may be sending a ACK groups. We calculate the delay time for the case where the frame came from the source of the message to its recipient (Figure1, sending frame N for the second time), but there was a distortion or loss of ACK.

For example, consider the parameters of Ethernet protocol families, namely Ethernet DIX frame types, with a maximum resolution of standard size for a receipt with a minimum length of 64 bytes of the frame without preamble. According to the standard packet interval between frames (Inter Frame Gap, IFG) should be set to 9.6 microseconds. [9] To send a frame with a minimum size (46 bytes of data, 18 bytes of overhead, Preamble of 8 bytes, 46 + 18 + 8 = 72 bytes = 576 bits) at a nominal rate of 10 MB/s during frame transmission will be 57.6 microseconds, if you add a pause - 57.6 + 9.6 = 67.2 microseconds. At the maximum frame size standard (1 500 bytes of data, 1500 + 18 + 8 = 1526 bytes = 12208 bits), the transmission will be 1220.8 microseconds, while the frame repetition interval is equal to 1220.8 +9.6 = 1230.4 microseconds.

If the frame received by the recipient of the message without distortion, and the receipt came from the distortion, then resend the frame and receiving an ACK will take time, equal to 67.2 = 1230.4 + 1297.6 microseconds. In case of loss of the ACK message source when the timer re-send the frame. The timer is set by the standard as the RTT (RTT - Round-Trip Time) - the time of turnover:

$$RTT = 2 * t_{net},$$

where t_{net} – is a network - data transmission time frame with the maximum size allowed by the protocol standard.

Then, the delay time is equal to 2 * 1230.4 +1230.4 + 67.2 = 3758.4 microseconds. Thus, the delay time may increase significantly, which is unacceptable in the conditions necessary to maintain the efficiency and reliability of data delivery, especially in the critical control systems.

In this regard, the article deals with the sending of receipts for the group to achieve a guaranteed source of receiving a receipt message.

The probability of a successful transmission of the frame p:

$$p = (1 - BER)^N \quad (1),$$

Where

N - the length of the frame.

Define the number of attempts to retransmit frames a given probability of successful transmission – u [1]:

$$P_{spec} = 1 - p^i,$$

where, P_{spec} – given the probability of the trusted of delivering frames, p^i – the probability in i step, thence,

$$u = \frac{\lg(1 - P_{spec})}{\lg p}$$

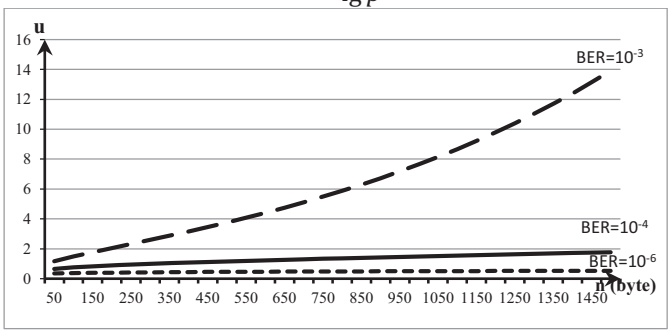


Figure2. The dependence of the number of required retransmission to a successful transmission of the frame at different values of BER depending on the frame size

Figure 2 for the P_{spec} = *0.97* and the BER = 10 $^{-3}$ with increasing frame size dramatically increases the number of repetitions. When BER <10 $^{-3}$ frame number of repetitions does not exceed two. If the calculations to take BER = 10 $^{-2}$, number of repetitions will increase dramatically, since the size of 500 bytes.

III. Determination "Real Speed Transfer Mineral Data"

Consider a situation where a receipt duplicated, ie source sequentially sends to the recipient group receipts. Based on calculations by the formula (2), we find that the size of the data frame is less than 100 bytes number of repetitions does not exceed two (Figure 2). In this regard, it can assumed that at least one of the two receipts source of the message received reliably.

Using the proposed approach [5] to calculate the RTT time and timeout:

$$RTT = \frac{N}{R} + \frac{2 * N_{ack}}{R} + \frac{S}{R_{sig.prop}} + 2 * T_{int} \quad (3),$$

where R - the nominal speed of link layer protocol,

T_{int} — Belt in the interval between shots,

$T_{int} = T_{DF} + T_{ack}$,

T_{DF} — The time spent on processing the data frame,

T_{ack} — The time spent on processing the acknowledgment

N_{ack} — number of bits in the frame of the receipt,

$R_{sig.prop}$ — signal propagation speed on the transmission medium,

$$R_{sig.prop} = c * \mu,$$

where c - the speed of signal propagation in vacuo

μ - ratio of the real propagation speed in vacuum.

For the most common cable systems have the following value μ:

Coaxial cable - μ = 0,66;

Twisted pair cat3 - cable - μ = 0,65-0,71;

Fiber-optic cable - μ = 0,66-0,78.

The channel noise conditions to evaluate the transmission time of the N bit data frame, we introduce the actual data transfer speed V in the case of sending a receipt. Thus it is necessary to take into account the likelihood of errors in the receipt [5]:

$$V = \frac{N-C}{T} * (p * p_{ack}) \quad (4),$$

where C - number of overhead bits in a frame,

$p_{ack} = (1 - BER)^{N_{\text{квит}}}$,

T - frame transmission time, including the time for the transmission of receipts,

$$T = \frac{N}{R} + \frac{N_{ack}}{R} + 2 * \frac{S}{R_{sig.prop}} \quad (5).$$

If you send two receipts in the expression (5) we suggest that one of the source of the message receipts will unerringly. V, and T are defined as

$$V = \frac{N - C}{T} * p \quad (6),$$

$$T = \frac{N}{R} + 2 * \frac{N_{ack}}{R} + 2 * \frac{S}{R_{sig.prop}} \quad (7)$$

The parameters in the expression (4) and (6) defined as follows: *C = 26 bytes, N_{ack} = 72 bytes, R = 10 Mb/s, $R_{sig.prop}$=1,98*10^{8} m/s, S = 100 meters*.

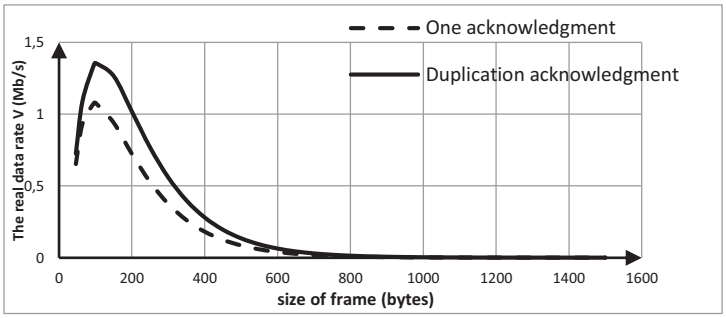


Figure 3. The actual useful data transmission rate with different amount of receipts and BER = 10 $^{-3}$

From Figure3 and calculations, (4) and (6) is visible gain in speed, since there is a possibility of distortion of the receipt, especially in noisy communication channels with BER = 10 $^{-3}$. With an increase in frame size reduces the information transfer rate also decreases and the speed difference between the two methods.

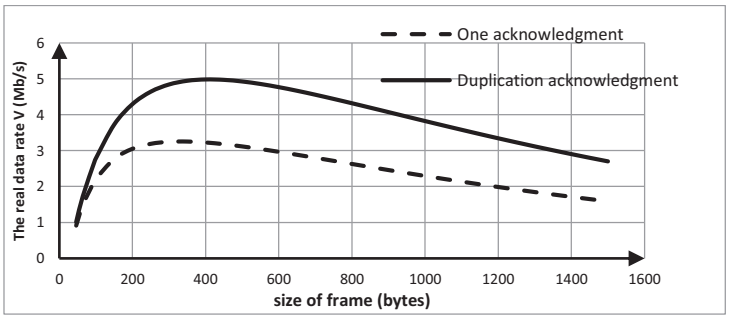


Figure 4. The actual useful data transmission rate at BER = 10^{-4}

With an increase in the frame size to 426 bytes (we assume that this optimum point [5], at which the maximum useful data transfer rate), the rate increases, there is a tendency gain in payload data transfer speed means receipts duplication and reaches its maximum (the difference between the rates in 426 bytes up to 1.81 MB / s), which is associated with reliable delivery receipts.

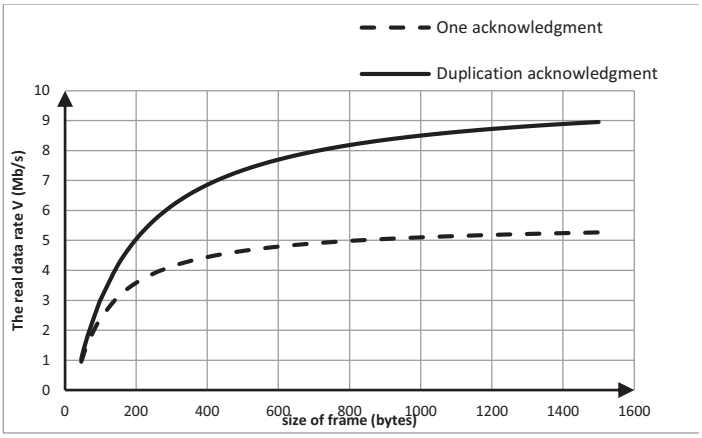


Figure 5. The actual useful data transmission rate at BER = 10^{-9}

Figure 5 is set to the BER = 10^{-9}, considered communication channel with low interference. Although the communication channel with low interference, the graph and calculations of expressions (4) and (6) the gain in speed is noticeable, confirming the efficiency of sending slips duplicated. However, the model cannot account for the load on the channel, as the channel is loaded with duplicate receipts. In order to determine the load of the communication channel, then consider the simulation model in the Matlab environment, using the Simulink library and the environment Communication Toolbox. The results obtained will be close to the actual figures used in the industry, as the Communication Toolbox uses the settings and parameters of the most popular link layer protocols.

IV. Experimental Research on the Effectiveness of Duplication Sending Acknowledgments

In the previous section for solving estimation link layer protocols performance when sending duplicate receipts were investigated mathematical models to calculate the actual transmission rate of user data via communication channels. To assess the reliability of the results in real data environment simulation in Matlab environment was performed (Simulink) work channel in standby mode. Using Matlab simulation model developed that realizes the data transmission mode using the method of "standby"

The basis chosen as a model of IEEE 802.11b standard data. The model (Figure 6) is an implementation of a data transmission system in a noisy communication channel, providing a nominal data rate of 11 Mbit / s [10]. In the model, Simulink built-in packages and Communication Toolbox used.

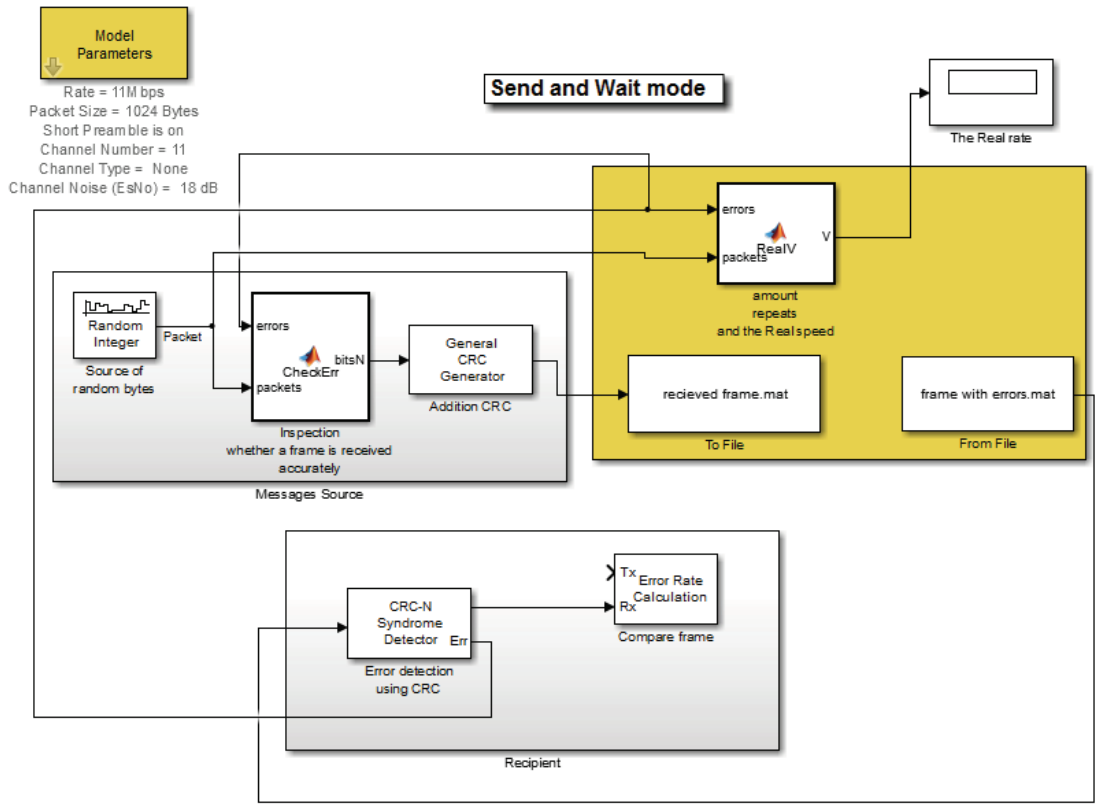


Figure 6. General view of the simulation data model the IEEE 802.11b standard

In accordance with the model of Figure 6 comprises the following blocks: the source of the message, the communication channel, the message recipient, validation unit frame reception and error calculator frames. Posts Source creates a frame with pseudo-random numbers using the Random Integer block. The Random Integer parameters given frame size and frequency of administration of frames per second (depending on the nominal rate of Protocol 11 Mbit / s). These data are converted (block

CheckErr) bits in the data block and transmitted to "link" (if the frame is the first). If subsequent frames transmitted, it is first checked faithful reception of the previous frame. In addition, a timer is started CheckErr unit after each frame transmission. The communication channel received bits written to the file. This file frame processed further in order to change some bits in error to simulate frames. Thus, when sending a message size 1,000,000 bytes BER = 10^{-3}, 8000 bit errors can be grouped in frames at their discretion to learn the reaction system at the single errors on multiple errors and packet errors. The model assumes that the maximum number of transmission attempts is five, that is the fifth frame is attempting to accurately transmitted (the number of attempts is chosen pseudo randomly). Block "number of repetitions and the actual speed," depending on the number of frames of repetitions determines the real speed to send the message using the formula:

$$V_{real} = \frac{packets*N*8}{time} \quad (8),$$

where V_{real} — he actual transfer rate,,
packets — the number of frames in the message,
N — frame size,
time — time spent on transmitting the message.

Interpretation of the measurement results

The output data block "real speed link protocol" (Figure 6) collected in a file for comparison with the results of analytical modeling. The maximum number of retries on errors in frames on the standard was set at fifteen. For example, the first time the source of messages sent to the frame, the recipient of the message, an error is detected in the frame, the source sends a NAK message, in turn, the source sends the frame again - may be the most fifteen such repeats.

Figure 7 shows a comparison of the results obtained for the analytical and simulation models at the level of BER = 10^{-3} to standby. Thus in 3000 it was generated channel frames including both correct and erroneous frames. Frame size, ranging from 46 bytes to 1 increased in increments of 500 bytes 2 bytes.

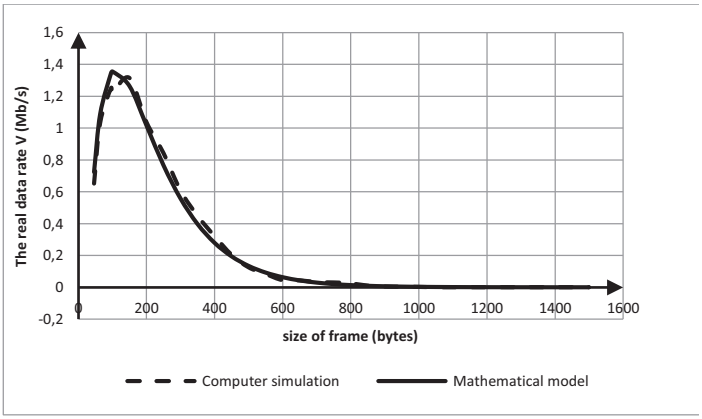


Figure 7. Comparison of the results of the analytical and simulation at the level of BER = 10^{-3} (Standby)

It can see that the simulation results are in good agreement with the results of analytical modeling. The difference between the rates in the two models is on average 1.02%.

V. Conclusion

You must submit the IEEE Electronic Copyright Form (The developed method of calculating the actual speed when you send two receipts based on the noise level registered in the communication channels, frame size and length of communication channels. Determined the method of calculation of the timer for the data transmission method. With this, the actual conditions for optimizing the data rate of standard data link layer protocols depending on the frame size and level of interference in the communication channels. The conducted simulation data using send two receipts can serve as the basis for a hardware implementation of the method of transmission. Interpret the results of the simulation show good agreement.

References

[1] Chatzimisios P., Vitsas V., Boucouvalas A., Tsoulfa M. Achieving performance enhancement in IEEE 802.11 WLANs by using the DIDD backoff mechanism INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS Int. J. Commun. Syst. 2007; 20: 23-41 Published online 16 May 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002 / dac.811

[2] Halsall F. Data Communications, Computer Networks and Open Systems. Addison-Wesley, 1996.-907pp.

[3] Bashtannik NA Evaluation performance computing complex information measurement and control system for special purposes. Dis. ... Kand.teh. Sciences. - Astrakhan - 2010. - 176 p.

[4] GOST 17657-79. Data transfer. Terms and Definitions. - Moscow: Publishing House of Standards, 1987. - 35

[5] Kushnazarov FI Development and research of models of performance evaluation of communication protocols for channel interference. Dis. ... Kand.teh. Sciences. - St. Petersburg - 2016. - 159 p.

[6] Letsky EK, Pankratov V., Yakovlev VV Information technologies in rail transport. Textbook for high schools zh.d.t. CMD M MEA Russia, 2000 - 680 p.

[7] Losev NN, VN Dmitriev layer protocol performance analysis medium access in wireless local area networks. Bulletin of the GTU. Ser. Management, Computer Science and Informatics 2011. № 2

[8] Configuring TCP Maximum Segment Size Electronic resource]. - Internet page. - mode

[9] Olifer, VG Computer networks. Principles, technologies, protocols / VG Olifer, NA Olifer. - Moscow, 1999

[10] 802.11 Working Group [electronic resource]. - Internet page. - Access: https://manta.iee.org/gropus/802/11

[11] Hamadulin EF Methods and means of measurement in telecommunication systems: a manual for schools / National issledovat. University (MIET). - Moscow: Yurayt 2014

.

Study of robust stability and quality of the two channel control system

Aizhan Erulanova, Gennadii Shadrin, Natalia Arinova
Faculty of Information Technology and Power Engineering
D. Serikbaev East Kazakhstan State Technical University
Oskemen, Kazakhstan
A_Erulanova@bk.ru, Shadrin.g.k@yandex.ru, NVarinova@mail.ru

***Abstract*. We discuss the problem of robust stability of the two-channel control system of object having the parametric uncertainty in its mathematical description. Uncertainty is concentrated in the coefficients of the equations. The regulator has been synthesized by the method of compensation of the dynamics of the object and disturbances for two-channel control system. We conducted model studies of the influence of changes in the coefficient of object transmission through two control channels on the robust stability and the quality of ACS. The system is robustly stable and the satisfactory quality of control is maintained in it.**

Index Terms - robust stability, parametric uncertainty, disturbances, automatic control system, weigher, method of compensation of dynamics of the object and disturbances, controlled object, two-channel control system, standard filter.

I. Introduction

In the real control tasks, the uncertainty is unavoidable, and the systems used must be operable in the presence of uncertainty. Such control is called robust control. If the model describes a physical object, then, as a rule, its parameters are not known precisely, and in many cases their values cannot be available in principle, because they can be changed during operation. It is a common case, when the equations structure describing the movement is assumed to be known exactly. Uncertainty is concentrated in the coefficients of the equations. In such situations, we can speak of parametric uncertainty [1].

The system is unavoidably under the influence of the disturbances, which make the unstable system inoperative.

The precise identification of the object model parameters is a difficult task. The structure and parameters of the object is always defined with a certain error. Also, objects parameters in the operation may vary within certain limits. This can lead to loss of stability and quality of ACS. Therefore, the creation of robust control methods is essential [2].

We will be interested in the question: whether the disturbed system is robust stable, whether the system provides a given value of the selected index of the quality - robust quality.

II. Description of Weight Metering Unit

The robust stability of the system is considered on the example of the previously developed two-channel metering device [3]. In the control system, this system created a mathematical model of two-channel object with a transport delay [4]. This model looks like:

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} f_x \\ f_y \end{bmatrix}, \quad (1)$$

In the course of the regulator development, the reverse model (2) of the controlled object and standard filter of the closed system (3) were needed.

$$\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \cdot \begin{bmatrix} \dot{x} \\ y_з \end{bmatrix}, \quad (2)$$

$$\begin{bmatrix} \dot{x}_ф \\ y'_ф \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_3 & \Phi_4 \end{bmatrix} \cdot \begin{bmatrix} x'_ф \\ y'_ф \end{bmatrix}. \quad (3)$$

where x, y, u – respectively, $n-, m-, m-$ dimensional vectors of state, output and control; f_x, f_y – vectors of disturbance influences on x and y; A, B, C – matrixes of the relevant dimensionalities; $y_з$ - vektor of setting at the system input; E, F, G, H - matrixes of the reverce model of the controlled object.

The regulator structure is determined by the equations of the form [5-8]

$$\left.\begin{aligned} \dot{x}_ф &= (\Phi_1 + \Phi_2\Phi_3)x_ф + \Phi_2\Delta y \\ u &= (G\Phi_1 + H\Phi_3) + G\Phi_2\Phi_3 x_ф + G\Phi_2\Delta y \end{aligned}\right\}. \quad (4)$$

For complete confidence in the system workability it is necessary to find out whether the system will maintain stability in the possible variations of its parameters relative to their calculated values.

We believe that the modified matrixes of the control object remains constant in the course of the system operation, i.e., system is stationary during operation. Let us consider the

interval uncertainty of object matrixes of the controlled object in the following form:

$$A_n = A + \Delta_A,$$
$$B_n = B + \Delta_B, \quad (5)$$
$$C_n = C + \Delta_C.$$

where $\Delta_A, \Delta_B, \Delta_C$ - matrixes determining the intervals of change of input matrix coefficients.

The regulator is calculated for the input matrixes of the controlled object.

The analytical study of robust stability and robust quality is complex quite and not demonstrative. We carried out our studies by computer simulation of a real system. The disadvantage of the computer simulation is that a particular object is simulated. In our case, the results of the study can be extended to weighers with other characteristics. This is confirmed by that the size of the dispenser mechanical components are associated with the elements of matrices by smooth single-valued dependencies.

When modelling the system, the matrix transferring function of the controlled object is taken in the general form:

$$\begin{bmatrix} Y_1(p) \\ Y_2(p) \end{bmatrix} = \begin{bmatrix} W_{11}(p) & W_{12}(p) \\ W_{21}(p) & W_{22}(p) \end{bmatrix} \times \begin{bmatrix} U_1(p) \\ U_2(p) \end{bmatrix}, \quad (6)$$

where

$$\begin{bmatrix} Y_1(p) \\ Y_2(p) \end{bmatrix} = \begin{bmatrix} k_{11}\frac{1}{Tp+1} & k_{12}e^{-p\tau}\frac{2v}{L}\left(\frac{2(1-e^{-\frac{p\tau}{2}})^2}{\tau p^2}\right) \\ k_{21} & k_{22}e^{-p\tau}\frac{2v}{L}\left(\frac{2(1-e^{-\frac{p\tau}{2}})^2}{\tau p^2}\right) \end{bmatrix} \times (7)$$
$$\times \begin{bmatrix} U_1(p) \\ U_2(p) \end{bmatrix}.$$

in the numerical form

$$\begin{bmatrix} Y_1(p) \\ Y_2(p) \end{bmatrix} = \begin{bmatrix} 0{,}135\frac{1}{Tp+1} & 0{,}08\frac{2v}{L}\left(\frac{2(1-e^{-\frac{p\tau}{2}})^2}{\tau p^2}\right) \\ 0 & 0{,}321\frac{2v}{L}\left(\frac{2(1-e^{-\frac{p\tau}{2}})^2}{\tau p^2}\right) \end{bmatrix} \times$$
$$\times \begin{bmatrix} U_1(p) \\ U_2(p) \end{bmatrix}. \quad (8)$$

When modelling, the following parameters are taken:

$\tau = \frac{L}{v} = 4\,c$ - time of delay;

$L = 1\,м$ - conveyor belt length;

$v = 0.250\,м/c$ -conveyor belt movement speed.

In the course of study of the robust stability, the transmission factors $k_{11}, k_{12}, k_{21}, k_{22}$ were changed by ±10%.

The weigher is shown in Figure 1.

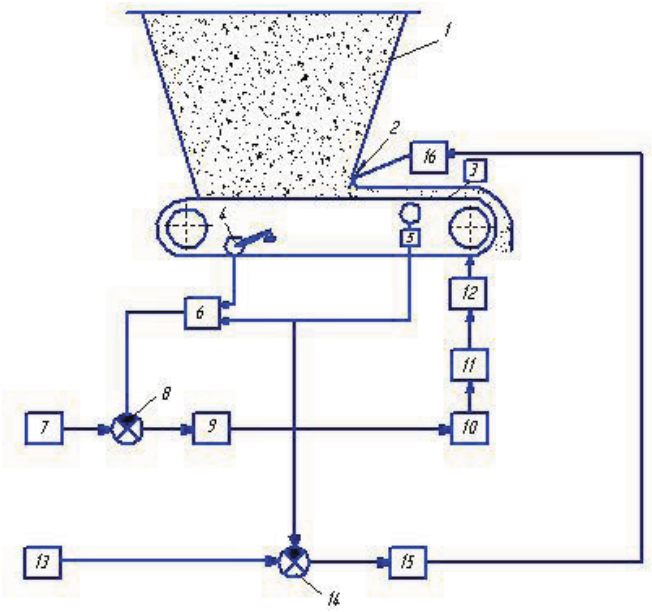


Fig. 1. Weigher functional scheme

Continuous weigher depicted in Figure 1 comprises: 1 - hopper 2 - flap 3 - conveyor 4 - conveyor belt speed sensor 5 – weight meter 6 - multiplier 7 - performance reference-input element 8 - performance comparator 9 - performance controller, 10 - device for controlling conveyor belt speed, 11 - motor 12 - conveyor reducer, 13 - linear load reference-input element, 14 - linear load comparator, 15 - linear load controller 16 - flap drive. [9-10]

III. Simulation Results

Two-channel control object simulation schemes for variation of coefficients in the first channel by ± 10% in program VisSim are shown in Fig.2 and Fig.3 [11]. System transient characteristics is presented in Fig.3 and Fig.5. Control actions are fed from multivariate regulator, scheme of which is shown in Fig.6. Table 1 shows values of coefficients of object transmission through two control channels, procedure of experiment, channel controlled object coefficients changes ± 10%, and the form of the corresponding transfer characteristics obtained by computer simulation.

TABLE I. RESEARCH OF CONTROL SYSTEM ROBUST STABILITY

№	Term of the experiment	Values of controlled object transmission coefficients		
		K_{11}	K_{12}	K_{22}
1	Channel 1 +10%	0.148	0.08	0.321
2	Channel 1 -10%	0.121	0.08	0.321
3	Channel 2 +10%	0.135	0.08	0.353
4	Channel 2 -10%	0.135	0.08	0.289

Table 1 shows that if there are different combinations of controlled object transmission coefficients in the first and the second channels (± 10%), the system keeps stability and good governance. This proves, that the applied synthesis method provides robust stability of the control system.

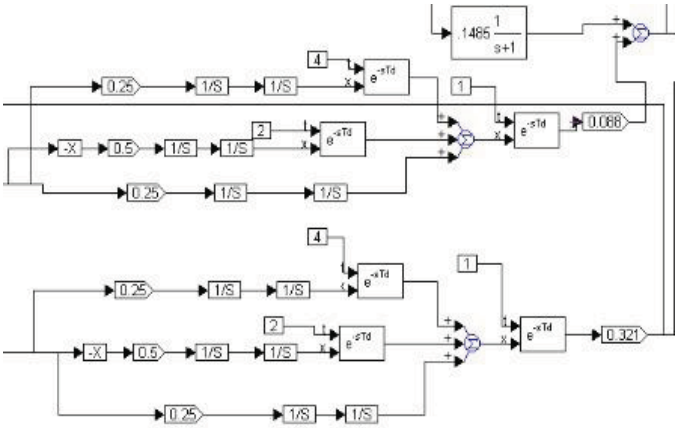


Fig. 2. Controlled object model during measuring transmission (Channel 1+10%)

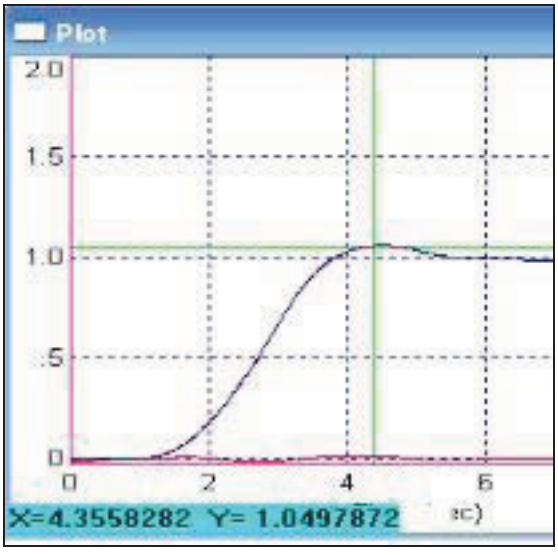


Fig. 3. System transient characteristics (Channel 1 +10%)

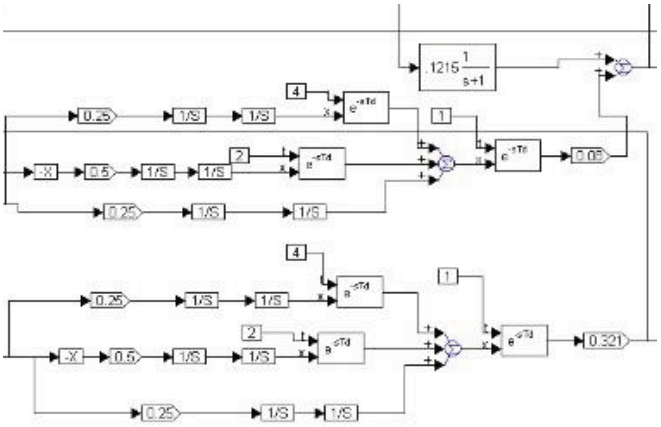


Fig. 4. Controlled object model during transmission coefficient changing (Channel 1 – 10%)

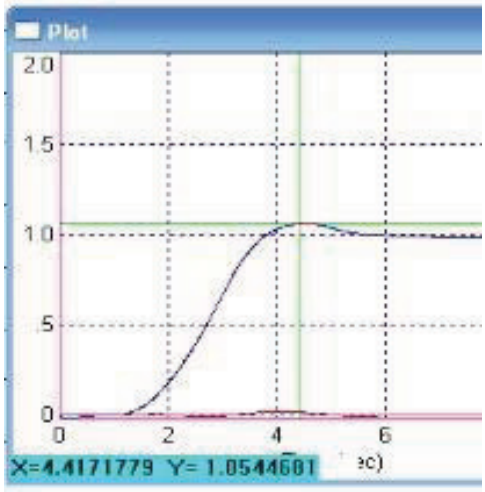


Fig. 5. System transient characteristics (Channel 1 –10%)

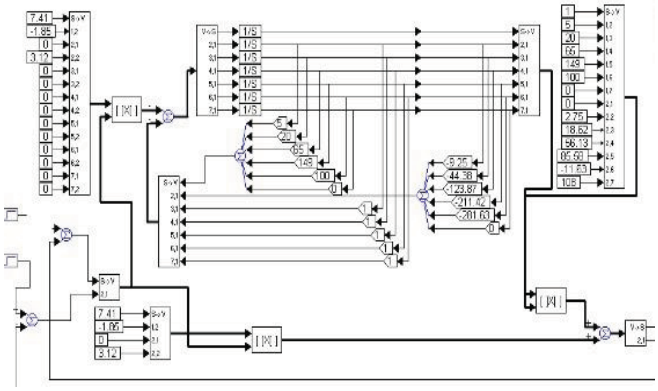


Fig. 6. Regulator of multidimensional system

CONCLUSION

The robust stability of the two-channel weigher control system is considered. It is shown, that if transmission coefficient combinations of an object of control in the first and the second control channels (± 10%) are different, the system is

robust stable and saves a good quality of regulation. Since actual functioning systems should not only be stable but also possess enough intensively damping transients. Research at various variations of the controlled object transmission coefficient shows, that the quality of management systems remains satisfactory. Regulation time is ≈4,61sek, dynamic error does not exceed 5%.

REFERENCES

[1] B.T. Polyak Robust stability and control / B.T Polyak, P.S. Shcherbakov - M.: Scienece, 2002. – 303 p.

[2] V.Ya Rotach. Theory of Automatic Control, M.: Publishing House MEI 2008.

[3] Simulation of two-channel bulk materials metering system / A.E. Erulanova, G.K. Shadrin, N.V. Arinova, I.A. Asangaliiev // Proceedings of the 9th International Conference «Application of information and Communication Technologies». – Roston – on – Don, 2015. – pp. 335-338.

[4] Zh.I. Batyrkanov, Mathematical model of continuous weigher / Zh.I.Batyrkanov, A.E.Erulanova // Kyr. State. Tehn. University. - 2012.- No.26.- pp. 140-144.

[5] Application of compensatory regulation algorithms by powerful drives by the example of the plate feeder drive / G.K.Shadrin, M.G.Shadrin, A.N.Shuklin, O.D.Masaylo // Proceedings of the 1st Intern. Scientific and Engineering. Conf. «Energy, environment, energy saving».- Ust-Kamenogorsk, 2005.- pp. 250-251.

[6] Y.M. Abdeev, Development of automation systems of technological processes in the production of lead and zinc based on anti-jamming algorithms / Yu.M.Abdeev, V.V.Terehin, G.K.Shadrin // Non-ferrous metal.- 1990. – No. 12.- pp.36-38.

[7] G.K. Shadrin, U.V. Mosolkov, M.G. Shadrin, 2013. Application of the method of object dynamics compensation and disturbances for the creation of level control algorithm in flotation machine. Journal of the East Kazakhstan Technical University after D.Serikbaev. Computational technologies. Institute of Computing Technologies, Siberian Branch of the Russian Academy of Sciences. Part 3: Ust-Kamenogorsk, pp. 49-54.

[8] G.M. Mutanov, Comparative analysis of the methods of development of automatic control systems / G.M.Mutanov, G.K.Shadrin, N.V.Arinova // Journ. Vostoch.-Kazahst. State. Tehn. Univ. after D.Serikbaeva.- 2010.- No. 2.- pp. 94-100.

[9] RK Patent No.24575. Continuous Dosing // Inventions. 2011. No. 9.

[10] RK Patent No.26500. A method of continuous dosing of bulk materials // Inventions. 2012. No.12.

[11] N.V. Klinachev Simulation of systems in program VisSim: help desk in the Russian lang. / N.V.Klinachev.- Chelyabinsk, 2001.- Access: http://model.exponenta.ru/ vsmhlpru.zip.- Screen Title.

[12] V.O.Andreev, A.I.Suzdaltsev, S.E. Tinyakov., 2002.Control system for technological complexes with transport delay. Automation and telecontrol. No.5. pp. 184-189.

[13] V. O.Andreev, A. I.Suzdal'tsev, S. E.Tinyakov, 2002. Control Systems of Processes with Transit Delays Look Inside Get Access. Automation and Remote Control, Volume 63, Issue 5, pp. 861-865.

Development a method for routing in heterogeneous telecommunication systems under uncertainty using fuzzy logic

Q.T. Tran, T.H.T Vo, V.N. Dmitriev, A.A. Sorokin
dept.: «Telecommunication» of the Astrakhan State Technical University
Astrakhan, Russia
E-mail address: sviaz@astu.org

***Abstract* - Existing approaches to routing in heterogeneous telecommunication systems under uncertainty do not take into account all the many factors that affect traffic distribution in the network and packet transmission between two nodes. The main problem is the complexity of building accurate mathematical models of the network for many reasons, the main of which is a complex system of mutual influence of the parameters. Given the larger number of parameters, the network status will enable us to provide a more accurate picture of the traffic distribution in the network, a picture of the occurrence of collisions and download sites. The availability of such information at the moment of making decision on selection of the optimal route of data transmission in accordance with some criterion to choose the optimal route. To solve this problem we need to propose a mathematical formalism to account for the collected parameters and decision which route is optimal from the point of view of these criteria. In this paper, an approach to the problem of routing in in a heterogeneous telecommunication systems under uncertainty based on fuzzy logic.**

Keywords – routing, telecommunication system, fuzzy logic, routing, heterogeneous.

I. Introduction

Currently there is a proliferation of data transmission systems with heterogeneous topological structure. Under the heterogeneity of topological structure refers to the availability in a distributed system transmission and processing of wired and wireless component. The widespread use of wireless means of transmission and processing of information leads to complicated topological structures of such systems. The complexity of the topology increases the factors affecting the efficiency of identifying and maintaining optimal relevance, according to specified criteria routes of information transfer.

As the decision on the choice of the optimal route is proposed to use fuzzy logic. Fuzzy logic is based on fuzzy set theory, which was started in 1965 by American mathematician Lotfi Zadeh. The main advantages of fuzzy logic is the possibility of considering multiple factors when deciding without building accurate mathematical model of the controlled system and the ease of writing rules fuzzy implication, which is set on intuitive human terms. We can say that fuzzy logic is a cross between a complicated mathematical analysis and simple intuition. You only have to create rules of conduct in the form of fuzzy conditional propositions of the form IF... THEN. Therefore, the fuzzy logic can be applied to control the routing process in wireless mesh networks, where it is necessary to consider many factors when choosing the best route.

II. Statement of the Problem

In the process of determining the optimal route of information transfer in a wireless network according to the QoS requirements, as the most important optimality criteria are: available bandwidth, delay of transmission packet, workload of buffer, number of intermediate nodes, probability of packet loss.

The structure of the developed fuzzy controller to determine the routing metrics in the heterogeneous network shown in Fig. 1, where: x_1 - available bandwidth of communication channels between a given pair of nodes, x_2 - delay of transmission packet, x_3 - workload buffer, x_4 - probability of packet loss, x_5 - number of intermediate nodes, R_k - set of fuzzy inference rules ($k = 1...N$, where N is the number of rules); B - output fuzzy set that determines the rating of the route.

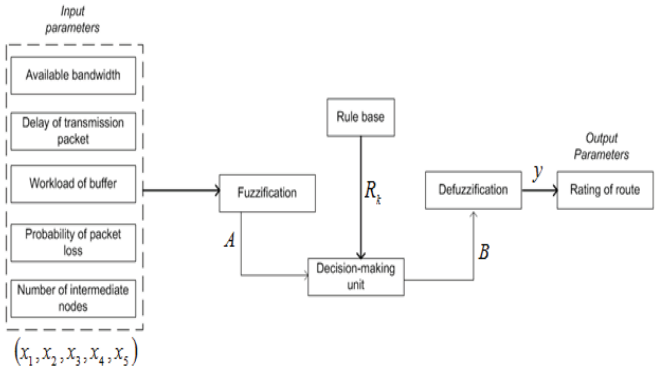


Figure 1. Structure of the developed fuzzy controller

Each input parameter of the controller is mapped to a linguistic variable with three terms:

- Available bandwidth: low, medium, high;
- Delay of transmission packet: bad, average, excellent;
- Workload of buffer: low, medium, high;
- Probability of packet loss: bad, average, excellent;
- Number of intermediate nodes: small, medium, high.

The output of each rule is an implication of the linguistic variable "RATING OF ROUTE", a set of values which also consists of three terms: "LOW", "MEDIUM" and "HIGH". The value "LOW" indicates that the route is poorly suited for transmission of traffic, or even not suitable, a value of "MEDIUM" indicates that the route is generally suitable for information transfer, and "HIGH" indicates that the route is best suited for the transmission of this information.

The total number of fuzzy rules N = $3^5 = 243$. The rule base is a set of fuzzy rules of the form:

$$R_k : if \ \left(x_1 = A_1^k \wedge x_2 = A_2^k \wedge x_3 = A_3^k ... x_5 = A_5^k\right)$$
$$then \ \ y_k = B_k$$

For example:

If the "Available bandwidth" - "SMALL" and "Delay of transmission packet" is "HIGH" and "Workload of buffer" is "HIGH" and "Probability of packet loss" is "BAD" and "Number of intermediate nodes" is "HIGH", then "RATING OF ROUTE", B_k - "LOW". The fuzzy output is defined thus:

Input:

$X = (x_1, x_2, x_3, x_4, x_5)$, this is

$$A \wedge A = A_1 \times A_2 \times A_3 \times A_4 \times A_5$$

The union of sets of rules, described by the relation:

$$\bigcup_{k=1}^{N} R^{(k)}, R^{(k)} : A^{(k)} \to B^{(k)}$$

Report generation is described by the ratio:

$$y_k \to B_k$$

The input of fuzzy logic controller the signal:

$$X = (x_1, x_2, x_3, x_4, x_5)$$

Membership function of a fuzzy set is equal to:

$$\mu_B(y) = \max_{k=1...N} \mu_{B^k}(y) \quad (1)$$

where:

$$\mu_{B^k}(y) = \mu_{A^k \to B^k}(x, y) =$$
$$= \mu_{A_1^k \times A_2^k \times A_3^k \times A_4^k \times A_5^k \to B^k}[(x_1, x_2, x_3, x_4, x_5), y]$$

Fuzzy implication is given by Mamdani rule:

$$\mu_{A \to B}(x, y) = \mu_R(x, y) = \mu_A(x) \bigcap \mu_B(y) =$$
$$= \min[\mu_A(x), \mu_B(y)]$$

Thus:

$$\mu_{B^k}(y) = \min[\mu_{A_1^k \times A_2^k \times A_3^k \times A_4^k \times A_5^k \to B^k}[(x_1, x_2, x_3, x_4, x_5), \mu_{B^k}(y$$
$$= \min[\mu_{A_1^k}(x_1), \mu_{A_2^k}(x_2), \mu_{A_3^k}(x_3), \mu_{A_4^k}(x_4), \mu_{A_5^k}(x_5), \mu_{B^k}(y$$

(2)

As a result of the expressions (1) and (2), we obtain:

$$\mu_B(y) = \max_{k=1...N} \min\{[\mu_{A_1^k}(x_1), \mu_{A_2^k}(x_2), \mu_{A_3^k}(x_3),$$
$$\mu_{A_4^k}(x_4), \mu_{A_5^k}(x_5), \mu_{B^k}(y)]\}$$

Defuzzification the output value of the controller (route rating) is produced by the method center of gravity:

$$y = \sum_{k=1}^{N} a_k \int_y \mu_{B^k}(y) dy \Big/ \sum_{k=1}^{N} \int_y \mu_{B^k}(y) dy$$

where $\mu_{B^k}(y)$-membership function rules output fuzzy sets of k-th rule on base of rules ; $k = 1,..., N$; a_k - point at which the membership function takes the value 1.

An example of constructing membership functions of terms of the parameter "Available bandwidth" shown in Fig. 2

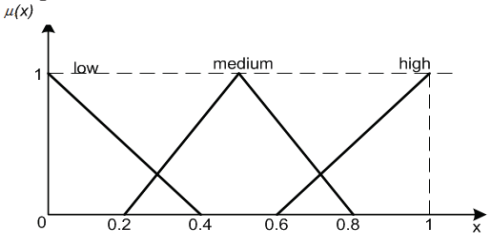


Figure 2. Example of membership functions of terms of the parameter "Available bandwidth"

III. Simulation and Results

The performance of the developed model was verified with the MATLAB environment using the fuzzy Logic Toolbox extension. The model structure is shown in Fig. 3.

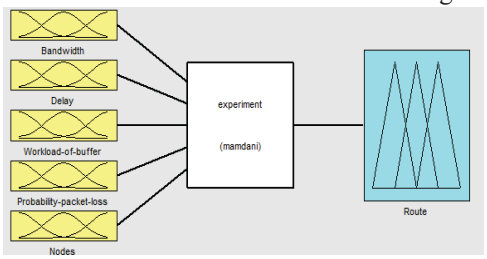


Figure 3. Model structure in MATLAB

The results of the model by assigning route metrics shown in table. 1

TABLE 1. RESULTS OF THE MODEL BY ASSIGNING ROUTE METRICS

NE	AB	D	PPL	WLB	N	RR
1	0.209	0.231	0.197	0.414	0.177	0.159
2	0.367	0.715	0.378	0.250	0.590	0.437
3	0.511	0.231	0.505	0.516	0.676	0.261
4	0.629	0.478	0.612	0.803	0.750	0.500
5	0.726	0.640	0.716	0.654	0.622	0.550
6	0.790	0.780	0.803	0.420	0.697	0.653
7	0.855	0.909	0.888	0.218	0.378	0.844
8	0.984	0.984	0.090	0.984	0.016	0.864

where: NE – Number of experience, AB - Available Bandwidth, D – Delay, PPL - Probability of packet loss, WLB - Workload of buffer, N – Nodes, RR – Rating of route.

Graphically, the results of the model by definition of the metrics is shown in Fig. 4, which shows the dependence of the output values, values of input parameters.

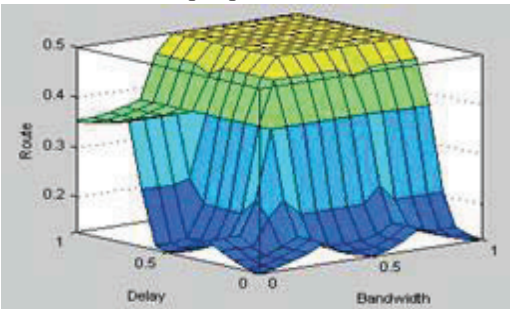


Figure 4. Results from the model in the MATLAB program

IV. CONCLUSION

In this paper, developed by the routing method in a heterogeneous telecommunications systems under uncertainty. This method is based on the use of fuzzy models for calculating route metrics. In the process of creating the model defines a set of input parameters: available bandwidth, delay of transmission packet, workload of buffer, probability of packet loss, number of intermediate nodes, and formed a knowledge base that describes the interactions between the input variables and their impact on the metric value of the route. The developed model and method of routing possibilities for improving the efficiency of traffic management in heterogeneous wireless telecommunication systems.

REFERENCES

[1] Shtovba S. D. Design of fuzzy systems by means of MATLAB. – Hot line – Telecom, 2007. – 288 p.

[2] A.A. Sorokin, V.N. Dmitriev, "Description of communication systems with dynamic network topology by means of model flickering graph", Bulletin of the Astrakhan State. tehn. Univ. Ser.: Management, Computer Science and Informatics, vol. 2, 2009, pp. 134–139.

[3] Zadeh L. Fuzzy sets / L. Zadeh // Information and control. -1965. – V.8. – P.338-353.

[4] A.Kaufmann and M.M.Gupta, Introduction to Fuzzy Arithmetic Theory and Application, Van Nostrand Reinhold, New York, 1991.

[5] G. J. Klir and T. A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice Hall, Englewood Cliffs, NJ, 1988.

[6] G. J. Klir and Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, Upper Saddle River, NJ, 1995.

[7] G. J. Klir, Ute H. St. Clair, and Bo Yuan, Fuzzy Set Theory, Prentice Hall, Upper Saddle River, NJ, 1997.

[8] T. J. Ross, Fuzzy Logic with Engineering Applications, McGraw-Hill, Hightstown, NJ, 1995.

[9] M. S. Stachowicz and M. E. Kochanska, Graphic interpretation of fuzzy sets and fuzzy relations, Mathematics at the Service of Man, edited by A. Ballester, D. Cardus, and E. Trillas, based on materials of Second World Conf., Universidad Politecnica Las Palmas, Spain, 1982.

[10] H. J. Zimmermann, Fuzzy Set Theory and Its Applications, 3rd ed., Kluwer Academic Publishers, Boston, MA, 1996.

Evaluation of the Available Wireless Remote Devices Subject to the Information Impact

Nurzhan Bazhayev, Ilya Lebedev, Irina Krivtsova, Mikhail Sukhoparov, Kseniya Salakhutdinova, Alexander Davydov, Yuriy Shaparenko

Saint Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University)
Saint-Petersburg, Russian Federation
nurzhan_nfs@hotmail.com, {lebedev, ikr}@cit.ifmo.ru, mikhailsukhoparov@yandex.ru,
kainagr@mail.ru, alex.davydov@mail.ru, y.shaparenko@inbox.ru

Abstract **— Considered the remote device is a wireless network as a queuing system. The analysis of a wireless network bridge is under attack, aimed at organizing "broadcast storm", in order to determine the availability of stand-alone units, the ability to carry out their tasks in the functional impact of the information. It identified a number of conditions for the organization of this type of attack by a potential malefactor of information security. Showing model depending functioning of the system in the normal state and the implementation of an information system on the impact from a potential malefactor. The analytical modeling of wireless network operation in normal mode, and during the attack on the organization of "broadcast storm." Disclosed is an experiment that provides obtaining statistical information on the remote wireless devices. The results of the experiment of a typical attack system, to transmit data via a broadcast packet network scan at different values of the noise intensity posts from the violator information security. The proposed model can be used to determine the specifications of the wireless ad hoc network of devices, making recommendations on the node configuration aimed at countering "broadcast storm."**

Index Terms—**Information security, wireless networks, multi-agent systems, vulnerability, accessibility of devices, information security model.**

I. Introduction

A permanent decline in the cost of wireless devices and improvement of quality indicators of time, energy and information characteristics of sensors makes these technologies very promising. The flexible architecture, increase the computing power of individual units can not only shift from standard wired devices, but also build self-sufficient with respect to multi-agent systems, carrying out the reception, processing, analysis of the received and transmitted data.

The implementation of self-organizing wireless networks is accompanied by the need to address other problematic issues of information security [1,2].

Among the main vulnerabilities can provide an opportunity to listen to the channel, the premise of "external" packages, the implementation of physical access to the malefactor's site, the lack of standardization of intelligent routing algorithms that take into account the state of the network. A large number of devices that provide an intelligent transmission, collection, processing data packets, their relative remoteness, autonomy of operation, dynamically changing topology, weak study models, methods and algorithms for concurrent detection of incorrect information from the compromised node determines the difficulty of creating the classic protection systems [3-5].

Some of the main potential vulnerabilities associated with the peculiarities of the functioning of the individual units. The need for the exchange of official information in the event of a number of internal and external events is sending broadcast packets between network nodes.

II. Research Statement

A typical network node includes a transceiver, a battery, a processing module which may be connected various detectors. In view of such a structure is necessary to effectively address the problems related to the conservation of energy, providing processing power and throughput characteristics of channels [6]. The combined solution of these problems leads to the presence of most of the protocols of a number of problematic issues implosion, and the blind overlay resources, making such technologies vulnerable to a series of attacks by hackers [7,8].

Based on the features and operation of wireless networks using the recommended settings to optimize operation of the remote node wireless sensor network may exercise "a broadcast storm" [9-11].

The technology of this attack is associated with vulnerabilities that lead to a large increase in the service pack on the network. In the simplest case, if you allow the rules defined by the system administrator, then wheeled traffic growth can be generated broadcast messages. Analysis [6,8,10] brings out a number of conditions in the configuration settings for this type of attack:

- the absence of restrictions on the time to live;
- the presence of rules that allow to transmit a frame to the broadcast address to all except the node from which he resigned;
- the introduction of devices, continuously generating the message.

Especially it should be noted that to carry out destructive influences the potential infringer may have minimal capacity

to dispatch bad frames. The result is a waste of resources to the reception, transmission, processing overhead, which is under load a wireless sensor network not only performs its functions, but also becomes unmanageable [9]. There is no possibility of rapid access and autonomous control devices that are non-stop responding to events in the network. There is a threat to the implementation of the availability of a wireless sensor network devices due to deliberate action on the part of the malefactor, to increase the number of broadcast and other service messages, resulting in blocked access to communication channels and nodes of the computer system. The vast majority of models describing the place in a wireless network processes that information does not include the possibility of exposure to a potential malefactor.

Thus, in order to ensure information security (IS) of the wireless sensor network, there is a problem of probabilistic assessment of the availability of the devices under attack such as "broadcast storm".

III. MODELING OF THE IMPACT ON THE SYSTEM

Carrying out an attack on the part of the malefactor is reduced to increase the intensity of the receipt of applications, leading to the inability to service the total message flow device. Such a condition can occur in case of a configurable device when filling a buffer having a predetermined volume or unavailable channels, resulting in a loss of the application. There is a threat to the availability associated with limitations on authorized access to network elements, stored information, information flows, services and applications due to events affecting the network [10].

To simplify the model we consider relatively simple, do not have high processing power for the wireless sensor network, for example, a Zigbee technology, with limited functionality, receiving and transmitting a small limited range of types of messages without a priority with pre-configured parameters. A service duration depends on the number of events in a predetermined time interval. Assuming that the device claims arrival process is a poisson process, and the duration of service is distributed exponentially, it becomes possible to consider the processes of collection, processing and transmission of information, as the queuing system M / M / 1 / n.

Features a hardware implementation of autonomous wireless remote nodes suggest a buffer of input messages, which allows you to store several messages received for processing.

Assuming that the process of receipt of applications is a Poisson process, and the duration of service is distributed exponentially, it becomes possible to consider the processes of collecting, processing and transmitting information, such as queuing system M / M / 1 / n.

Applying the theory of queuing systems can be assumed that the probability of packet loss during transmission from A to B via one device C will be determined by the formula (1):

$$P_c = 1 - \rho^m \frac{1-\rho}{1-\rho^{m+1}},\ \rho = \frac{\lambda}{\mu} \qquad (1)$$

where in λ - the input flow, μ is the intensity of the service, m - the size of the input buffer processing device.

In the process of transmitting information in a wireless network information flow passes through several such devices. To estimate the probability of packet loss, passing by k units expression (1) takes the form:

$$P_c = (1 - \rho^m \frac{1-\rho}{1-\rho^{m+1}})^k \qquad (2)$$

Most probabilistic models implemented evaluation systems does not imply the existence of a potential malefactor, whose actions are aimed at exploiting vulnerabilities used protocols and system components.

However, a certain openness and accessibility of the network allow an malefactor to perform actions to increase the intensity distribution of frames that do not contain the correct information, a bad checksum, the wrong title, which lead to an unjustified waste of resources on the part of the system.

Fig.1 shows the effect on the sequence of the chain of devices that transmit data packets to the intensity λ_p. An malefactor on the network increases the number of events causing the generation of broadcast frames, with intensity λ_{sh}.

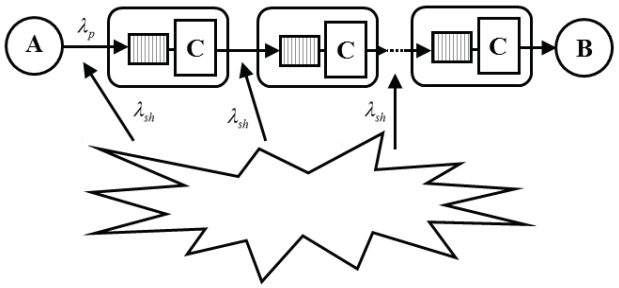


Fig. 1. Impact on the chain of devices

Received broadcast frame is processed by a wireless network and is relayed to the next node. In the presence of a powerful transmitter, a plurality of nodes at the same time affecting the wireless network, there is a growing junk traffic network.

Using expressions (1) and (2) the probability of packet loss in the network, subject to a broadcast storm, is given by:

$$P_u = \prod_{k=1}^{r} \left(1 - \left(\frac{\lambda_p + k\lambda_{sh}}{\mu} \right)^m \frac{1 - \left(\frac{\lambda_p + k\lambda_{sh}}{\mu} \right)}{1 - \left(\frac{\lambda_p + k\lambda_{sh}}{\mu} \right)^{m+1}} \right),$$

(3)

where in λ_p - is the intensity of the useful traffic, λ_{sh} - the intensity of the noise of traffic, μ - is the intensity of service, *n* - the size of the input buffer processing device, *k* - is the number of devices found in the path of the package.

On Fig.2, Fig.3, Fig.4 show the dependence of the probability of loss from the package, λ_p is the intensity of the useful traffic, λ_{sh} is the intensity of the noise of traffic, μ is the intensity of service.

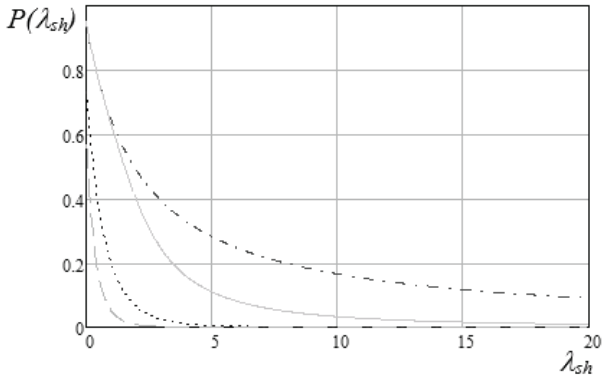


Fig. 2. Dependence of the probability of loss of information when changing the package, λ_{sh} - is the intensity of the noise of traffic

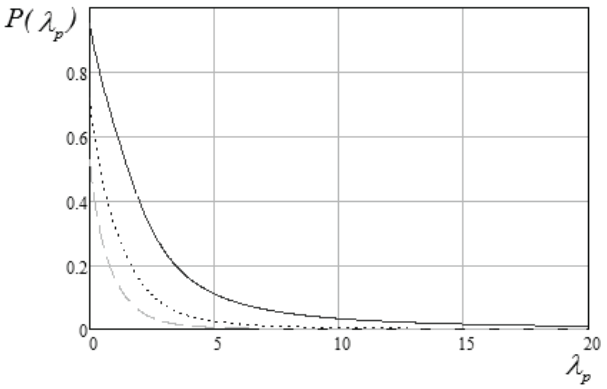


Fig. 3. Dependence of the probability of loss of information when changing the package, λ_p - is the intensity of the useful traffic

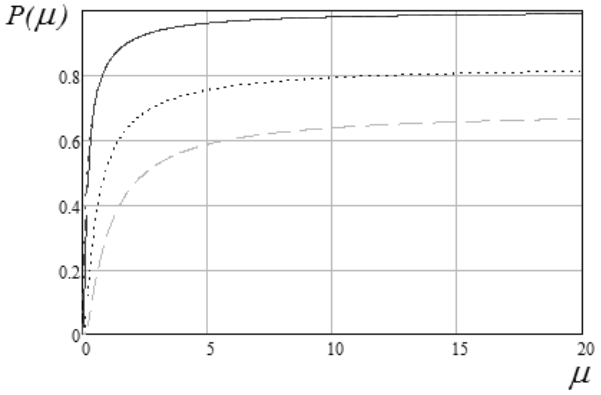


Fig. 4. Dependence of the probability of loss of the information packet at change, μ - is the intensity of service

Despite the possibility of limiting the admitted assumptions characteristic of the mathematical apparatus of Queueing theory, such models allow us to estimate the probability state of the system and levels of indicators specific to its operation in harsh conditions and environments, taking into account the violator. With regard to consideration the type of attack, the model takes into account the possibility of the person carrying out the impact on the intensity of the events causing the generation of broadcast packets.

IV. Experiment

To implement the experiment was configured wireless network based devices Telegesis, presented in Fig.10, consisting of several units. From device A to device B transmitted messages at a rate of 250 kbit/sec. A node C contained sniffer generating broadcasts.

The purpose of the experiment was to obtain quantitative availability terminal. As used metric percent of lost packets. Each node can accept messages only from two sites, one of which was the sniffer sends broadcast packets, and the other - the node providing data traffic by relaying all received packets. On the terminal node received packets are analyzed and determined by the statistics of lost and unrecognized messages.

The percentage of lost and unrecognized information packets in the configured communication channel for the speed of 250 kbit/sec at a frequency of generation of messages, query nodes in the network, in the second of the sniffer is shown in Fig.6 and Fig.7.

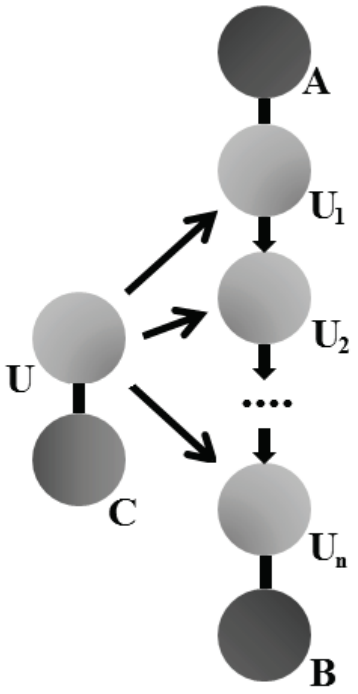


Fig. 5. The scheme of the system for the experiment

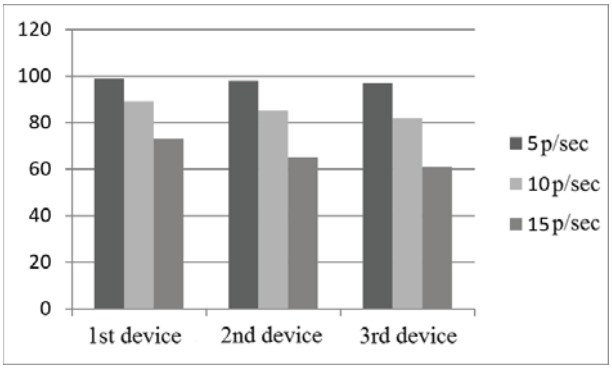


Fig. 6. A packets loss p (%), depending on the number of devices, the transmission frequency broadcast type AT + N

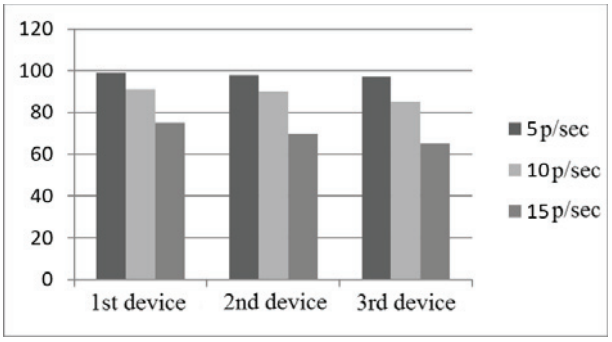


Fig. 7. A packets loss p (%) depending on the number of devices, the transmission frequency broadcast type AT + SN: 00

V. Analysis of the Results of the Experiment

Comparing charts theoretical probability of packet P_{loss} (Fig. 8) with histograms in Fig.6 and Fig.7, you can make a qualitative idea of the proximity of theoretical and experimental distributions of packet loss probability for different values λ_{sh}.

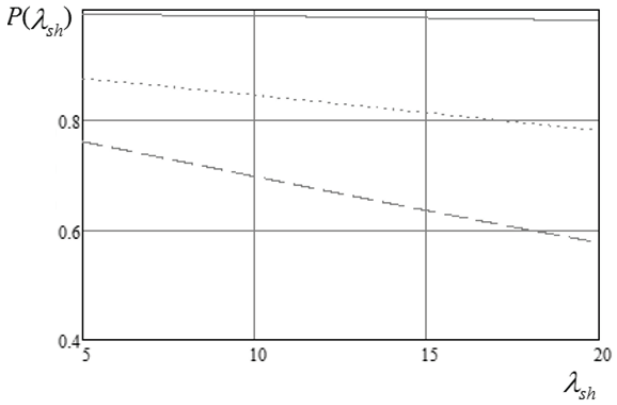


Fig. 8. Model according to the experimental values

Experimental values used in the model and natural experiment are shown in Table I

TABLE I. Experimental Values

Parameter	Value for the experiment
λ_p	120 messages per second
λ_{sh}	1 to 16
μ	180 messages per second
n	10 messages
k	1, 2, 3

Fig.9 shows histograms of the experimental values of packet loss for different values λ_{sh} for k= 3 devices.

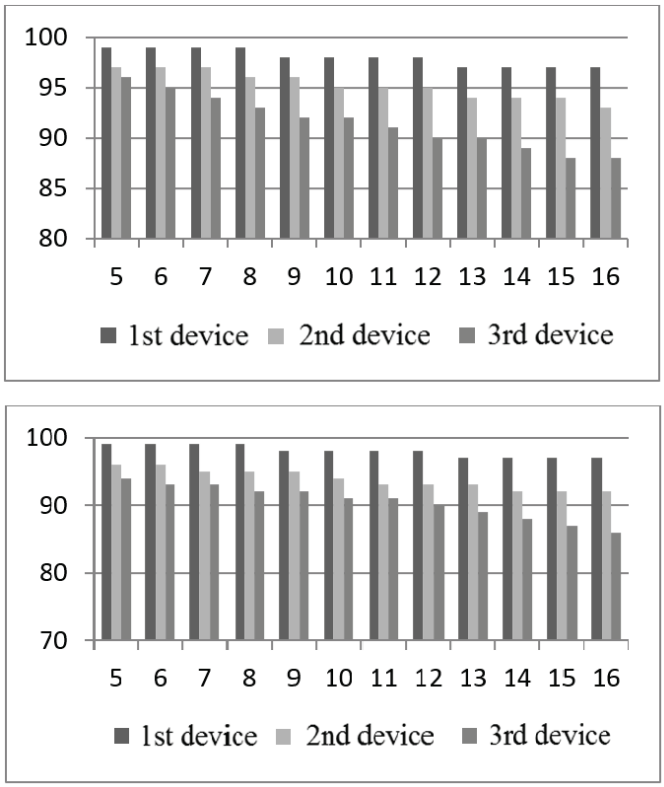


Fig. 9. The experimental values of losses parquet ρ (%) of broadcast messages for k = 3

Testing statistical hypothesis that received the experimental distribution is different from the theoretical, carried out at level of significance $\alpha = 0{,}05$ for Pearson (χ^2 - criterion):

$$\chi_v^2 = \sum_{j=1}^{l} \frac{(n_j^* - np_j)^2}{np_j} \quad (4)$$

when the number of degrees of freedom $\nu = 11$, n = 100 sample volume, number interval l=12.

It is known [12] that the number of events in np_j formula (4) in the intervals λ_{sh} expected parameter values may be equal to 2 or 3, if $l \geq 10$. Therefore, in accordance with the formula (3), for example, k = 3 $\lambda_{sh} = 5$. The critical value for the corresponding table [12] found a criterion and still $\chi^2_{cr} = 19.68$. The experimental value χ^2_{ex} criterion is calculated using the formula (4), equal to 0.4795 and 1.051 for broadcast messages AT + IN and AT + N: 00, respectively, while the inequality $\chi^2_{ex} < \chi^2_{cr}$.

It follows that at the level of significance $\alpha = 0{,}05$ can be argued that the discrepancies between the theoretical and experimental P_{loss} probability distributions of packet loss in the network, a broadcast storm-prone, at various specified values of intensity noise of traffic λ_{sh} not statistically significant.

Thus, the assumption of the Poisson flow of the process of receiving applications and exponential distribution service time confirmed by the data of the statistical analysis of wireless networks load.

VI. Conclusion

The widespread emergence of wireless networks, the ability to detect them outside the controlled area, making them an attractive target for attempts at various kinds of attacks. A potential malefactor with the scanner of radio, protocol scanner software to decode the dongle has sufficient capacity for the organization of eavesdropping, radio coverage and creating a false access point wireless sensor network.

The implementation of a large number of projects on the basis of technology Bluetooth, ZigBee, WiFi, their use in intelligent transport systems, local area networks, sensor networks makes it necessary to provide the required level of security circulating in their data [13-17].

The proposed model makes it possible to explore the availability of devices, wireless network vulnerable to attack, "broadcast storm" based on the performance of the intensities of the transmitted and received data messages, allowing you to select various settings and thresholds lengths, the buffer size during system configuration, the number of packets, limiting "the limb of one or other system resources."

The proposed model takes into account the characteristics of violator behavior and does not require a significant investment of computational resources, setting up special computational experiments in identifying common patterns in the behavior of the system.

References

[1] P. Kumar, M. Ylianttila, A. Gurtov, S.-G. Lee, H.-J. Lee, An Efficient and Adaptive Mutual Authentication Framework for Heterogeneous Wireless Sensor Networks-based Applications, MDPI Sensors, 14(2), 2732-2755, 2014.

[2] Sridhar, P., Sheikh-Bahaei, S., Xia, S., Jamshidi, Mo. Multi agent simulation using discrete event and soft-computing methodologies // Proceedings of the IEEE International Conference on Systems, Man and Cybernetics 2, 2003, pp. 1711-1716

[3] Page J., Zaslavsky A., Indrawan M. Countering security vulnerabilities using a shared security buddy model schema in mobile agent communities // Proc. of the First International Workshop on Safety and Security in Multi-Agent Systems (SASEMAS 2004), 2004. – P. 85–101.

[4] I. Zikratov., E. Kozlova., T. Zikratova. *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [A vulnerability analysis of robotic systems with swarm intelligence] // *Nauchno-tekhnicheskiy*

vestnik informatsionnykh tekhnologiy, mekhaniki i optiki. – 2013. – № 5 (87). – P. 149–154.

[5] Zikratov, I. Lebedev, A. Gurtov, Trust and Reputatio Mechanisms for Multi-agent Robotic Systems, Lecture Notes in Computer Science Volume 8638, 2014, pp 106-120. (Proc. of the Internet of Things, Smart Spaces, and Next Generation Networks and Systems)

[6] Wyglinski, A.M., Huang, X., Padir, T., Lai, L., Eisenbarth, T.R., Venkatasubramanian, K. Security of autonomous systems employing embedded computing and sensors//IEEE Micro 33 (1) 2013, art. no. 6504448, pp. 80-86

[7] Lebedev I.S., Korzhuk V.M. The Monitoring of Information Security of Remote Devices of Wireless Networks // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2015, Vol. 9247, pp. 3-10

[8] M. Prabhakar, J. N. Singh, G. Mahadevan Nash equilibrium and Marcov chains to enhance game theoretic approach for vanet security. // International Conference on Advances in Computing, ICAdC 2012; Bangalore, Karnataka; India; 4 July 2012 through 6 July 2012, Volume 174 AISC, 2013, pp 191-199

[9] Korzun D.G., Nikolaevskiy I., Gurtov A.V. Service Intelligence Support for Medical Sensor Networks in Personalized Mobile Health Systems // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2015, Vol. 9247, pp. 116-127

[10] Recommendation ITU-T X.805. *Security architecture for systems providing end-to-end communications.*

[11] Nikolaevskiy I., Lukyanenko A., Polishchuk T., Polishchuk V.M., Gurtov A.V. isBF: Scalable In-Packet Bloom Filter Based Multicast // Computer Communications - 2015, Vol. 70, pp. 79-85

[12] E.I. Kulikov. Prikladnoy statisticheskiy analiz [Statistical analysis]. Study guide for universities. M.: Goryachaya liniya-Telecom, 2008. – pp. 464.

[13] S. Komov. Terms and definitions in the information security field. – M., AS-Trasт, 2009, pp. 304.

[14] Kumar P., Gurtov A.V., Linatti J., Ylianttila M., Sain M. Lightweight and Secure Session-Key Establishment Scheme in Smart Home Environments // IEEE Sensors Journal - 2015, Vol. PP, No. 99, pp. 1

[15] Al-Naggar Y.,Koucheryavy A. Fuzzy Logic and Voronoi Diagram Using for Cluster Head Selection in Ubiquitous Sensor Networks // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 14th International Conference, NEW2AN 2014 and 7th Conference, ruSMART 2014 Saint-Petersburg, Russia, August 27–29, 2014, Proceedings. Springer, LNCS 8638, – PP. 319–330.

[16] Chehri A.,Hussein T. Moutah Survivable and Scalable Wireless Solution for E-health and Emergency Applications. // In EICS4MED 2011. Proceedings of the 1st International Workshop on Engineering Interactive Computing Systems for Medicine and Health Care. Pisa, Italy. – 2011. – PP. 25–29

[17] Koucheryavy Y.A., Ometov A.Y., Andreev D.S. On the role of wireless technologies in the development of the Internet of Things. // Information Technology and Telecommunications. – 2014. – N 3 (7). – PP. 31–40. (In Russian)

Monitoring of Qualitative Changes of Network Traffic States Based on the Heteroscedasticity Effect

Aleksandr Skatkov, Aleksey Bryukhovetskiy, Victoria Shevchenko, Dmitry Voronin
Sevastopol State University,
Sevastopol, Russian Federation
{AVSkatkov, Bryukhovetskiy, VIShevchenko, Voronin}@sevsu.ru

***Abstract*—The proposed approach is based on the use of the heteroscedasticity effect, which is widely known in econometrics. The authors believe that it can give significant positive results for the detection of changes of network traffic states while analyzing the dispersion of the random errors in the regression equations. This method will allow to solve the problem of the definition of supply and demand on the services maintenance in the cloud environments.**

***Index Terms*—Cloud computing environment, network traffic monitoring, heteroscedasticity, adaptation.**

I. Introduction

Rapidly expanding level of modern computerization leads to the need for a more widespread use of different software tools (including cloud-computing technologies). Cloud computing is an information technology to provide easy network access to a common pool of configurable computing resources. Servers, data storage systems, applications and services are the examples of such resources. There are special procedures that can quickly activate and provide these resources with minimal loss of management or minimal interaction with the service provider [1, 2].

The complexity and scale of the network infrastructure predetermine high-level requirements for management tools automation that are used to ensure reliability and high performance of the network. The solutions of such problems are based on systematic analysis and cloud processes interpretation. Such type of analysis requires special traffic monitoring systems existence.

Recently, a number of important theoretical and practical results was got in the scientific direction of traffic monitoring. For example, in [3] software system Unihub has been proposed. This is a SaaS cloud calculation system and it allows users to use a Web-browser for remote work with interactive graphic Linux-based applications that support dynamic requirements for the computing resources distribution. These solutions make it possible to perform available resources balancing and to increase the limit of overall system's performance.

The paper [4] examines the problem of person-hours amount estimation needed for the design of functional system requirements in service-oriented architecture. For this purposes a stochastic model for information system's services identification has been proposed. It allows to estimate the expected number of elements and links under the limitation of input data minimization.

In the work [5] A.N. Shiryaev represents probabilistic and statistical methods of decision theory. The problem of significant but rare information situations detection is extremely important for cloud systems.

The main objectives of modern systems monitoring are the following [6,7]: scalability and openness of software provision, a detailed system analysis for bottlenecks search, collection of information about the network elements, system event monitoring, etc. In [7] a scalable method for trust management estimation is presented. It is based on the implementation of the distributed online monitoring system using the model of the extended semi-state machine.

The data describing the user's behavior in the cloud is widely used for the following purposes: energy consumption optimization [8], thin clients working delays reducing [9], virtual machines launching optimization and applications implementation using virtual desktops on remote servers.

The results of the "state of the art" analysis are the following. The complexity and scale of information and telecommunication systems increase continuously. Thus, the requirements to the cloud services' quality also significantly increased. Thereby, the following tasks are to be solved:

- To provide adaptive control of network traffic states changing in cloud environment;
- To provide the ability of network traffic states monitoring;
- To develop decision-making support system for resources allocation in terms of stochastic dynamic processes of cloud functioning;
- To propose criteria that are suitable for evaluation of service quality monitoring. These criteria should provide optimization task solving under conditions of a priori information deficit.

The proposed new scientific approach develops the well-known solutions and allows to overcome a number of requirements that are specific for the implementation of network traffic state control methods. The known solutions do not provide the required quality of network traffic states control, particularly, with their help it is rather difficult to track the moments of significant changes in traffic properties (such as intensity, dispersion, unsteadiness, etc.). In order to resolve these conflicts the authors propose to introduce the concept of

heteroscedasticity, which is widely used in econometrics. This term is understood as the assumption that the dispersion of the random deviations are significantly heterogeneous, which leads to a violation of the principles of the correct use of multiple regression methods [10].

The proposed method of traffic state assessment is based on the heteroscedasticity effect and has the following advantages:

- Sensitivity to changes in the traffic state;
- Low computational complexity;
- Adaptability to external impacts.

For further material exposition it is necessary to introduce a number of important definitions. A-event is an abnormal phenomenon, which may be caused by changes in demand and supply of the services among the market participants. G-effect is the process of heterogeneity evaluation of network traffic state, which can be used for the external influences detection that leads to the emergence of the A-event. G-detector is a device for the G-effect detection.

The proposed method of the traffic state assessment is relevant for the cloud systems having "customer-broker-provider" reference architecture [11].

II. Research Goals and Objectives

The goal of the research is to evaluate the qualitative changes in the traffic properties based on the G-effect. To achieve this goal, it is necessary to develop methods and models that can be used for the study of the efficiency of the network traffic states changes detection. Nonparametric methods of statistical processing, simulation techniques, methods of decision-making and interface managing technologies are the basic components of the proposed approach [12]. The essence of the effect, used for the network traffic state recognition, is achieved by the analysis of the random error's dispersion for the regression equations. The assumption about the constancy of the remainders' dispersion is known as the assumption of homoscedasticity [12]. If this assumption is violated (and the dispersion of residuals is not constant), then it is considered that the assessments are heteroskedastic, i.e. G-effect takes place.

The proposed method is aimed to be widely used in distributed environments aimed for scientific cooperation [11, 13]. In accordance with the principles of the system approach, a set of functional tasks is to be solved to meet the requirements of the designed system. They are the following:

- The system structure and functional modules development;
- The adaptive testing technology implementation;
- The efficient interactive mode provision (needed for decision-making person's activities automation);
- Simulation model of the network node development (needed for the compensation of a priori information deficiency).

The structure of the system and its implementation details are given in next Section IV.

III. Applicability Research of Nonparametric Statistical Criteria

The monitoring process of multidimensional stochastic network traffic has the following features: the complexity of full-scale monitoring systems creation and information deficit. Thus, the proposed simulation tool of hypothesis testing is useful for the statistical criteria effectiveness investigation. An effective plan of experiments can provide risks evaluation for different combinations of input parameters, describing the tested hypotheses.

The two types of errors can be done during the process of statistical hypothesis testing on heteroscedasticity existence. They are called errors of the first and the second kinds. The probability of making a mistake of the first kind is denoted as $P(H1|H0)$ and is called the level of significance α. The probability of making a mistake of the second kind is denoted as $P(H0|H1)$ and is called as β. The decision on the hypothesis validity is based on the knowledge about a random values sample that describes the output parameters of the system. The sample size is denoted as n. The presence of the disturbance in this random sample leads to the complexity of the task of confident and uncertain recognition zones identification.

IV. Scheme for Network Node State Monitoring

Distributed computer network consists of a set of nodes. A model of each node can be considered as queuing system of the following type: $G/G/K/L$. The traffic-monitoring scheme for a single node is presented on the fig.1.

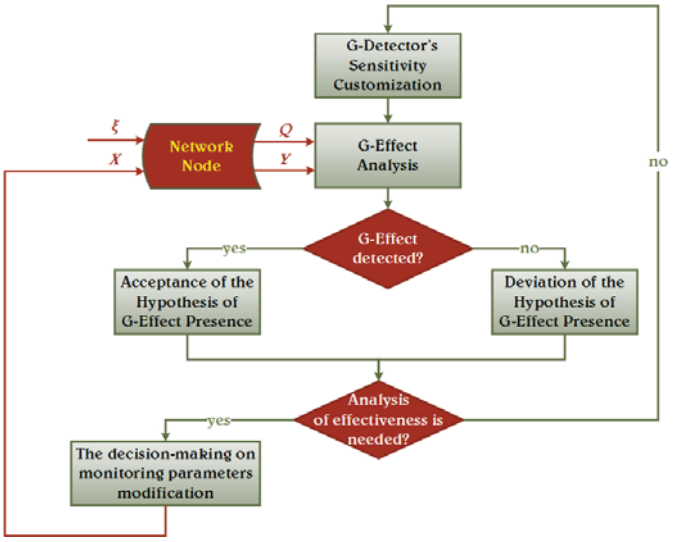


Fig. 1. The traffic monitoring scheme for a single node aimed for G-effect detection

The following designations are to be described:

- X – a vector of numerical characteristics of multidimensional input traffic flow;
- Y – a vector of numerical characteristics of multidimensional output traffic flow;

- Q – a vector of the loss flow characteristics;
- ξ – a vector of numerical characteristics of the random perturbations;
- λ_i – the intensity of the flow entering the i-th channel;
- μ_i – the intensity of the requests processing in the i-th channel;
- Ω – the learning sample;
- n – size of the sample;
- p – the level of significance;
- C_i – the recognition test;
- S_j – zones of certain/uncertain of G-effect recognition;
- ε_1 and ε_2 – normally distributed random variables with mathematical expectation equal to zero.

The functioning of the node is described as follows. Multi-dimensional flows, denoted as X and ξ, enter the node. The components of vector Y can be described by the equation of the regression (Eq. 1), and the components of vector Q can be described by the equation of the regression (Eq. 2).

$$y_i = f_i(x_1, \ldots, x_n) + \varepsilon_1 . \quad (1)$$

$$q_i = \varphi_i(x_1, \ldots, x_n) + \varepsilon_2 . \quad (2)$$

Generally, the dependences denoted as f_i and φ_i are non-lineal, but their linearization can be performed by using the approximate methods.

If it is necessary, the adjustment can be carried out with the help of the learning mode of the system. The aim is to adapt the setting of the following model parameters: the volume of training sample, confidence level, the first and second kind of error values limitations, periods of classification rules modification, etc. The G-effect detection is carried out in the process of traffic state changes evaluation using the values of residuals of the regression model. The specific form of the regression equation can be determined during the detailed problem formulation. It should be done separately for each individual case. In the simplest case, it is possible to use the equations of auto-regression. As a result, hypothesis of G-effect existence is accepted or rejected. Then the errors of the first and the second type are estimated and risk analysis of the accepted hypothesis is performed.

V. Monitoring System Settings Scheme

Monitoring system settings scheme is presented on the fig. 2. It consitsts of the following components:

- Preprocessing and validation modules;
- Module of configuring, testing and optimization of the adjustment parameters;
- Network traffic states classification module;
- Adaptation of decision rules' parameters to the current traffic state module;
- Decision-making support module for the formation of the output signal of the current traffic state, etc.

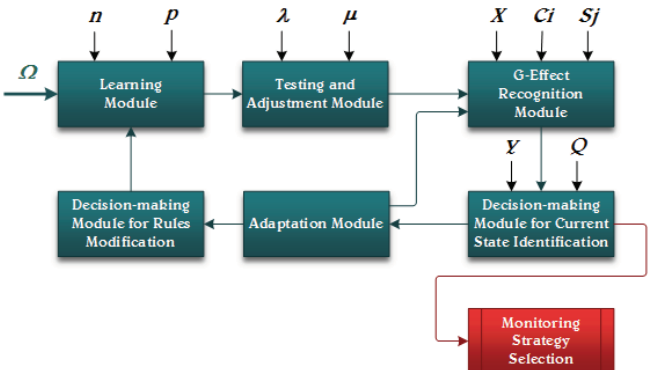


Fig. 2. Scheme of the single node traffic monitoring to determine the G-effect

Operation of the system supports multiple measurement mode for different time intervals. In order to adapt parameters of decision rules to the current traffic state, the system testing is carried out in the background mode. The controlled environment is unsteady, thus the informative value of decision rules decreases gradually.

Adaptive monitoring strategy selection is the main goal of the process of the monitoring system. Monitoring system's quality evaluation bases on the values of hypotheses $P(H0|H0)$, $P(H1|H1)$, on first and second kinds of errors: $P(H1|H0)$,$P(H0|H1)$. Important parameters for monitoring strategy selection (that influence the value of hypotheses) are the following:

f – the frequency of network traffic state monitoring,

$f \in [f_{min}, f_{max}]$,

Δt – the depth of monitoring (the duration of control),

$\Delta t \in [\Delta t_{min}, \Delta t_{max}]$,

$\underline{k_n}$ – the fullness of monitoring (the ratio of the number of controlled communication channels to the total number of channels), $kn \in [kn_{min}, kn_{max}]$.

p_{ij} – the confidence level using the criterion C_i for zone S_j,

n_{ij} – the volume of controlled sample using criterion C_i for zone S_j,

k_{ij} – resources utilization factor using criterion C_i for zone S_j,

z_{ij} – costs for network traffic state changes recognition using criterion C_i for zone S_j,.

As a result, evaluation matrix can be offered for decision-maker (table 1).

TABLE I. Evaluation Matrix for Decision-Maker

Criteria for recognition	G-effect recognition zones (f, Δt, kn)		
	S_1	S_j	S_q
C_1	…	$< p_{1j}, n_{1j}, k_{1j}, z_{1j} >$	…
C_i	…	$< p_{ij}, n_{ij}, k_{ij}, z_{ij} >$	…
C_m	…	$< p_{mj}, n_{mj}, k_{mj}, z_{mj} >$	…

VI. Scenario of the Experiment on the G-effect Detection

G-effect detection is a multistage process that includes the following steps: to configure the law of the input stream distribution (denoted as λ); output characteristics calculation (Y, Q); regression construction; collection of statistical data on the dynamics of changes in balances dispersions; decision-making on the G-effect existence (based on Spearman test). The scheme of computational experiments is presented on the fig. 3.

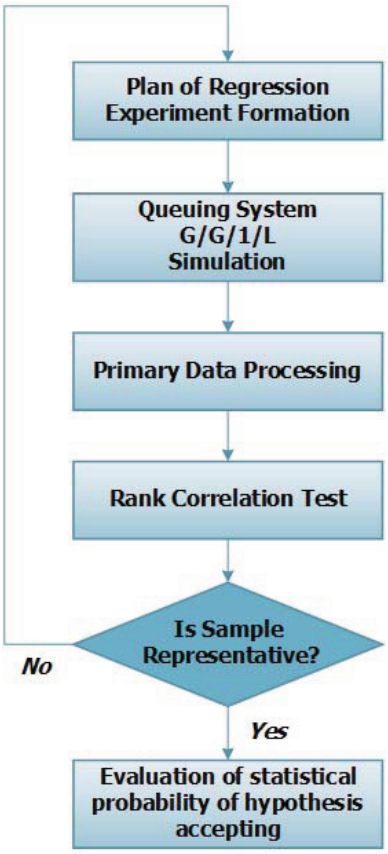


Fig. 3. The scheme of computational experiments

Technique of the experiment has the following steps:

- To construct regression dependences for the output streams of processed and lost requests;
- To determine the conditions of the G-effect occurrence. In other words, it is necessary to define sustainable G-detector threshold for the level of significance that is equal to 0.01. The introduced assumption stipulates that the distribution function of network traffic submit the normal distribution. Functional dependencies, denoted as $M(\lambda)$, $\sigma(\lambda)$, $M(\mu)$ are given. During the experiment, it is possible to vary the value, denoted as $\sigma(\mu)$.

The results of traffic experiments on the G-effect existence detection can be used as the input data in decision support systems. This will improve the effectiveness of decisions on the statistical criteria choosing. This raises the problem of the assured and uncertain recognition zones detection. To solve this problem the authors propose to evaluate the influence effect of the parameters (denoted as Δt, ρ, n) on the statistical sensitivity and statistical stability of the system.

VII. Conclusion

The proposed approach allows to adapt monitoring system to traffic state changes. The plans for further research contain the development of decision support systems for the G-effect detection in cloud environments.

Acknowledgment

This work was supported by the Russian Foundation for Basic Research (project № 15-29-07936).

References

[1] P. Mell and T. Grance. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology, 2010.

[2] Ceilometer developer documentation. Web: https://ceilometer.readthedocs.org/en/latest/measurements.html.

[3] D.A. Grushin and N.N. Kuzyurin, "Load balancing in Unihub SaaS system based on user behavior prediction" (in Russian), Proceedings of the Institute for System Programming, Vol. 26 (Issue 5), 2014, pp. 23–34.

[4] G. Tsiperman "The stochastic model of information system services identification process (in Russian), Proceedings of the Institute for System Programming, Vol. 26 (Issue 5), 2014, pp.7–28.

[5] A.N. Shiryaev. "Probabilistic - statistical methods in decision theory", 2014, P. 144.

[6] Homepage of Zabbix. The Enterprise-class Monitoring Solution for Everyone, Web: http://www.zabbix.com/ru/ .

[7] J. Lopez, S. Maag and G. Morales. "Scalable Evaluation of Distributed On-line Network Monitoring for Behavioral Feedback in Trust Management", Proceedings of the Institute for System Programming, Vol. 26 (Issue 5), 2014.

[8] T.D. Nguyen "Prediction-based energy policy for mobile virtual desktop infrastructure in a cloud environment", Inf. Sci. 2015, № 6 (319), pp. 132–151.

[9] M. Suznjevic, L. Skorin-Kapov and Humar I. "User behavior detection based on statistical traffic analysis for thin client services", Advances in intelligent systems and computing. Springer International Publishing, 2014, pp. 247–256.

[10] F. Liu, J. Tong, J. Mao and R. Bohn, "NIST Cloud Computing Reference Architecture". Recommendations of the National Institute of Standards and Technology, Sept. 2011.

[11] A. Skatkov and V. Shevchenko, "Expansion of reference model for the cloud computing environment in the concept of large-scale scientific researches" (in Russian), Proceedings of the Institute for System Programming, Vol. 27 (Issue 6), 2015, pp. 285–306, DOI: 10.15514/ISPRAS-2015-27(6)-18

[12] N. Sh. Kremer, "Econometrics", 2006, P. 311.

[13] T. Duy, Y. Sato and Y. Inogushi, "Performance evaluation of a green scheduling algorithm for energy savings in cloud computing", Parallel and Distributed Processing, Workshops and PhD forum, Atlanta, April 2010, pp.1-8.

A Network Platform for Creating Digital Entrepreneurship in Cloud Environment based on Big Data

Heresh Beyadar
Department of Management, Boukan Branch,
Islamic Azad University,
Boukan, Iran
heresh.beyadar@gmail.com

Maryam Askari, Ararat Askari
Department of Management, Boukan branch,
Islamic Azad University,
Boukan, Iran
mmaskari77@yahoo.com, askary.ararat@gmail.com

***Abstract*—Nowadays, due to the high capacity of data and contents produced by a variety of organizations, individuals, and equipment, we are faced with "Big Data" phenomenon, the management, control, and process of which is above the capability of software tools during an endurable and predictable time. Cloud computing provides giant computing resources based on the demand which gains importance through Big Data. The growth of cloud computing and cloud-data storage is the main causes of the emergence and development of Big Data. In this respect, it can be said that cloud environment provides a good ground for managing, storing, and processing of big data.**

Due to the constant development of the use of information and communicative technologies, today, digital entrepreneurship is developing by using these instruments. In this paper, a network butterfly platform for manufacturing digital entrepreneurship within a cloud environment and based on big data is presented. Using this platform within the cloud environment causes the world's famous brands obtain high value and create several thousand businesses over a short period of time. The main goal of this platform is to exploit technologies of gathering, processing, searching, and visualizing data for extraction, discovery, and access to the hidden patterns, unknown correlations, and other useful information from extremely voluminous data for producing digital entrepreneurship.

Keywords—Digital entrepreneurship, knowledge age, network butterfly ecosystem, application, reverse engineering

I. Introduction

Today, due to the high capacity of data and contents generated by a variety of organizations, individuals, and equipment, we are confronted with a phenomenon named "Data Explosion". The development of communicative technologies (e.g., internet technologies) and information technologies (e.g., different kinds of electronic services) has confronted the world with "data deluge" phenomenon. Big Data industry develops from these emerging phenomena. Big Data or voluminous data is a set of data requiring its own special management practices due to their large size and highly spacious capacity. Such data are so spacious and voluminous on which such different operations as gathering, storage, management, and processing cannot be carried out by ordinary data-management software's and tools [1,6].

Cloud is an abstract image of a vast network and a mass, the capacity of which is not known, and the amount of its processing

Resources is not definite. The time and space dimensions of each of its components is also unknown. We do not know where in the mass the software and hardware are located, but we know what it presents. Based on this, it can be said that cloud space offers a good ground for management, storage, and processing of big data. Entrepreneurship is a notion existed since the creation of humankind. A review of entrepreneurship literature indicates that the concept of entrepreneurship was first introduced by economists. The concept of digital entrepreneurship also belongs to that kind of entrepreneurship which defines a job related to the IT fields by using IT tools and covers and implements it on the web environment. In his paper, a network butterfly platform is presented for producing digital entrepreneurship in a cloud environment and based on big data [2,7].

By studying several applications through reverse engineering technique for this platform, we understand that such companies as Microsoft, Sony, Amazon, and so on held a systemic look toward developing digital entrepreneurship at big data age. By utilizing the presented platform, these brands have been able to create thousands of businesses within digital revolution.

II. Big Data and Digital Entrepreneurship

Today, multi-Terabyte data sets (each Tera byte =1000 Gigabytes) to multiple Petabytes (each Petabyte = 1000 Terabytes) are called Big Data. Data which is produced from various sources such as social technology and media, clicks,

photos, videos, transactions, and etc. Three indicators of capacity, speed, and variety are used to define Big Data.

- Data capacity: The minimum data capacity is Terabyte in Big Data.
- Data variety: There is a combination of voluminous data as structural, semi-structural, and non-structural in Big Data.
- Data speed: Production speed, expiration, and use of data in Big Data is much higher than traditional approach.

The amount of data which can be stored and processed, nowadays, and its results can be utilized was not that much imaginable in the past. Outcomes and effects of this technological change are not only limited to the processors' world but also affect all areas of human life including digital entrepreneurship [8].

III. Cloud environment and digital Entrepreneurship

Cloud computing is a computing model based on large computer networks such as internet which provides a new pattern for supply, consumption, delivery of IT services (including hardware, software, information, and other shared computing resources), and use of the internet. This means that access to the IT resources at demand time and based on the amount of user demand is delivered to the user in a flexible and scalable way through the internet. In other words, cloud computing provides vast computing resources based on the demand which gains importance via the expansion of Big Data. The growth of cloud computing and cloud data storages is the main cause of emergence and progress of Big Data. Based on this, cloud environment provides a good ground for such diverse operations as gathering, storage, management, and processing of Big Data. In this environment, one of the processing outputs is digital entrepreneurship [6,7].

IV. Styling Digital Entrepreneurship Challenge at The Age of Big Data

Digital entrepreneurship is applicable at different hardware, software, information, and communication levels. There are many opportunities available for entrepreneurship across each level, and, more interestingly, entrepreneurship at information level is much easier and practical than entrepreneurship at other areas. For example, in nuclear power station area, due to limited application, there are less ideas; while, within the information and communication area with noticeable wideness, many ideas will be created [8].

On the one hand, due to the high capacity of data and contents generated by a variety of organizations, individuals, and equipment, today, we are confronted with a phenomenon named "Data Explosion". On the other hand, the development of communicative technologies (e.g., internet technologies) and information technologies (e.g., different kinds of electronic services) has confronted the world with "Big Data" phenomenon [6].

According to this, one of the challenges of world's famous brands is generating entrepreneurship based on Big Data. Cloud environment, today, helps to form an appropriate ground for different operations as gathering, storage, management, and processing of Big Data so that the digital entrepreneurship can be generated by presenting the proposed platform [4,8].

V. A Systemic Look at The Digital Entrepreneurship in The Cloud Environment

The well-known companies across the worlds such as Google, Yahoo, IB, and etc. have a systemic view toward digital entrepreneurship in cloud environment; that is, they contain a platform for digital entrepreneurship which is truly well-established and long-standing. Ideas constitute the system's input and entrepreneurship forms its output. On the basis of this, digital entrepreneurship platform can be constituted of society and different levels, and it should be noted that these components are interwoven and continuous. In the future technologic world, companies, organizations, and states can get come closer, via this platform, to the digital entrepreneurship standards at Big Data age and step towards the global society for generating digital entrepreneurship. Those developing countries without a systemic look at entrepreneurship and not using a certain platform were confronted with various limitations. The following figure presents a systemic look based on the Big Data in a cloud environment since cloud is a ground (floor) for different operations on the Big Data [9].

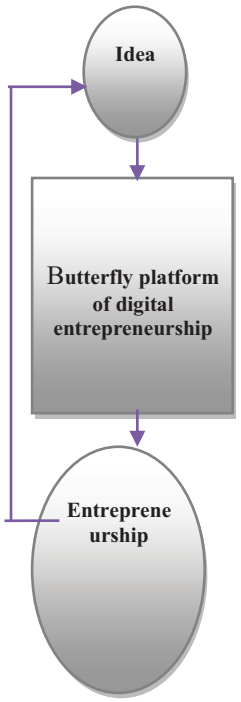


Fig. 1. Systemic look based on the Big Data on digital entrepreneurship in cloud environment

VI. The Proposed Platform

In this study, a network butterfly platform for developing entrepreneurship based on Big Data in the cloud environment is presented in the following figure.

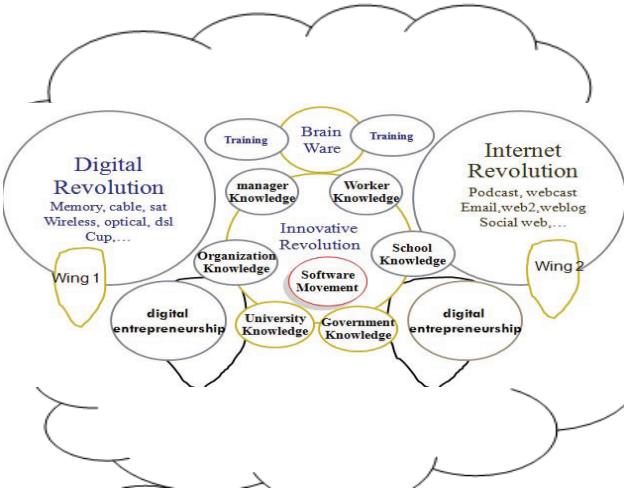


Fig. 2. Big Data in The Cloud Environment

The presented platform, in this study, has a butterfly-like structure which is composed of two wings: digital revolution and internet revolution. Digital revolution moves forward based on the Moore Law, i.e., greater power and lower cost. In other words, in digital revolution, wire and wireless hardware including processors, memories, cables, and satellites should be developed with regard to the Moore Law, but the internet revolution deals with the development of communicative techniques such as Podcast, Webcast, E-mail, Weblog, social networks, electronic magazines, digital movies, and so on. Innovative revolution includes community knowledge, government knowledge, university knowledge, organization knowledge, and school knowledge, which leads to the creation of Big Data. The headquarter of this platform is made up of brainware which is constantly being trained, but the final output of the presented platform in the cloud environment leads to the creation of digital entrepreneurship. In general, if we can, in the digital world based on this platform, develop information technology components such as software, hardware, and brainware in a parallel manner, we can achieve success in the area of generating digital entrepreneurship. Differently put, if digital revolution's wing is large and internet revolution's wing is small, the butterfly will not be able to fly. Furthermore, if brainware and/or innovative revolution is not developed, the butterfly cannot move the digital revolution and internet revolution wings.

Therefore, if a government, company, or organization considers this platform as a base for developing digital entrepreneurship, they will make progress in the area of digital entrepreneurship production.

VII. Big Data, Cloud Environment, and Proposed Platform

Based on the proposed platform, the innovative revolution includes community knowledge, government knowledge, university knowledge, organization knowledge, school knowledge, and so on, which ends in the generation of Big Data that needs their own special way of management practices due to their vastness and very high capacity. Such data are so voluminous and extensive that such operations as gathering, storage, management, and processing cannot be implemented on them via software and customary data management tools. Since processing and storage resources are not limited within the cloud environment, cloud environment is a good ground for Big Data technologies and proposed platform to create digital entrepreneurship [8].

All Big Data technologies including data gathering, data processing, data searching, and data analysis are applicable in the cloud environment. That is why the cloud environment is a good infrastructure for using the proposed platform for creating digital entrepreneurship. In other words, if the proposed platform is utilized in the cloud environment, its components, that is, software, hardware, and btrainware will be developed in a parallel manner, and we can achieve success in the area of digital entrepreneurship production.

In what follows, by applying reverse engineering, we propose several applications for the proposed platform based on the Big Data. By presenting these applications, we come to the conclusion that the global community is moving in the direction of this platform. In this regard, companies, governments, and famous brands should utilize the proposed platform due to the presence of Big Data in the cloud environment, and users can easily have access to this platform.

VIII. . Reverse Engineering and The Proposed Platform

Nowadays, the world's famous brands such as Sony, Yahoo, Google, Microsoft, Intel, Oracle, HP, AOL, and Dell, and so on working on the software, hardware, and brainware industry have been successful in the areas of digital entrepreneurship production. The reason for this success is using the proposed platform by such companies for developing digital entrepreneurship. This platform allows these companies to become a brand in the computer industry, receive great value within a short time, and create thousands of businesses. For instance, Sony Company which has turned to a brand in the computer industry, today, produces thousands of products, each product produces thousands of businesses, and each business creates several thousand creativities and innovations [9].

IX. . The Proposed Platform Based on The Big Data

Yesterday world was detached and hierarchical, but today world and tomorrow world is connected and integrated, which has led to the creation of Big Data and production of cloud environment. Accordingly, about 15 billion internet connections involving various types of mobile devices, cars, TVs, and embedded systems along with hundred HegzaByte information will be existed on the internet by 2015. Therefore, the present era is the era of Big Data and sharing massive information on the large networks.

It is possible that different companies network these platforms together. Accordingly, they can apply different topologies for creating digital entrepreneurship [4].

X. . The Proposed Platform and Reverse Engineering

Today, Nokia and Microsoft companies, each separately uses the proposed platform based on Big Data in cloud environment. According to the following figure, it is possible that the above companies turn the proposed platform to network in the cloud environment and make the bus topology the basis for developing digital entrepreneurship.

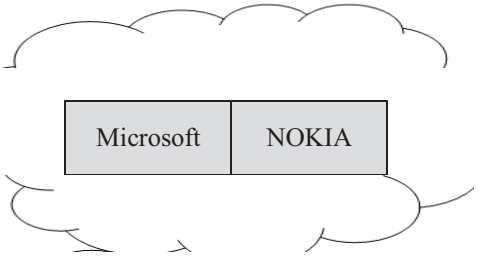


Fig. 3. The Cloud Environment (Nokia and Microsoft)

Within this application, we can also observe that Samsung and Microsoft companies use the bus topology of the proposed platform for developing digital entrepreneurship.

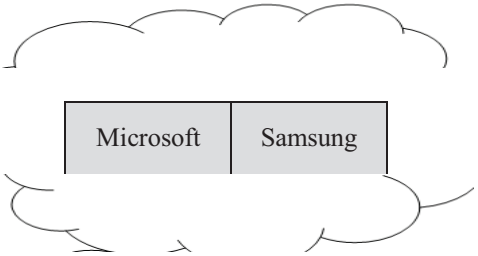


Fig. 4. The Cloud Environment (Samsung and Microsoft)

Application 3 indicates that today's proposed platform is also applied in such universities as MIT.

MIT University has networked the platform of several companies such as Sapynt, Mavstore, Microsoft, NetRicker, Akamay, and Hewlett Packard for implementing MIT OCW (free learning resources) project. By using the proposed platform in the cloud environment, these companies were able to create thousands of businesses across the world. In this project, the star topology of the proposed platform was used.

The needs are recognized through the entrepreneurship, and the progress would be made through meeting the needs. That is why the industrial countries have made substantial progress in the proposed platform components based on Big Data and in the cloud environment. Through the proposed platform, these countries both have dominated the world's economy and turn to information authority, and they direct the world culture toward their own desires. The proposed platform is an important factor in the economic growth and development and leads to the social benefit through the government, exploitation of unexplored opportunities, creation of new manufacturing lines, and the maximum use of capacities and capabilities [8,9].

XI. Conclusion

The development of communication technologies (internet technology) and information technologies (different kinds of electronic services) faced the world with "Big Data" phenomenon. Moreover, management, control, and processing of this pool of data is above the capability of software tools over an endurable and expected time. Cloud computing offers vast computing and storing resources based on the demand, which gains significance via Big Data. The growth of cloud computing and cloud data storages is the main cause of emergence and progress of Big Data. Accordingly, it can be said that cloud computing provides a suitable ground for management, storage, and processing of Big Data. In this paper, a network butterfly platform in the cloud environment makes the well-known world brands gain considerable value and create thousand businesses in a very short period of time. On the one hand, the main purpose of this platform was to exploit such technologies as gathering, processing, searching, and visualization of data for extraction, discovery, and access to the covert patterns, un-known correlations, and other useful information from highly massive data for creating digital entrepreneurship. On the other hand, the proposed platform is a constant platform leading to employment, improving the quality of life, reducing social anxiety, exploiting the resources, and activating productivity [6,8].

References

[1] Brynjolfsson E, Kahin B (2002). Understanding the digital economy. MIT Press, Cambridge.

[2] Timmons JA, Spinelli S (2004). New venture creation: entrepreneurship for 21th century. 6th edition, New York: Irwin/McGraw Hill.

[3] Kuratko DF, Hodgetts R (2004). Entrepreneurship: theory, process and practices. 6th edition, NY: Thomson/Southwest.

[4] Drucker P (1986). Innovation and entrepreneurship. HarperCollins, p. 277.

[5] Coulson-Thomas, C. The knowledge entrepreneur - How your business can create, manage and profit from intellectual capital Kogan Page Publishers, 2003, p. 240.

[6] Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M. A view of cloud computing. Communications of the ACM. 2010 Apr 1;53(4):50-8.

[7] Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Konwinski A, Lee G, Patterson DA, Rabkin A, Stoica I, Zaharia M. Above the clouds: A berkeley view of cloud computing.

[8] Zhang Q, Cheng L, Boutaba R. Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications. 2010 May 1;1(1):7-18.

[9] T OGRAPH B, MORGENS YR. Cloud computing. Communications of the ACM. 2008 Jul;51(7).

[10] Nurmi D, Wolski R, Grzegorczyk C, Obertelli G, Soman S, Youseff L, Zagorodnov D. The eucalyptus open-source cloud-computing system. InCluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on 2009 May 18 (pp. 124-131). IEEE.

Smart Control System of Thermal mode of Stabilization Column

Gulyaev S.V.
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences
Moscow, Russia
svgul@inbox.ru,

Abstract: **This paper considers the work of a smart adaptive control system of thermal mode of furnace firing heating and stabilization column, which separated hydrocarbon fractions of natural gas. The considered control system minimizes fuel gas consumption and increases the life of means of automation.**

Keywords: ***adaptive optimal control, gas processing, simulation, PID controller***

I. Introduction

In the chemical industry as a raw material commonly used natural gases, which, depending on the origin are divided into three groups: the pure gas, gas condensate and gas-oil. These gases are a mixture of hydrocarbons, the composition of which depends on the origin of the gas. A mixture mainly composed of methane and other heavier hydrocarbons - ethane, propane, butane, and others. When using natural gas for chemical synthesis purposes requires the separation of natural gas fractionation. One way such a separation process is distillation-based separation of components of the mixture differ in boiling point as a result of interaction of countercurrent mixture of vapor and liquid mixture [1]. This interaction takes place in the stabilization column. The stabilization column is loaded feedstock broad fraction of light hydrocarbons (RM) and condensate mixture of propane-butane (CMPB). The output stabilization column obtained hydrocarbon gas containing light fraction and the concentrate of pentane-hexane fraction (PHC). Part PHC preheated fuel gas in the furnace firing heating returns to the stabilization column. For normal operation of the stabilization column is necessary to strictly maintain therein a predetermined temperature range. In this paper, we propose a system of automatic control of thermal mode of the process of separation of the components of the gas mixture, which provides not only the stabilization of the temperature, but also the optimization of fuel gas consumption and increase the resource of technical means of automation.

II. Problem Statement

In known control systems of temperature mode stabilization column is commonly used control circuit controlling the supply of fuel gas for heating concentrate PHC entering the column. Consumption reflux (CMPB condensate) is controlled by a separate independent regulator [2]. In this paper we propose a system for regulating a thermal mode of, including a control loop CMPB feed. The proposed control system operates on the basis of information extracted from the temperature sensors located in the stabilization column and furnace firing heating, where the measured temperature of the furnace and preheated condensate PHC. Furthermore, PHC condensate preheating, the temperature stabilization process is carried out in a column also influence the regulator control valve which meters the flow of condensate CMPB, supplied to the stabilization column. Since the initial hydrocarbon feedstock RM (natural gas liquids) and condensate CMPB may have different chemical and technological characteristics (the percentage of hydrocarbon fractions, temperature, etc.) offered by the control system adapts to changes in these parameters. In addition to temperature stabilization objectives, the proposed control system minimizes fuel gas consumption and increases the life of the technical means of automation in the system.

Requirements for thermal control system:

1. The required level of temperature in the stabilization column should be provided fuel gas at minimum cost.

2. Operating temperature furnace firing heating regime should ensure its maximum speed, while carrying out the limitations of the temperature inside the furnace.

3. The required level of temperature in the stabilization column must be provided at the changing chemical and technological characteristics of hydrocarbons RM and condensate CMPB.

4. Mechanical actuators, which generally have the smallest resource in comparison to other resource elements of automation to run in a minimum number of switching operation in which the required control accuracy is ensured.

The first and second of these requirements can be satisfied by implementing a certain kind of structure of the control system, the third - the introduction of the system adaptation units, the fourth - in the use of this system, modified regulators that need to provide high quality control processes in piecewise constant control signals.

III. Proposed Structure Of The Optimal Control System

The block diagram of an optimal adaptive control system of thermal mode of stabilization column is shown in Fig. 1.

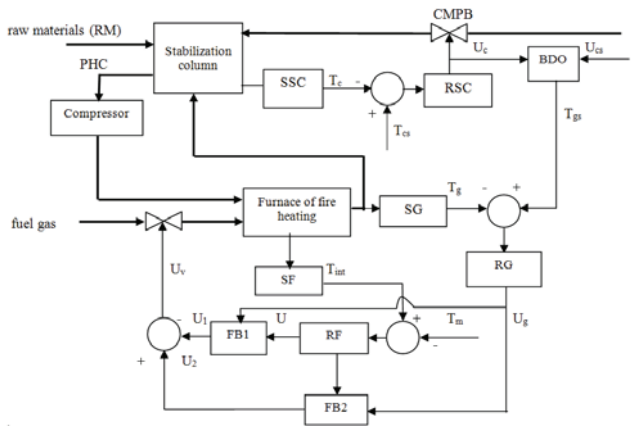


Fig. 1. Block diagram of smart control system of thermal mode of stabilization column and furnace firing heating

Designations:

SSC – temperature sensor of stabilization column.

RSC – temperature regulator of stabilization column.

BDO – block of dynamic optimization.

SG – temperature sensor of gas output stream.

SF – temperature sensor of furnace firing heating.

RG – temperature regulator of gas output stream.

RF– temperature regulator of furnace firing heating.

FB1 – functional control unit of regulator RG.

FB2 – functional control unit of regulator RF.

U_c – control signal of adjusting valve CMPB.

U_{cs} – set point of control signal of adjusting valve CMPB.

T_c – temperature of stabilization column.

T_{cs} – set point of temperature of stabilization column.

T_{gs} – set point of temperature of gas output stream.

T_g – temperature of gas output stream.

T_m – maximum temperature of furnace firing heating.

T_{int} – temperature of furnace firing heating.

U_g – control signal of adaptive regulator RG.

U – control signal of adaptive regulator RF.

U_1, U_2 – output signals of FB1 and FB2.

U_v – control signal of adjusting valve fuel gas.

IV. Operation Of Units And Control System Of Thermal Mode Of Stabilization Column As A Whole

The primary optimization objective to be pursued by the system of automatic control of the furnace firing heating and stabilization column is the problem of minimizing the consumption of fuel gas, subject to the required characteristics of the temperature regime of the process. Such a problem can be solved if a predetermined level of temperature in the column is maintained at a minimum flow of condensate therein CMPB. The amount of supply (consumption) CMPB condensate is determined by the value of U_c RSC controller output. Therefore, the smaller the signal U_c, in which the stabilization of T_c on T_{cs} level occurs, the less heat will need to flow PHC produced in the combustion of the furnace firing heating. The required amount of heat in a stream PHC to stabilize the temperature at the level of T_{cs} which is extracted from the furnace firing heating, determined by the level T_{gs} of stabilization of output gas flow temperature. If this level is to link a certain way with the minimum acceptable value U_{cs} flow CMPB control valve control signal, the task set is to minimize the fuel gas consumption will be solved. This connection is achieved dynamic optimization unit (BDO). This unit is a PI regulator, the entrance of which is the mismatch between the specified minimum value of the valve control signal supply CMPB U_{cs} and the current value of the signal. The output of the PI controller is the setpoint temperature stabilization T_{gs}. Block diagram PI controller with an input filter of the first order is shown in fig. 2.

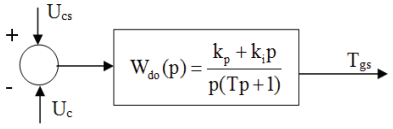


Fig. 2. Block diagram of dynamic optimization block

Thus, a cascade of stabilization column temperature control scheme in which the stabilization column temperature T_c is controlled by the cold condensate CMPB inner loop with RSC controller. Outer loop comprising regulators RG, RF BDO and controls the supply of fuel gas to maintain the desired temperature T_g of the output gas stream from furnace firing heating. Value U_{cs} is selected so that the control loop with RSC controller could quickly offset the decrease in the temperature stabilization of the column with respect to the value T_{cs}. The values of k_p, k_i and T - tuning parameters BDO These parameters are chosen so that the rate of change of setpoints T_{gs} was significantly lower tempo change temperature T_c and T_g. From the transfer function of fig. 2 that, for example, when $T_g > T_{gs}$, balance $T_c = T_{cs}$ will be achieved with $U_c > U_{cs}$. Therefore, from the logic block BDO setpoint stabilization temperature T_{gs} will decrease until it becomes the equality $U_c = U_{cs}$. Thus, the logic of block fig. 2 in conjunction with that shown in fig. 1 structure of the control system performs the task of minimizing fuel gas consumption. The obtained value T_{gs} is the setpoint for the temperature T_g stabilization circuit containing a temperature sensor SG and regulator RG. The second important task that should be considered to solve the automatic control system that is the problem of the protection of the furnace firing heating from overheating during transients when controlling the temperature T_{int}. During overheating T_{int} temperature is greater than the allowable level of T_m. Eliminate overheating can be the introduction of a special control loop containing a temperature sensor SF and the regulator RF.

Joint operation logic circuits SG, SG and SF, RF is next. In cases where T_{int} is much less than T_m, the control loop operates with SG, RG, and the output of

the regulator RF has a constant minimum possible value. In the control valve for regulating the fuel gas in this case, the signal passes through a RG functional blocks FB2, FB1 and the output signal is constant blocks. When the danger of overheating furnaces when T_{int} tends to T_m, RF controller output signal begins to grow, it is a warning signal the emergence of a dangerous situation. To prevent such danger should operate the control valve of fuel gas circuit not SG, RG, and on the information about the furnace temperature T_{int} by loop containing SF, RF. With this change, the furnace control structure control signal from regulator RG does not pass through the block FB2, and the output of this block assumes a constant value equal to the control signal of the controller when switching structure. Note that after changing the control structure output of the regulator RG for some time continues to grow. In the new control structure, the output control signal RF goes through the function block FB1 and, together with the constant signal of block FB2 controls regulates the fuel gas valve. In this case, there is a stabilization of the temperature of the furnace at the level of T_m and the gas flow temperature T_g with a lower rate of growth continues, seeking to the target temperature T_{gs}. Therefore, at a time controller output signal RG begins to decrease, which means that the approach to the transition temperature T_g to T_{gs}. At the time of reducing the output signal of regulator RG should return to the original management structure, as in this case the risk of overheating of the furnace will not. When you return to the structure of the function block FB1 stores the signal that controls the fuel gas regulates valve and to block FB2 and the regulator RG is set to zero initial condition. Operation function blocks FB1 and FB2 can be represented as:

$$U_1 = \begin{cases} U_v(t_i) \quad npu\ U_1(t) = 0 & t_i \le t < t_{i+1} \\ U_1(t) \quad npu\ U_g'(t) \ge 0, U_1(t) > 0 & t_{i+1} \le t < t_{i+2} \\ U_v(t_{i+2}) \quad npu\ U_g'(t_{i+2}) < 0, U_1(t_{i+2}) & t_{i+2} \le t \end{cases} \quad (1)$$

$$U_2 = \begin{cases} U_g(t)\ npu\ U_g(t_i) = 0, U_1(t) = 0 & t_i \le t < t_{i+1} \\ U_v(t_{i+1})\ npu\ U_g'(t) \ge 0, U_1(t) > 0 & t_{i+1} \le t < t_{i+2} \\ U_g(t)\ npu\ U_g'(t_{i+2}) < 0, U_1(t) = 0 & t_{i+2} \le t \end{cases} \quad (2)$$

where ti, ti+1, ti+2 - switching time of the control structures, Uv(ti), Uv (ti+1) and Uv(ti+2) – stored at these times the value of valve control signals, i=1;2; - natural numbers. Regulators RSC, RG and RF are linear PID controllers implementing the control law

$$u(t) = k_p \varepsilon(t) + k_i \int_0^t \varepsilon(\tau) d\tau + k_d \dot{\varepsilon}(t) \quad (3)$$

where ε(t) – error, kp, ki, kd – tuning parameters. To solve the problem of minimizing the number of actuators included in these regulators added a controlled quantizer fig. 3.

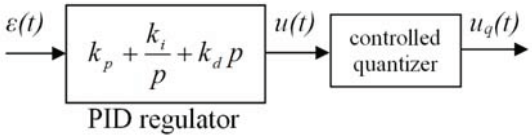


Fig. 3. Block diagram of regulators RSC, RG, RF

In accordance with [3] controlled quantizer (4) generates a time-quantized signal $u_q(t)$ with persistence equal segments $u_q(t)$ function, ie

$$u_q(t) = \begin{cases} k_p \varepsilon(t_i) + k_i \int_0^{t_i} \varepsilon(\tau) d\tau + k_d \dot{\varepsilon}(t_i)\ \ npu\ t_i \le t < t_{i+1} \\ k_p \varepsilon(t_{i+1}) + k_i \int_0^{t_{i+1}} \varepsilon(\tau) d\tau + k_d \dot{\varepsilon}(t_{i+1})\ \ npu\ t = t_{i+1} \end{cases} \quad (4)$$

The value of quantization $t_{i+1} - t_i = T_q > 0$ defines a time of constant control signal $u_q(t)$. Depending on the time T_q position of the working body of the actuator will change more or less often, increasing resource in any event this mechanism. With a relatively large values of T_q position of executive body does not change often, which brings in an additional delay of the control system and, as a consequence, to a deterioration of its dynamic properties. In the case of small values of T_q position of executive body changes often has a negative impact on the resource of the body. Therefore, there is always a trade-off value T_q, in which there are relatively high and the dynamic properties of control systems and a relatively large share of the actuator. From this it follows that the parameter T_q is a function of the dynamic characteristics of the process and in each case has its optimum value. To solve the problem of adaptation linear PID controllers RSC, RG and RF complemented block of estimation of parameter processes and block of PID settings calculation in accordance with the block diagram of fig. 4. The challenge of adaptation and automatic settings PID decided on the basis of the results obtained in the maximum degree of stability control systems, published in [4-6].

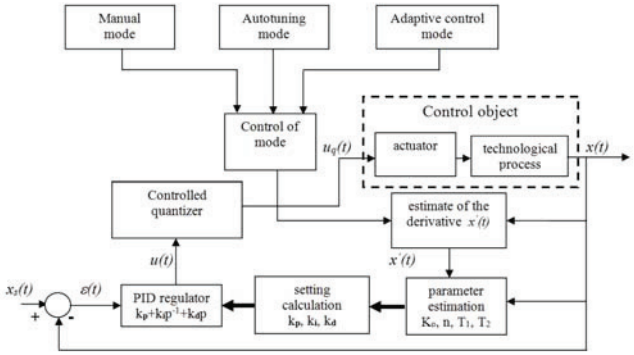


Fig. 4. Block diagram of adaptive regulators RSC, RG, RF

Designations:

$x_s(t)$ – set point.

$x(t)$ – controlled variable.

$\varepsilon(t)$ – control error.

$u(t)$ – control signal.

It is assumed that the behavior of the control object is approximated with sufficient accuracy, the transfer function of the form

$$W(p) = \frac{K_o}{(T_1 p + 1)^n (T_2 p + 1)} \qquad (5)$$

where K_o – gain, T_1 и T_2 – time constants, n – the number of inertial units, $T_2 >> T_1$.

Adaptive controllers operate in three different modes. In manual mode to manually set the control signal *u(t)*, in which there is a desired value of the controlled parameter *x(t)*.

The auto-tuning mode is run mode automatically determining tuning parameters of PID control by forming an additional step control action Δ_{st},

$$u(t) = \begin{cases} u_{beg} & t < 0 \\ u_{beg} + \Delta u_{st} & t \geq 0 \end{cases} \qquad (6)$$

which is not more than 5% of the range of variation of the control action u u(t). Such effects do not adversely affect the process. Parameters of the model (5) is estimated [4] keypoint functions x (t) and x'(t), where x (t) - the reaction of the object on a stepwise control action (6) in the blocks «estimate of the derivative x'(t)» and «parameter estimation Ko, n, T_1, T_2». In the block «setting calculation k_p, k_i and k_d » in accordance with [5] determined the optimum setting k_p, k_i and k_d of PID according to the degree of stability. In the adaptive control mode continuously without any control actions carried tuning PID parameters. Adjustment is carried out in a closed loop with a short control pulse impact, which is formed at a certain time and summed with the current PID control in accordance with [4]. The positive properties of the system laid down in setting control problem, provided the temperature regulators capabilities RSC, RG and the RF, members of the control loops considered a thermal mode of control system.

V. Simulation Of Control Thermal Mode Of Furnace Firing Heating And Stabilization Column

To investigate the performance of the proposed control system was carried out simulations of the furnace firing heating and stabilization column. Below are the results of the simulation of the proposed control system and the traditional linear, in which no commutation logic block structures and block optimization. Temperature of the stabilization column was maintained at level $T_{cs} = 100$. The limiting value of the furnace temperature taken $T_m = 300$, the optimal value of condensate CMPB regulator output signal is received $U_{cs} = 0.1$.

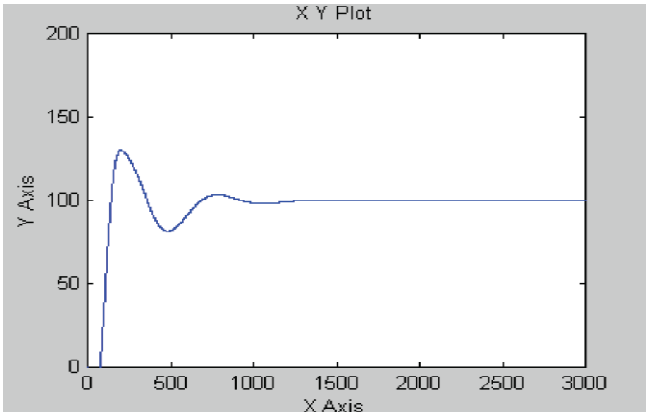


Fig. 5. Proposed control system. Temperature of stabilization column Tc

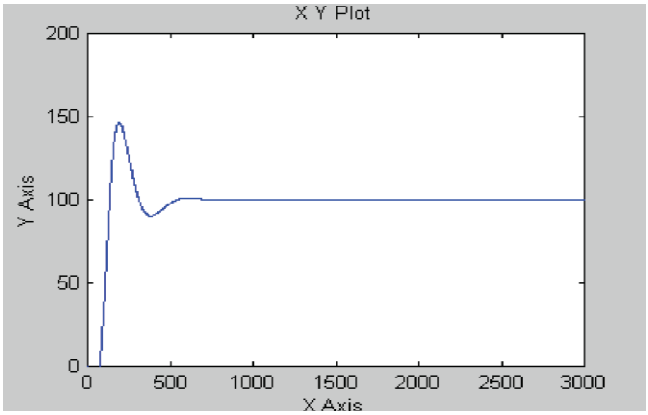


Fig. 6. Linear control system. Temperature of stabilization column Tc

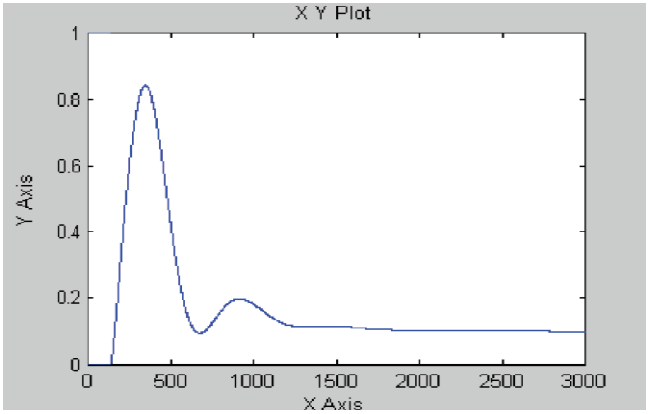


Fig. 7. Proposed control system. Output signal Uc

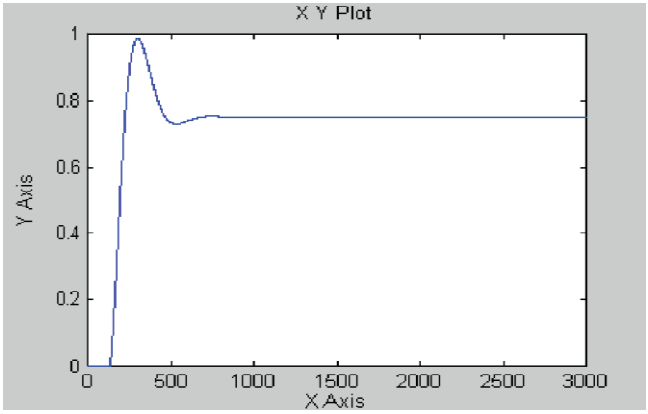


Fig. 8. Linear control system. Output signal Uc

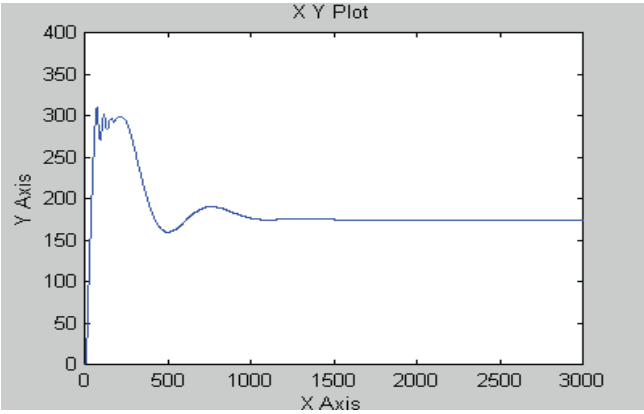


Fig. 9. Proposed control system. Temperature of furnace Tint

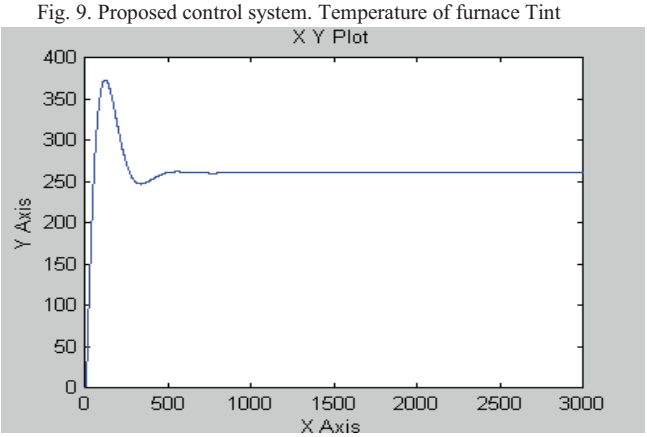


Fig. 10. Linear control system. Temperature of stabilization column Tc

From these graphs fig. 5 - fig. 10 should that the proposed control system is better than the linear system, since furnace temperature in the linear system is 75^0C greater than in proposed system, temperature ejecton in a linear system is greater than the proposed system and the control signal is greater than optimal by 65%. Made of approval of the simulation results confirm the findings that the proposed system of automatic control of thermal mode in the furnace firing heating and the stabilization column minimizes fuel gas consumption and provide optimal temperature conditions in the furnace firing heating. To check the efficiency of the controlled quantizer simulated operation control loop temperature of stabilization column at different values of the quantization time T_q.

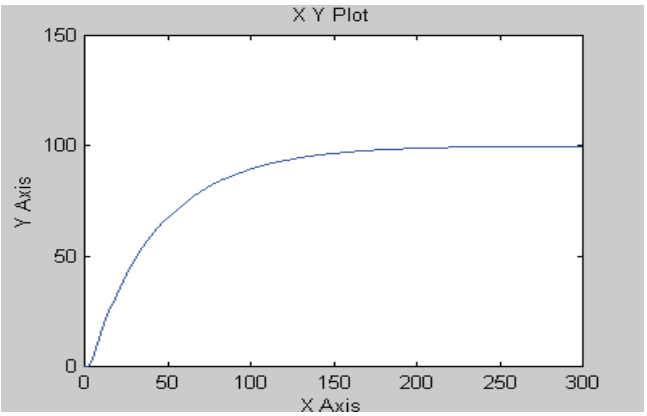


Fig. 11. Temperature control process Tc when sampling Tq = 1 sec.

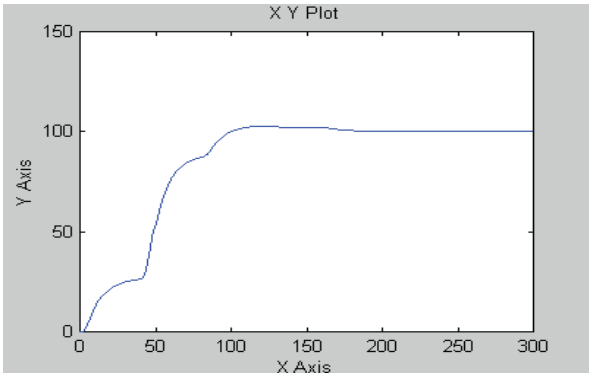


Fig. 12. Temperature control process Tc when sampling Tq = 40 sec.

From these graphs 11 - fig. 12 that the decrease in the number of such inclusions to 40 times without changing the control time, changes only the character of the transition, in most practical cases is acceptable. Thus, the assertion made above about the possibilities offered by the system of automatic control of thermal conditions in the furnace firing heating and stabilization column simulation results are confirmed.

VI. Conclusion

Proposed control system of thermal mode of furnace firing heating and stabilization column intended to separate gas hydrocarbon fractions.

The control system optimizes the stabilization column, reducing process gas consumption for heating the column. Controlled quantizers used in regulators control system to optimize the number of inclusions of technical automation devices, increasing thus, the system resource. A simulation of the control system, which confirms the high quality of the proposed control system.

References

[1] Skoblo A. I., Molokanov Yu. K., Vladimirov A. I., Shchelkunov V. A. - Processes and devices of oil and gas processing and petrochemistry Moskow, Gubkin University, 2012

[2] Bespalov A. V., N. I Kharitonov. Control systems of chemical and technological processes. Moskow, IKTS OF "AKADEMKNIG", 2007

[3] Gulyaev S.V., Pashchenko F.F., Malakhov V.A. Optimal adaptive control system of gas wells. Problemyi mashinostroeniya i avtomatizatsii. Engineering & Automation Problems. 2015. No 3. P. 90-98.

[4] A.M. Shubladze. Sufficient Conditions of Extrema in Systems with Maximum Degree of Stability I. Avtomatika i Telemekhanika. 1997. No 3. P. 93-105.

[5] Gulyaev S.V., Cherepova T.I., Shubladze A.A., Shubladze A.M. Dynamic Object Control Systems Optimal with Respect to the Degree of Stability. Control Sciences. 2003. No 3. P. 2-9.

[6] Shubladze A.M., Gulyaev S.V., Malakhov V.A., Olshvang V.R., Kuznetsov S.I. Automatic Setup and Adaptation in Industrial PID Controllers. Instruments and Systems: Monitoring, Control, and Diagnostics. 2007. No 12. P. 26-30.

Developing a Computation Algorithm for Approximation Solution of Linear Fredholm-Stieltjes Integral Equations of Second Kind Using Generalized Trapezoid Quadrature Method

Isa Muslu, Ph.D.
Applied Mathematics & Computer Department
International Ataturk-Alatoo University
Bishkek, Kyrgyzstan
isa.muslu@iaau.edu.kg

Moussa Mahamat Boukar, Ph.D.
The Department of Computer Engineering
Nigerian Turkish Nile University
Abuja, Nigeria
musa.muhammed@iaau.edu.kg

***Abstract*— Integral Equations (IE) are used in many areas especially in engineering to solve problems. Fredholm integral equations has a wide usage area among integral equations. There are a lot approximated solution methods for Integral equations and the quadrature methods is one of them. The main purpose of this paper is to develop a computation algorithm for approximation solution to the Linear Fredholm Stieltjes Integral Equations of Second Kind using the quadrature method, the Generalized Trapezoid Method (GTM). A few examples are presented at the end to indicate the efficiency and accuracy of the developed algorithm. The ready software package Maple is used for implementations.**

***Index Terms*—Linear Fredholm integral equation; generalized trapezoid rule; Stieltjes integral equations; linear integral equations; quadrature methods; maple.**

I. INTRODUCTION

Integral equations have been one of the fundamental tools in various areas of science such as mathematics, physics, biology and engineering etc. [1-4]. Fredholm integral equations are the most frequently investigated integral equations and there are numerous studies about exact and approximate solutions of them. Stieltjes Integral equations are a kind of integral equations which has a function as a variable and Fredholm type is one of the major class. However, in this paper, Linear Fredholm-Stieltjes integral equations of the second kind are considered. An approximation solution using GTM and computer implementation is consider in this work. In the first part, the general theory for approximation is given. In the second part, an algorithm is developed and computer implementation is done. [5]

II. PROBLEM STATEMENT

Linear Fredholm-Stieltjes integral equations of second kind can be defined as follows

$$u(x)=f(x)+\lambda\int_a^b k(x,s)u(s)dg(s) \qquad (1)$$

here $a\le x\le b$, $a\le t\le b$, and $k(x,t)$-kernel function of the integral, $u(x)$-unknown function to be find, $f(x)$-given function, $\alpha(t)$-strictly increasing function and $\lambda\in R$. If $\alpha(t)=t$, then it is Linear Fredholm Integral Equations of Second Kind [6-7].

In this work it is considered in case of $\alpha(t)\neq t$ .The Approximation solution method quadrature, Generalized Trapezium Method, to compute Stieltjes integrals is given in [7] and the method will be considered here to solve the equation.

III. APPROXIMATION SOLUTION ALGORITHM

Let's consider the equation (1) and define the function $g(s)$ as follows

$$g(s)=\varphi(s)-\psi(s),\ s\in[a,b] \qquad (2)$$

Here $\varphi(s)$ and $\psi(s)$ are known increasing functions. Substituting (2) in (1) yields,

$$u(x)=\lambda\int_a^b K(x,s)u(s)d\left(\varphi(s)-\psi(s)\right)+f(x),x\in[a,b] \qquad (3)$$

Opening parenthesis we receive,

$$u(x)=\lambda\int_a^b K(x,s)u(s)d\varphi(s)-\int_a^b K(x,s)u(s)d\psi(s)+f(x),x\in[a,b] \qquad (4)$$

If we apply generalized trapezoid method given in (14) to (3) we get the form

$$u(x)=\frac{\lambda}{2}\sum_{i=1}^{n}[K(x,x_i)u(x_i)+K(x,x_{i-1})u(x_{i-1})]\left[g(x_i)-g(x_{i-1})\right]+f(x)+E(T) \qquad (5)$$

when the function $g(s)=\varphi(s)-\psi(s)$ is written as a difference of two functions, equation (5) becomes as

$$\begin{cases} u(x)=\dfrac{\lambda}{2}\sum_{i=1}^{n}\left[K(x,x_i)u(x_i)+K(x,x_{i-1})u(x_{i-1})\right]\left[\varphi(x_i)-\varphi(x_{i-1})\right]- \\ \quad -\dfrac{\lambda}{2}\sum_{i=1}^{n}\left[K(x,x_i)u(x_i)+K(x,x_{i-1})u(x_{i-1})\right]\left[\psi(x_i)-\psi(x_{i-1})\right]+ \\ \quad +f(x)+E(T) \end{cases} \tag{6}$$

Here $E(T)$ indicates the error term stem from the generalized trapezoid approximation method. If the function $\bar{u}(x)$ is taken as approximation solution of the function $u(x)$ in (6), then the solution is found as

$$\bar{u}(x)=\frac{\lambda}{2}\sum_{i=1}^{n}\left[K(x,x_i)\bar{u}(x_i)+K(x,x_{i-1})\bar{u}(x_{i-1})\right]\left[\varphi(x_i)-\varphi(x_{i-1})\right]-$$
$$-\frac{\lambda}{2}\sum_{i=1}^{n}\left[K(x,x_i)\bar{u}(x_i)+K(x,x_{i-1})\bar{u}(x_{i-1})\right]\left[\psi(x_i)-\psi(x_{i-1})\right]+f(x)$$

or in opened and reorganized form

$$\begin{aligned}\bar{u}(x)=&\frac{\lambda}{2}K(x,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\frac{\lambda}{2}\sum_{i=1}^{n-1}K(x,x_i)\left[\left(\varphi(x_{i+1})-\varphi(x_{i-1})\right)-\left(\psi(x_{i+1})-\psi(x_{i-1})\right)\right]\bar{u}(x_i)+\\&+\frac{\lambda}{2}K(x,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x)\end{aligned} \tag{7}$$

is found. Since the equation is satisfied at every quadrature point and if $x=x_i$ is written in the equation, the value

$$\bar{u}(x_i)=\frac{\lambda}{2}\sum_{j=1}^{n}\left[K(x_i,x_j)\bar{u}(x_j)+K(x_i,x_{j-1})\bar{u}(x_{j-1})\right]\left[\left(\varphi(x_j)-\varphi(x_{j-1})\right)-\left(\psi(x_j)-\psi(x_{j-1})\right)\right]+f(x_i)$$

is found. If it is reorganized, then gets the value

$$\begin{aligned}\bar{u}(x_i)=&\frac{\lambda}{2}K(x_i,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\sum_{j=1}^{n-1}K(x_i,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\bar{u}(x_j)\right]\\&+\frac{\lambda}{2}K(x_i,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x_i)\end{aligned}$$

Replacing the natural numbers from 0 to n in the last equation gives,

$$\begin{aligned}\bar{u}(x_0)=&\frac{\lambda}{2}K(x_0,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)\\&+\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_0,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)+\\&+\frac{\lambda}{2}K(x_0,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x_0)\\ \bar{u}(x_1)=&\frac{\lambda}{2}K(x_1,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_1,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)+\\&+\frac{\lambda}{2}K(x_1,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x_1)\end{aligned} \tag{8}$$

$$\cdots\cdots\cdots\cdots$$

$$\begin{aligned}\bar{u}(x_{n-1})=&\frac{\lambda}{2}K(x_{n-1},x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_{n-1},x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)+\\&+\frac{\lambda}{2}K(x_{n-1},x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x_{n-1})\end{aligned}$$

$$\begin{aligned}\bar{u}(x_n)=&\frac{\lambda}{2}K(x_n,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_n,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)+\\&+\frac{\lambda}{2}K(x_n,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)+f(x_n)\end{aligned} \tag{3.7}$$

an algebraic equations system in (n+1) unknown. If $\bar{u}(x_i)$ is taken common parenthesis on the left hand side and $f(x_i)$'s send to the right hand side of the equation and arrange in (8), then the expression

$$\begin{aligned}&\left(1-\frac{\lambda}{2}K(x_0,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\right)\bar{u}(x_0)-\\&-\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_0,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)-\\&-\frac{\lambda}{2}K(x_0,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)=f(x_0)\end{aligned}$$

$$\begin{aligned}&-\frac{\lambda}{2}K(x_1,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)+\\&+\left(1-\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_1,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\right)\bar{u}(x_j)-\\&-\frac{\lambda}{2}K(x_1,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\bar{u}(x_n)=f(x_1)\end{aligned} \tag{3.8}$$

$$\vdots$$

$$\begin{aligned}&-\frac{\lambda}{2}K(x_n,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right]\bar{u}(x_0)-\\&-\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_n,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]\bar{u}(x_j)+\\&+\left(1-\frac{\lambda}{2}K(x_n,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]\right)\bar{u}(x_n)=f(x_n)\end{aligned} \tag{3.11}$$

is found and after some operations and abbreviations the (3.12)
expression

$$A_{00}=\frac{\lambda}{2}K(x_0,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right],$$

$$A_{0j}=\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_0,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]$$

$$A_{0n}=\frac{\lambda}{2}K(x_0,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]$$

$$A_{10}=\frac{\lambda}{2}K(x_1,x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right],$$

$$A_{1j}=\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_1,x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]$$

$$A_{1n}=\frac{\lambda}{2}K(x_1,x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right] \tag{3.13}$$

$$\vdots$$

$$A_{(n-1)0}=\frac{\lambda}{2}K(x_{(n-1)},x_0)\left[\left(\varphi(x_1)-\varphi(x_0)\right)-\left(\psi(x_1)-\psi(x_0)\right)\right],$$

$$A_{(n-1)j}=\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_{(n-1)},x_j)\left[\left(\varphi(x_{j+1})-\varphi(x_{j-1})\right)-\left(\psi(x_{j+1})-\psi(x_{j-1})\right)\right]$$

$$A_{(n-1)n}=\frac{\lambda}{2}K(x_{(n-1)},x_n)\left[\left(\varphi(x_n)-\varphi(x_{n-1})\right)-\left(\psi(x_n)-\psi(x_{n-1})\right)\right]$$

$$A_{n0}=\frac{\lambda}{2}K(x_n,x_0)\Big[\big(\varphi(x_1)-\varphi(x_0)\big)-\big(\psi(x_1)-\psi(x_0)\big)\Big],$$

$$A_{nj}=\frac{\lambda}{2}\sum_{j=1}^{n-1}K(x_n,x_j)\Big[\big(\varphi(x_{j+1})-\varphi(x_{j-1})\big)-\big(\psi(x_{j+1})-\psi(x_{j-1})\big)\Big]$$

$$A_{nn}=\frac{\lambda}{2}K(x_n,x_n)\Big[\big(\varphi(x_n)-\varphi(x_{n-1})\big)-\big(\psi(x_n)-\psi(x_{n-1})\big)\Big]$$

is received. This last system can be written as a matrix equations system as follows.

$$\underbrace{\begin{pmatrix} 1-A_{00} & A_{01} & \cdots & A_{0(n-1)} & A_{0n} \\ A_{10} & 1-A_{11} & \cdots & A_{1(n-1)} & A_{1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{(n-1)0} & A_{(n-1)1} & \cdots & 1-A_{(n-1)(n-1)} & A_{(n-1)n} \\ A_{n0} & A_{n1} & \cdots & A_{n(n-1)} & 1-A_{nn} \end{pmatrix}}_{A} * \underbrace{\begin{pmatrix} \bar{u}(x_0) \\ \bar{u}(x_1) \\ \vdots \\ \bar{u}(x_{n-1}) \\ \bar{u}(x_n) \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_{n-1}) \\ f(x_n) \end{pmatrix}}_{F} \quad (9)$$

Matrix equation system (9) can be written in short form as A.U=F (10). The approximation solution of the Linear Fredholm-Stieltjes integral equation (1) by finding out the unknown $\bar{u}(x_i)$'s in (9) and substituting in (7).

IV. COMPUTER ALGORITHM AND SAMPLES

The following figure 1 is the inputs for the developed solution algorithm in Maple for the equation.

```
K := (x, s) -> 1 + x^2*s:    f := x -> -x^2/4 - x/6 + x*sqrt(x):
φ := x -> sqrt(x): ψ := x -> x^(1/3): n := 5: a := 0.0: b := 1.0: λ := 0.1:
is( |λ| < 1 / (maximize(K(x,s), x = a..b, s = a..b) * |φ(b) - φ(a)|) )
true coef_of_h := 1
```

Fig. 1.Inputs for the developed solution algorithm in Maple

The following figure 2 is the solution algorithm developed in Maple software. This is the general algorithm to solve equations easily.

```
Solver_General := proc(K, φ, ψ, f, a, b, λ := 1, n := 100)
    local h := (b-a)/n, x :: list, A :: list, Id, Temp1, i, j, C, F, Pr, u :: list, U;
    for i from 0 to n do
        x_i := a + i*h;
    end do;
A_0 := λ/2 * K(x, x_0) * (φ(x_1) - φ(x_0) - ψ(x_1) + ψ(x_0));
A_n := λ/2 * K(x, x_n) * (φ(x_n) - φ(x_{n-1}) - ψ(x_n) + ψ(x_{n-1}));
for i from 1 to n - 1 do A_i := λ/2 * K(x, x_i) * (φ(x_{i+1}) - φ(x_{i-1}) - ψ(x_{i+1})
    + ψ(x_{i-1})); end do;
Id := IdentityMatrix(n + 1);
Temp1 := Matrix(n + 1);
for i from 0 to n do for j from 0 to n do Temp1(i + 1, j + 1) :=
subs(x = x_i, A_j); end do end do;
C := Id - Temp1;
Determinant(C);
F := Matrix(n + 1, 1);
for i from 0 to n do F(i + 1, 1) := f(x_i) end do;
Pr := Multiply(MatrixInverse(C), F);
for i from 0 to n do u_i := Pr(i + 1, 1) end do;
U := x -> add(A_i * u[i], i = 1..n) + f(x);
    return U(x);
end proc:
```

Fig. 2.Solution algorithm for the general equation

Example 1: Investigate the approximate solution of the Lineer Fredholm-Stieltjes İntegral Equation

$$u(x)=-\frac{1}{4}x^2-\frac{1}{6}\cdot x+x\sqrt{x}+\int_0^1\big(1+x^2s\big)u(s)\,d\Big(\ln\big(1+\sqrt{s}\big)\Big)$$

using generalized Trapezoid Method

Solution 1: Because of the difficulties with hand calculations, solution is done for n=4 and the results for other n's, n=4,16,64,256,512 ve 1024 the computer solution algorithm is used and the results are shown in table 1.

Here the inputs are,

$$K:=(x,s)\to 1+x^2\cdot s:\ ,\ f:=x\to -\frac{1}{4}\cdot x^2-\frac{x}{6}+x\cdot\sqrt{x}:$$

$$\varphi:=x\to\ln\big(1+\sqrt{x}\big):\ ,\ \psi:=0:\ ,\ n:=4:$$

$$a:=0.0:\ ,\ b:=1.0:\ ,\ \lambda:=1:$$

$$\text{For n=4, h}=\frac{\text{b - a}}{4}=0.25$$

$$x_1=0, x_2=0.25, x_3=0.50, x_4=0.75, x_5=1$$

Approximation solution of the equation is

$$A_0:=\frac{\lambda}{2}\cdot K(x,x_0)\cdot\big(\varphi(x_1)-\varphi(x_0)-\psi(x_1)+\psi(x_0)\big)$$

$$0.2027325540$$

$$\bar{u}(x)=0.2027325540\cdot\bar{u}(x_0)+\big(0.2673999983+0.6684999960x^2\big)\cdot\bar{u}(x_1)+$$
$$+\big(0.1091728042+0.5458640210x^2\big)\cdot\bar{u}(x_2)+\big(0.7917359200+0.5938019400x^2\big)\cdot\bar{u}(x_3)$$
$$+\big(0.3466823205+0.3466823205x^2\big)\cdot\bar{u}(x_4)$$

If the system is converted into matrix equation system and solve, then the coefficients of $\bar{u}(x_i)$ is determined as

$$\bar{u}(x_0)=0.357572869322348, \bar{u}(x_1)=0.434575488978139,$$
$$\bar{u}(x_2)=0.602470071969703, \bar{u}(x_3)=0.825115499197490,$$
$$\bar{u}(x_4)=1.08961478408291$$

If the determined coefficients are substituted in the previous equation, then the approximation solution is found as

$$0.285081208283520-0.101291418539437x^2-\frac{1}{6}x+x^{3/2}$$

Exact solution, Approximation solution and the determinant of the formed matrix are given for the other values of n in the Table I.

TABLE I. EXACT, APPROXIMATE SOLUTIONS AND DETERMINANT OF THE FORMED MATRIX

n	det(C)	ũ(x)	u(x)
4	0.2539	$0.2851-0.1013\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$
16	0.2615	$0.2748-0.1268\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$
64	0.2621	$0.2880-0.1287\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$
256	0.2622	$0.2965-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$
512	0.2622	$0.2992-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$
1024	0.2622	$0.3011-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$	$0.3058-0.1289\cdot x^2-\frac{1}{6}\cdot x+x\sqrt{x}$

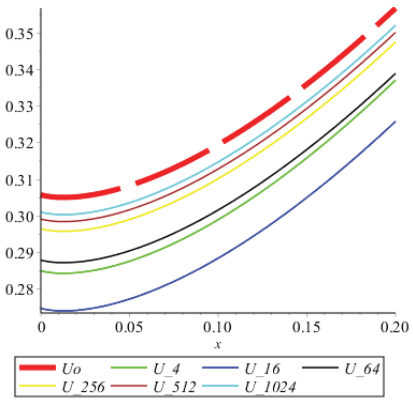


Fig. 3.Exact and Approximation solution of the equation in example 1 in the interval [0, 0.20]

Exact and approximation solutions of the equation for different values of n in example 1 in [0, 0.20]. Here the interval is taken very small to see the behavior of the graphs of the functions, clearly. Red dashed curve is the graphic of the exact solution. The blue one is the graphic of approximation solution for n=4 and n=16, 64, 256, 512 and 1024, respectively. It can be seen easily from the graphic that the graph for N =1024 is very close to the exact solution and also seen that the method is very effective to find approximate solution.

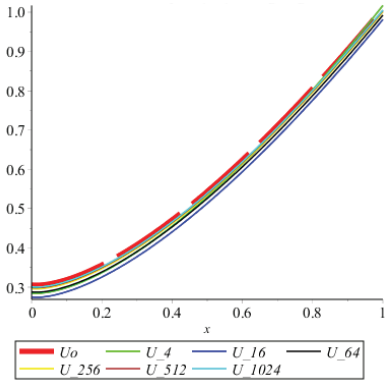


Fig. 4.Exact and Approximation solution of the equation in example 1 in the interval [0, 1]

Exact and approximation solutions of the equation in example 1 in [0, 1]. Red dashed curve is the graphic of the exact solution. The blue one is the approximation solution for n=4 and n=16, 64, 256, 512 and 1024, respectively.

Example 2: Investigate the approximate solution of the Lineer Fredholm-Stieltjes İntegral Equation

$$u(x)=\int_0^1\left(\sqrt{x}\sqrt{s}+x^2\cdot s^2\right)u(s)\,d\left(\sqrt{s}-\sqrt[3]{s}\right)-\frac{x+1}{x^2+1}-\sin(x)+x\sqrt{x}$$

using generalized Trapezoid Method.

Solution 2: The computations is done for n=4 and the results for n=4,16,64,256,512, and 1024 are shown in the table.

Here

$$K:=(x,s)\to\sqrt{x}\cdot\sqrt{s}+x^2\cdot s^2:\ ,\ f:=x\to-\frac{x+1}{x^2+2}-\sin(x)+x\cdot\sqrt{x}:$$

$$\varphi:=x\to\sqrt{x}:\ ,\ \psi:=x\to\sqrt[3]{x}:\ ,\ n:=4:\ ,\ a:=0.0:$$

$$b:=1.0:\ ,\ \lambda:=1:$$

For n=4, $h=\frac{b-a}{4}=0.25$

$$x_1=0,x_2=0.25,x_3=0.50,x_4=0.75,x_5=1$$

The approximation solution is

$$A_0:=\frac{\lambda}{2}\cdot K(x,x_0)\cdot\left(\varphi(x_1)-\varphi(x_0)-\psi(x_1)+\psi(x_0)\right)$$

0

$$A_n:=\frac{\lambda}{2}\cdot K(x,x_n)\cdot\left(\varphi(x_n)-\varphi(x_{n-1})-\psi(x_n)+\psi(x_{n-1})\right)$$

$$0.02126744630\sqrt{x}+0.02126744630x^2$$

$$A_i:=\frac{\lambda}{2}\cdot K(x,x_i)\cdot\left(\varphi(x_{i+1})-\varphi(x_{i-1})-\psi(x_{i+1})+\psi(x_{i-1})\right)$$

$$-0.02164843620\sqrt{x}-0.002706054525x^2$$

$$0.03090962872\sqrt{x}+0.01092820404x^2$$

$$0.03749619140\sqrt{x}+0.02435449072x^2$$

$$\bar{u}(x)=\left(-0.2164843620\sqrt{x}-0.2706054525x^2\right)\cdot\bar{u}(x_1)+$$
$$+\left(0.3090962872\sqrt{x}+0.1092820404x^2\right)\cdot\bar{u}(x_2)+$$
$$+\left(0.3749619140\sqrt{x}+0.2435449072x^2\right)\cdot\bar{u}(x_3)+$$
$$+\left(0.2126744630\sqrt{x}+0.2126744630x^2\right)\cdot\bar{u}(x_4)$$

If the system is converted into matrix equation system and solve, then $\bar{u}(x_i)$ coefficients are found as

$\bar{u}(x_0) = -.500000000000000, \bar{u}(x_1) = -.756685463022728,$

$\bar{u}(x_2) = -.838733329218575, \bar{u}(x_3) = -.781589585816435,$

$\bar{u}(x_4) = -.598603115938023$

Replacing the determined coefficients into the previous equation gives the approximation solution

$$-0.0515812711607650\sqrt{x} - 0.0388841927734743x^2 - \frac{x+1}{x^2+2} - \sin(x) + x^{3/2}$$

The approximation solutions and determination of the formed matrix are shown in the table TABLO II.

TABLE II. APPROXIMATION SOLUTIONS AND DETERMINANT OF THE FORMED MATRIX

n	det(C)	ū(x)
4	0.8978	$-0.0516\sqrt{x} - 0.0389 \cdot x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$
16	0.8863	$-0.0741\sqrt{x} - 0.0437x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$
64	0.5644	$-0.5900\sqrt{x} - 0.2650x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$
256	0.5648	$-0.5895\sqrt{x} - 0.2648x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$
512	0.5648	$-0.5895\sqrt{x} - 0.2648x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$
1024	0.5648	$-0.5894\sqrt{x} - 0.2648x^2 - \frac{x+1}{x^2+2} - \sin(x) + x\sqrt{x}$

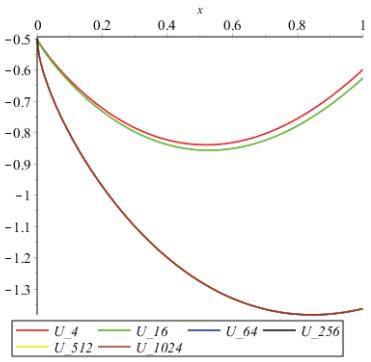


Fig. 5.Approximation solution of the equation in example 2 in [0, 1]

Approximation solutions of the equation in example 2 are given for different values of n in the interval [0,1]. Here to see the behaviors of graphics of the equation the interval is taken very small. Since the exact solution is not known, the estimation about exact solution can be done by looking behaviours of graphics. Here dark red indicates the approximation solution for n=4 and for n=16, 64, 256, 512, and 1024, respectively.

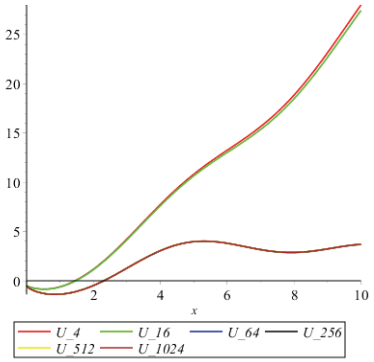


Fig. 6.Approximation solution of the equation in example 2 in the interval [0, 10]

Approximation solutions of the equation in example 2 are given for different values of n in the interval [0,10]. It is clearly seen that how the graph behaviors. It can be seen that the solution functions gets closer ans closer when N increases.

V. CONCLUSION

In many circumstances, it can be very difficult to find exact solution of The Integral Equations and it is required to find the approximate solutions. It can be seen through the numerical implementations, the proposed method is efficient and accurate to estimate the solution of these equations, also, it can be seen that when h increases, the absolute errors decrease to small values and the approximation solution approaches to the exact solution. It is planned, as a future work, to make implementation of the efficient proposed method to the other kinds of Stieltjes Integral Equations.

REFERENCES

[1] A.J. Jerri, Introduction to integral equations with applications, 2nd ed., Wiley pub, 1999, Canada

[2] A.M. Wazwaz, Linear and Nonlinear Integral Equations: Methods and Applications, Springer, New York, 2011

[3] V. K. Dzyadyk, Approximation Methods for Solutions of Differential and Integral Equations, Koninklijke Wohrmann, 1995, Netherland

[4] L.M. Delves, J.L. Mohamed, Computational methods for integral equations, Cambridge University Press, 1985, New York, USA

[5] I.Muslu, Lineer Fredholm-Stieltjes Integral Denklemlerinin Genellstirilmis Trapezoid Metodu Ile Yaklasik Cozumu, Ph.D Thesis, 2015, Biskek, Kyrgyzstan

[6] H. Wolfgang, Integral Equations; theory and numerical treatment, Birkhauser, Basel, 1995, Germany

[7] I.Muslu, A.Asanov, S.Yanik, Approximation Solution of Linear Fredholm-Stieltjes Integral Equations of Second Kind Using Generalized Trapezoid Quadrature Approximation Method, *Исследования по интегро-дифференциалным уравнениям*, B.46, 2014, pp.195-199.

Towards optimization of Availability and Cost in Selection of Geo-distributed Clouds Datacenter

Hasan Ziafat
PhD Candidate, Department of Computer
University of Kashan
Kashan, Iran
Ms.ziafat@gmail.com

Seyed Morteza Babamir
Associate Professor, Department of Computer
University of Kashan
Kashan, Iran
babamir@kashanu.ac.ir

***Abstract*— With increasing data clouds in different geographical areas, the availability of a datacenter and the cost of using the datacenter are two concerned factors of clouds users. The present research aims to present a method using K-means clustering and NSGA-II multi-objective algorithm to maximize availability and minimizes cost in selecting a datacenter. The proposed approach was applied to some real geographically distributed datacenters. Results showed that proposed approach outperforms greedy and random common algorithms.**

***Index Terms*— Geo-distributed Datacenter, Clustering, the K-means Algorithm, Multi-Objective Optimizing Algorithm, NSGA-II Algorithm.**

I. INTRODUCTION

Conventional clouds typology serves as a typology in which one or few datacenters are situated in small geographic zone. In this state, the distance between two datacenters and the final users has been potentially increased [1]. However its weakness is that it is highly risky to have a failure point and strengths is that the centralized datacenter is characterized with less operating costs and fast communication among datacenters. In the distributed cloud typology, a good number of datacenters are extended in the vast geographic zones and the users receive the data from a close datacenter and communicate with one another through internet [2]. The geo-distributed datacenter is usually of hybrid cloud type and focuses on infrastructures as the services and in comparison to the traditional datacenters they are small and large regarding size and number respectively. Each cloud system contains tens of geo-distributed datacenters. For example, Google has around 36 datacenters [3] and Amazon Web-service has more than 30 datacenters [4] all over the world. Considering geo-distribution of datacenter in different parts of the world, the present study focuses on selecting optimal datacenter based on resource characteristics that user requests with following conditions:

1) A datacenter is selected for the user that has the least geographic distance to the user`s position.
2) A datacenter is selected for the user that provides the highest availability coefficient for the user.
3) A datacenter is selected for the user with the least costs.

Each of mentioned conditions may negate the other, however, the aim is to select a datacenter that observes all of the mentioned conditions in an optimal way.

A. The Innovation

The present study introduces some innovations as the following:

1) So far, virtual machines allocation algorithms to the datacenters have focused on datacenters, but in the present study the focus is on geo-distributed datacenter.
2) In most previous works, the datacenter was selected randomly, but in the present study the data center position as well as geographic places were real and were borrowed from [5].
3) In the previous works, one user`s request was examined while here the massive/ volume request was also examined.

The present study used the hybrid clustering method and multi-objectives *NSGA-II* algorithm to select geo-distributed datacenter for improving response time to the user in an optimal way. The difference between multi-objective optimization and single-objective optimization is that in the later one the best response is usually the very fitness function maximum size but in the multi-objective optimization there is no single response to optimize all objectives simultaneously since there are conflicts regarding the objectives (that is one must decreases and the other increases). Therefore, the optimal response is the points in the best distance from the points to optimize every single objective individually.

The rest of the study is organized as the following: section 2 introduces the related works. Section 3 presents the proposed method. Section 4 includes the implementation and related experiment. Section 5 evaluates the proposed method efficiency and section 6 summarize the data and concludes.

II. RELATED WORKS

Related works on resource allocation are presented in three models [6]:

1) The Reservation Model: in this model, the user purchases a class of resource for a period of the time for example for one year.
2) The Spot Market Model: in this model, one-side auction market, the resources are consumed with less costs and flexibilities.
3) On-demand Access Model: in this mode, the user demands particular number of resources and pays the costs according to their prices.

The present study is limited to the third class that is On-demand Access Model. Many works have been conducted on resource allocation on centralized clouds, but here the works on geo-distributed cloud are reviewed.

In [7], the datacenter was presented as geo-distributed and the network aware resource allocation algorithm. Here, a comparison also made with traditional datacenter. In this study the virtual machines are homogeneous. The datacenter structure is like a tree and the 2 approximation method was applied to select optimal resources for user`s request, also the objective

was decreasing the maximum distance or latency among the datacenter. At the end, the heuristic algorithm was applied to divide user`s tasks among selected datacenter and compared the proposed method with greedy and random methods. The present study selected two experimental sets, one with 100 requests for 10 to 50 resource and the other 500 requests for 10 to 20 resources that in the datacenter section, the proposed algorithm (approximation 2) operated on average 79% better than the greedy and random algorithms and in task division among resource, datacenter traffic decrease in the greedy algorithm 10.2% better than random and in heuristic algorithm 4.6% better than in greedy algorithm.

In [8] network aware allocation resource algorithm was introduced and attempted to minimize communication cost and latency among servers. This study used two-dimensional knapsack algorithm. Since the resource allocation is NP-hard Problem, thus the study applied the heuristic algorithm called MCNVMA (Network-aware VM allocation algorithm based on Maximum Clique) and aimed to minimize the most communication latency in datacenter to solve the problem. Here, the MCNVMA algorithm was compared to the random and greedy algorithm and the typology and the user`s request was taken randomly and it was shown that the communication costs in datacenter is less than in two other algorithms. In previous works on recourse allocation in geo-distributed clouds, the main focus was on one parameter, for example reducing traffic or latency and other parameters were overlooked that is, there have not been any studies about combining effective parameters important to cloud providers and final users. Moreover, in present works, introducing massive request to the proposed algorithm was also overlooked. The present study aims to evaluate its proposed method for numerous users who request cloud as the same time.

III. Proposed Method

One of the most important conditions in resource allocation is cost of resource. In geo-distributed cloud structure, considering the datacenter extension, the first priority is finding the closest geographic point to the user`s position. Because selecting the closest datacenter to the user is one of the most important factor for reducing latency in geo-distributed cloud structure. If the mentioned datacenter meets the user`s need based on the number of requested resource with the least cost and the most availability, then the resource allocation operation to the user is accomplished. But sometimes, because the resource of datacenter is busy, it is necessary that communication with other centers be made in the same geographic zone. Therefore, by datacenter clustering structure, the user selects the cluster that has the closest characteristics to the user`s request. The datacenter structure similar to user`s request is put in one cluster to, if needed, meet some of the user`s request that cannot be met by the first datacenter.

A. Parameters of the Problem

The parameters considered for the datacenter selection in the present study are as the following: the mean of availability and cost.

1) *Availability:* that is the probability of access, whenever the user requests the resource on the datacenter how probable he accesses the resource.
2) *Cost:* the amount of money the user pays for his request for each resource based on memory unit, the processor and the bandwidth.

Base on the information form EC2 Amazon datacenter [9], the resources were classified regarding cost and availability. Equation 1. shows the datacenter classification.

$$Cost = \begin{cases} General & < 1\$ \\ Special & 1\$ \le Cost < 3\$ \\ Professional & \ge 3\$ \end{cases} \qquad (1)$$

$$Availability = \begin{cases} General & < 99 \\ Professional & \ge 99 \end{cases}$$

In the above mentioned classification, the general resource is defined for the user`s general affairs, the specialized resource for affairs that need optimal computing and professional resource for affairs the need graphic processes. In the proposed method, the characteristic of each datacenter are maintained in a record format so that whenever a datacenter is requested with particular characteristic by the user, the most appropriate one is selected and allocated. This Table used bite structure to specify datacenter cost type. Table I shows a sample of a datacenter characteristics. As it shown, the datacenter presents its resource with two general and professional price rate.

TABLE I. A Datacenter Characteristics Record

Availability	Cost			ID_DC
98	1	0	1	0001

Once datacenter state specified, the clustering structure is applied so that the similar datacenter is put in one class according to relation (1) characteristics. Up to now different clustering algorithms have been introduced, and among partition based clustering, the K-Mean algorithm [10] is also introduced that based on Table II, regarding time complexity in comparison to other partition-based algorithms it is the most appropriate one and it is appropriate regarding large scale data clustering [11].

TABLE II. Partition-Based Clustering Algorithm Comparison

Algorithm	Complexity time	Appropriate for large scale data
K-means[10]	$Low \quad O(knt)$	**Yes**
K-medoids[12]	$High \quad O(k(n-2)^2)$	**No**
PAM[13]	$High \quad O(k^3 \times n^2)$	**No**
CLARA[14]	$Middle \quad O(ks^2 \times k(n-k))$	**Yes**
CLARANS[15]	$High \quad O(n^2)$	**Yes**

B. Using K-Means Algorithm

This algorithm is one of the basic methods for other clustering methods such as Meta heuristics and phase clustering [16]. Although this algorithm is old enough and a good number of state-of-art clustering algorithms have been introduced, it is widely used because it is simple, easy to implement, and efficient [17]. The *K-Means* algorithm have been implemented in different methods and has similar processes. In the *K-means* clustering, Equation 2 as the fitness function is the finishing condition for the repeat algorithm that is the repeat continues until this relation become minimum. This relation is called Euclidean Distance and shows the sum of data distance from the related cluster.

$$E = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| X_i^{(j)} - c_j \right\|^2 \qquad 2$$

In the above equation $\| \ \|$ is the distance parameter between points to C_j center of j^{th} cluster, k is the number of clusters, X is the fitness vector for specifying the distance and n is the number of data. The K-means algorithm aims to minimize the distance between the components of a cluster and to maximize the distance between each separated cluster. In the proposed method,

after clustering it is necessary to user`s request be referred to the cluster near his geographic position and then in the related cluster, the closets datacenter, with characteristics appropriate to user`s request, be selected and allocated to the user. To select the closest datacenter to the user, and considering availability and costs parameters there is a need to use multi-objective evaluation algorithm. Among known multi-objective evaluation algorithms with two objectives, the *NSGA-II* algorithm has been able to have less response time, high availability, appropriate convergence, and high throughput [18].

C. Multi-Objective Optimizing NSGA-II Algorithm

The *NSGA-II* Algorithm was established in 2002 by Deb and his colleagues. This algorithm applied two concepts as non-dominated sorting and crowding distance [19].

1) The non-dominated sorting: In *NSGA-II* method the members of population are ordered based on non-dominating concept. In the non-dominated sorting, the members of the population who are not dominated in comparison to other members of that generation are ranked as 1 and are put in the first front and are excluded from the population. This procedure continues until all the members are put in different fronts.

a) The Pareto Dominance Concept: the $x \in R^n$ vector based on equation 3 dominate vector $y \in R^n$ that is x< y:

$$\forall i \in \{1,2,\dots,m\}: f_i(x) \le f_i(y) \ \ and, \\ \exists j \in \{1,2,\dots,m\}: f_j(x) < f_j(y) \qquad 3$$

The solution $x \in R^n$, based on equation 4 is the Pareto-optimal if and only if:

$$\nexists\, y \in R^n, y < x \qquad 4$$

2) Crowding Distance: after fronting the points, another parameter is used to evaluate the points present in one front and that parameter is called the crowding distance. The crowding distance is an alternative for *shared* in *NSGA* and that is why it is called *NSGA-II*. For each *x* present chromosome in the target front, *cd(x)* is attributed if it is equal to distance of that point from its adjacent points. First for each *K* fitness function, the points that have Min max values of this fitness function, the indefinite distance value is allocated and for other points *(i=2,3,...,(n-1))* the Equation 5 is used and the crowding distance are aggregated around the point according to Equation 6.

$$cd_f\left(x_{[i,f]}\right) = \frac{f_k\left(x_{[i+1,k]}\right) - f_k\left(x_{[i-1,k]}\right)}{f_k^{max} - f_k^{min}} \qquad 5$$

$$cd(x) = \sum cd_f(x) \qquad 6$$

In Equation 5, f_k is the fitness function and f_k^{max} and f_k^{min} are the maximum and the minimum for the function. The crowding distance for members of the population in each front is measured separately. The comparison among the distances is only made among the members of each front. In the proposed method, the *NSGA-II* algorithm shows the datacenter appropriate to the user`s request in ordered ranked and the user selects based on his number of request resource from the ranked datacenter.

IV. Implementing the Proposed Method

Based on the mentioned materials in the previous sections the procedures are divided into several phases: 1) normalizing datacenter data, 2) regulating a rule for specifying the number of clusters, 3) implementing *K-means* algorithm for arranging datacenter inside each cluster, 4) Find the closest Datacenter to user, 5) creating user`s request and specifying datacenter appropriate to the user`s request through *NSGA-II* algorithm. In the following each of these phases are described.

A. Phase one: Normalizing

The present study first randomly makes information from datacenter with longitude and latitude values, availability, and costs (for each processer unit, memory, and bandwidth). Since parameters do not have same material normal distribution among the data may not be fulfilled. Considering different normalizing methods, such as Min Max normalization, Z-score normalization, normalization by decimal scaling and so on and also the data type the present study selected the Z-score normalization because it can control the normal off distance data effect better [20]. Moreover, by using z-score normalization the mean absolute deviation is applied instead of standard deviation so that normal off distance data effect is lessen more. Thus, based on Equation 7, first the mean absolute deviation is measured and then by Equation 8 the data is normalized.

$$MAD = \frac{\sum |X_i - \mu_X|}{N} \qquad 7$$

$$norm\_data = \frac{X_i - \mu_x}{MAD} \qquad 8$$

In the above relation *N* is the number of whole data, μ_X is the mean for data and X_i is the data value.

B. Phase two: Regulating a Rule for Specifying the Number of Clusters

The United Nation has divided all countries into 9 geographic zones [21]. In this phase, 3844 datacenters in the world are classified based on their geographic zones [5]. The degree of the datacenter in each geographic zone is shown is Figure 1.

In each class based on mentioned parameters in 3 and Equation 9 the number of clusters are 9.

$$Count\ of\ Cluster\ = Cost\ St. \times Avail.\ St. \\ = 3 \times 3 = 9 \qquad 9$$

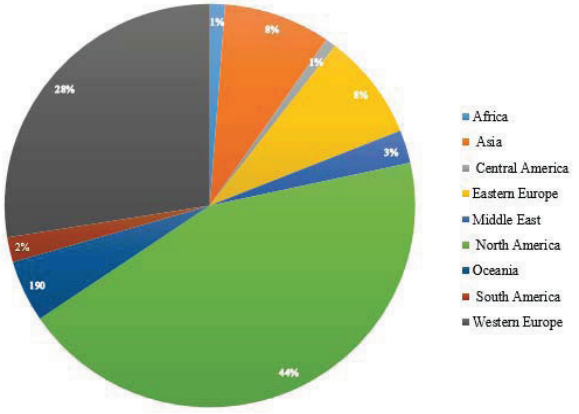


Fig. 1. 3844 datacenter distribution in 9 geographic zones

C. Third Phase: Clustering by K-Means Algorithm

The defined datacenter for each geographic zone in the previous section must be clustered by K-means algorithm. In this problem we have 9 zones and each zone`s datacenter must be divided among 9 clusters. In first level, for each 9 clusters one primary datacenter must be selected, then distance from all datacenter to each cluster`s primary datacenter be measured. To measure the distance from datacenters to head clusters, there are different method and based on Equation 2 the Euclidean

Distance was applied. Then based on the results, the datacenter with lesser distance to the primary head cluster is put in that cluster. Again, among the cluster members a new center is selected trough mean and the process repeats. The algorithm finishes when no more move made in the datacenter from one cluster to another. Otherwise, it again goes to the first level and the level is repeated. To specify the primary datacenter for each cluster, considering the geographic zone of the user for service request from cloud provider, the closets datacenter is selected for sending to user request. The reason to select this is that the faster the user`s request get to the service, the faster would be response process and the user`s request would meet faster as well. Although it is possible that the closest datacenter cannot meet the needs, the primary datacenter can send the user`s request to its same cluster datacenter and meet the user`s need because it is supposed to get a cluster from the datacenter.

D. Fourth Phase: Find the closest Datacenter to user

According to user`s longitude and latitude, the closets geographic point to him is selected. To specify the closest datacenter to the user, the user`s geographic distance to all datacenters considering longitude and latitude fields of the user and datacenter is measured. The geographic distance between two points on the land is measured by Equation 10 [22].

$$a = sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) * \cos(\varphi_2) * sin^2\left(\frac{\Delta\lambda}{2}\right)$$
$$c = 2 * atan^2\left(sqrt(a), sqrt(1-a)\right) \qquad 10$$
$$distance = R * c$$

In above Equation φ is the latitude and λ is the longitude and R radius of the planet (7371 kilometers [23]). For all datacenter this distance is measured. The closets geographic zone regarding distance is specified as the selected zone for cluster. Then based on user`s request the closets cluster to the user is selected based on requested parameters.

E. Fifth Phase: Implementing the NSGA-II Algorithm

In implementing the NSGA-II algorithm, the degree of appropriateness (or advantage) of each chromosome or datacenter is selected and is gained by the fitness function. Each chromosome is allocated a value based on SLA. Two QoS criteria of service as positive criterion and negative criterion are considered for the datacenter. Increasing values with positive criteria such as availability and decreasing values with negative criteria like cost is useful for the users. The fitness function must present increasing positive criteria and decreasing negative criteria. In addition, fitness function reflects user`s preferences as well. Equations 11 and 12 show the positive and negative criteria set respectively:

$$\zeta_i^-(CS) = \{Cost\} \qquad 11$$
$$\zeta_i^+(CS) = \{Availability\} \qquad 12$$

In the above Equations ζ_i^- and ζ_i^+ are positive and negative values of i^{th} datacenter. Generally to avoid SLA violation it is better to maximize the positive value and minimize the negative value. Moreover, user`s preferences for different criteria must be taken into account in the fitness function. Equation 13 shows the fitness functions.

$$F(CS^+) = \sqrt{(U_{Av} - D_{Av})^2}$$
$$F(CS^-) = \sqrt{(U_{Co} - D_{Co})^2} \qquad 13$$
$$F(CS) = [F(CS^+) - F(CS^-)]$$

In the above relation U_{Av} and U_{Co} are the availability value and cost value requested by the user and D_{Av} and D_{Co} are the availability value and mean costs value in the datacenter. The final value of fitness function is gained by two $F(CS^+)$ and $F(CS^-)$ output.

V. Simulation Result

This section shed lights on efficiency of the proposed algorithm. The proposed algorithm is applied for optimal selection of datacenter to host the resource requested by the user. This algorithm helps to select the most optimal datacenter based on user`s request. To do so, the proposed method was compared to two other algorithms as greedy and random algorithm. Three algorithms were implemented by MATLAB software 2015 and Cloudsim Simulator. In the random algorithm, first the user position is identified, then a case is selected from present datacenters in the user`s geographic zone and the user is directed to use that resource. Whenever the random selected datacenter cannot meet the user`s request, then again another datacenter is selected for the user. If datacenter resource of the zone be busy, the request is put in the line and waited for resource to get free in the datacenter zone. When resource get free, again form datacenter with free resource, the random sample is selected for the user. This continues until a datacenter for the user`s request be found. And finally the response time to the user request is measured. The greedy algorithm process is like that of random algorithm but it is different in that among present datacenter, it selects the biggest datacenter regarding free resource and introduced parameters in 3-A. In the following section, to measure the algorithm efficiency, the following scenarios are observed and each scenario are implemented and compared in the random, greedy, and proposed algorithms.

1) In the first scenario the number of users varies from 10000 to 100000 users and the number of fixed requested resource is 15. Figure 2-(a) shows the rate of cost of selected datacenter for the user by random, greedy, and proposed algorithms. It shows that the cost of random algorithm is more than that of two other algorithms. The proposed algorithm (*NSGAII_ Cluster*) could reduce the cost 82 percent and 76 percent in comparison to random algorithm and greedy algorithm respectively. Of course when the number of users is little the proposed algorithms does not operate optimally. The reason that cost reduces in the proposed method in comparison to greedy and random ones is the optimal datacenter selection and better decision making by multi-objective genetic algorithm.

2) In second scenario number of users is fixed with 50000 and the number of request resource varies from 5 to 50. As Figure 2-(b) shows the proposed algorithm (*NSGAII_ Cluster*) could reduce the costs 61 percent and 26 percent in comparison to random and greedy algorithm respectively.

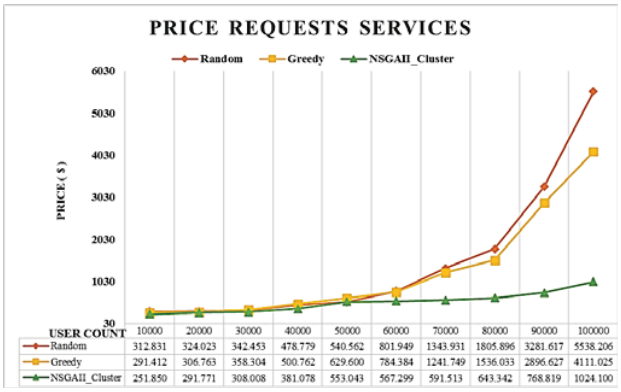
In above mentioned scenarios the resource is homogeneous and the datacenters are variable.

VI. Concluding Remarks

Selecting resource is integral and indispensable part of cloud management especially when the datacenter is geo-distributed. A high number of small datacenters with different geo-distribution makes geo-distributed cloud. The present study aimed to select the best datacenter based on user`s request among geo-distributed datacenters. Our proposed algorithm is a combination of K-Means clustering and the multi-objective optimal NSGA-II algorithm. One innovation in the present study is using a real datacenter for the simulation and the proposed method results were implemented for two scenarios as 1) the variable number of users and 2) the variable number of requests. To compare the results of proposed method, the random and

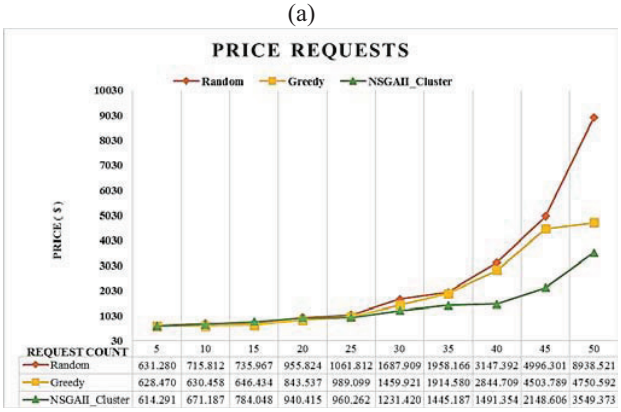
greedy algorithms were also implemented and it was shows the proposed method selected the best datacenter based on user`s request from geo-distributed datacenters better than random and greedy algorithms. Here are some suggestions for future works:

- The proposed algorithm be implemented with more complicated request from the user with more parameters and heterogeneous resource.
- Multi-objective algorithms or even many-objective algorithms be applied and the result be compared with results from the present study.



USER COUNT	10000	20000	30000	40000	50000	60000	70000	80000	90000	100000
Random	312.831	324.023	342.453	478.779	540.562	801.949	1343.931	1805.896	3281.617	5538.206
Greedy	291.412	306.763	358.304	500.762	629.600	784.384	1241.749	1536.033	2896.627	4111.025
NSGAII_Cluster	251.850	291.771	308.008	381.078	553.043	567.299	591.513	643.342	768.819	1024.100

(a)



REQUEST COUNT	5	10	15	20	25	30	35	40	45	50
Random	631.280	715.812	735.967	955.824	1061.812	1687.909	1958.166	3147.392	4996.301	8938.521
Greedy	628.470	630.458	646.434	843.537	989.099	1459.921	1914.580	2844.709	4503.789	4750.592
NSGAII_Cluster	614.291	671.187	784.048	940.415	960.262	1231.420	1445.187	1491.354	2148.606	3549.373

(b)

Fig. 2. Cost Rate, in Three Algorithms in (a) First Scenario (b) Second Scenario

REFERENCES

[1] Mohney D, Centralized vs. distributed data centers: Green, redundant, both, http://www.greendatacenternews.org/articles/473515/centralized-vs-distributed-data-centers-green-redu/, Access date: 03/26/2016.

[2] Testa S, Chou W, The distributed data center: front-end solutions, IT professional, 2004, 6, pp. 26-32.

[3] Miller R, Google Data Center FAQ, http://www.datacenterknowledge.com/archives/2012/05/15/google-data-center-faq/, Access date: Nov 28, 2015.

[4] Amazon, Amazon Global Infrastructure, http://aws.amazon.com/about-aws/global-infrastructure/, Access date: 02 Dec, 2015.

[5] Cooper B, Data Center Map, http://www.datacentermap.com/, Access date: 05/24/2016.

[6] Shang S, Wu Y, Jiang J, Zheng W, An intelligent capacity planning model for cloud market, Journal of Internet Services and Information Security, Innovative Information Science & Technology Research Group, 2011, 1, pp. 37-45.

[7] Alicherry M, Lakshman T. Network aware resource allocation in distributed clouds. The 32th IEEE International Conference on Computer Communications (INFOCOM), 2012, pp. 963-971.

[8] Yao Y, Cao J, Li M, A network-aware virtual machine allocation in cloud datacenter, In: Network and Parallel Computing, Springer,2013, pp 71-82.

[9] Amazon, Amazon EC2 Pricing, http://aws.amazon.com/ec2/pricing/, Access date: 04/05/2016.

[10] MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967, pp. 281-297.

[11] Xu D, Tian Y, A Comprehensive Survey of Clustering Algorithms, Annals of Data Science, 2015, 2, pp. 165-193.

[12] Park H-S, Jun C-H, A simple and fast algorithm for K-medoids clustering, Expert Systems with Applications, 2009, 36, pp. 3336-3341.

[13] Kaufman L, Rousseeuw PJ, Partitioning around medoids (program pam), Finding groups in data: an introduction to cluster analysis, 1990, pp. 68-125.

[14] Kaufman L, Rousseeuw PJ, Finding groups in data: an introduction to cluster analysis: John Wiley & Sons, 2009.

[15] Ng RT, Han J, CLARANS: A method for clustering objects for spatial data mining, Knowledge and Data Engineering, IEEE Transactions on, 2002, 14, pp. 1003-1016.

[16] Alpaydin E, Introduction to machine learning: MIT press, 2014.

[17] Jain AK, Data clustering: 50 years beyond K-means, Pattern recognition letters, 2010, 31, pp. 651-666.

[18] Marcel Cremenea MS, Denis Pallezc, D. Dumitrescu, Comparative analysis of multi-objective evolutionary algorithms for QoS-aware web service composition, Applied Soft Computing, 2015, 39, pp. 124–139.

[19] Jiang Z, Zuo L, Mingcheng E, Study on multi-objective flexible job-shop scheduling problem considering energy consumption, Journal of Industrial Engineering and Management, 2014, 7, pp. 589.

[20] Han J, Kamber M, Pei J, Data mining: concepts and techniques: Elsevier, 2011.

[21] World Regions Based on United Nations Country Grouping, http://www.internetworldstats.com/list1.htm#AF Access date: 05/13/2016.

[22] Veness C, Calculate distance, bearing and more between Latitude/Longitude points, http://www.movable-type.co.uk/scripts/latlong.html, Access date: 05/13/2016.

[23] Wikipedia, Earth radius, https://en.wikipedia.org/wiki/Earth_radius, Access date: 05/13/2016.

On the Methods of Collecting and Storing Big Network Traffic

Ramiz Shikhaliyev
Institute Information Technology ANAS
Baku, Azerbaijan
ramiz@science.az

***Abstract*— Collecting and storing network traffic of computer networks (CNs) is one of the major stages of the monitoring process. However, collecting and maintaining full network traffic in the modern CNs is a very complex problem. With rising speed and scale of the CNs and network traffic size, maybe needed petabytes storage a day. There are various methods for network data collecting and storing. Their correct choice can significantly reduce collected data size and required storage size. The article examines the issues of network data collection and storage.**

***Index Terms*— computer network, monitoring, network traffic, network traffic collection, network traffic storage.**

I. Introduction

Today, computer networks (CNs), especially the Internet, have become a global infrastructure for interactive, secure and everywhere accessible services. In order to provide high QoS (Quality of Service), effective monitoring infrastructure needed. In this case, the most proper method for the infrastructure of the CNs monitoring is a passive monitoring [1]. Passive monitoring finds the general condition of the CNs and safety, as well as the QoS level of provided services, etc. For this, it is necessary to constantly collect, store and analyze the network traffic, i.e.to constantly carry out the collection, storage and analysis of large data scale. However, this is a very difficult task, especially when monitoring of large CNs. Since, with rising speed and scale of the CNs and network traffic size, a day petabytes storage maybe needed. However, to conduct a comprehensive analysis of the condition and safety of the CNs, it is necessary to make the collection and storage of all traffic data. For example, at the CNs security monitoring collecting all network packets is needed to detect malicious activity or viral attacks (e.g., worms), etc.

It is known, that while passive monitoring of the CNs the large size of monitoring data collected. This leads to the problems associated with their storage, which cuts the efficiency of the analysis. Therefore, along with the need to develop new methods for analyzing large network traffic, the real problem is the development of new approaches to collect and store large network traffic for the CNs monitoring. Thus, reduction of network traffic feature dimension space, which is used to monitor the CNs [2] is very important. The article examines the issues of network data collection and storage.

II. Network Traffic Collection Methods

Today, available monitoring tools collect different types and size of information. At the same time, there are three main methods of traffic collection, which have different requirements to memory size: all packets' collection; network flow collection; and so-called enhanced flow collection.

The aim of all packets collection is the collection of all network traffic, which is generated by computers and devices of the CNs, while each packet header and transmitted information, is collected and stored for further analysis. This is collected data provides analysts with complete information about the traffic, i.e. about packet headers and transmitted information. Therefore, this monitoring data collection method may be the most multipurpose, since a large size of information can be intensively stored and processed [3].

Network flow is defined as IP-packets passing through the observation point in the network during a certain time interval. All packets belonging to a particular flow have a set of common properties. Requirements to IP-packet streams are defined in RFC 3917 [4] and according to the given definition the network stream is a set of streams of network packages for which the following conditions are provided:

– occur during the same time period;
– have the same source address and port number;
– have the same destination address and port number;
– use the same protocol.

Thus, if not to consider information transferred in packages and information of some fields heading of packages and to unite some packages then the data size is decreased, which leads to the reduction of the required memory for storage of flows. However, it leads to the reduction of quality of the analysis of network traffic [5].

Collection of the expanded flow includes collecting all packets and network flow. Thus, information of the packets headers or information transferred in packages is added to flow information. At the same time, the expanded flow may also contain more other information about any external sources, such as the geographic location of the source and destination IP address. Therefore, in some expanded flow collecting solutions, this information is considered as metadata [6].

Most of current research in the field of collecting network traffic is devoted to the questions of collecting packages in high-speed networks with minimum, data loss and to the

compression of data after collecting that is to decrease in size. For example, in [7, 8] the authors discuss questions of transformation of data for their effective storage and processing and monitoring in a cloud respectively. In works [9, 10] the authors offer approach to applications programming on data collection in high-speed networks based on standard hardware. And in work [11] for the full analysis of network traffic the authors offer a method of aggregation of flows.

III. Network Traffic Storage Methods

Other problem of the effective network traffic analysis is a storage of collected data which has to be stored long enough and reliably, and analytics may use them when needed. At the same time, depending on the place and the way of data storage the required memory size for storage can be much definitely changed. Also, the problems can be arisen connected with the administration and service, etc. However, data can be stored locally in the organization, in a cloud or other external storage, and various ways of data storing can be used, such as: files (e.g., logs); database, and their combinations. Each of these methods has its own aspects.

Usually, in most organizations CNs collects network traffic at several points. Therefore, it is very important to choose the place of the physical place of collected data.. For example, the centralized storage of all collected data in one place can simplify the management and analysis of data, but it requires a data transfer to the center, resulting in inefficient use of bandwidth network transmission channels. And this way of storage is impractical in terms of data security, since the unauthorized removal of data can occur when compromising storage. An alternative to the centralized data storage is the distributed data storage, but in this approach is complicated the process of data analysis, as well as administration and maintenance. One of distributed data storage types is cloud storage [12, 13], which can also implement data collection.

Most network traffic collecting tools record the obtained data in the files (log files) and usually have their own file formats. It is very important to know the format of the stored file to easily arrange the transfer of data between data collection and analysis applications, as the majority of them support certain file formats. However, there are some common formats (e.g., pcap), which are supported by most data collection and analysis applications. Thus, the format of the file may define the necessary size for file storage. Despite the fact that differences between the formats in small data are insignificant, but in large size data the choice of this or that format is essential. The reduction of memory size required to store the data can also be achieved by data compression, which can make the data storage and analysis more efficient. An efficient compression algorithm can not only cut the required disk space for data storage, but also reduce the time required to retrieve data from the disc. For example, lzo1x data compression algorithm can reduce the size of the records to about 50% [14].

Some data collection tools for storage data can use the database. Applications that support only files, can store it's own database or using data mining application and the analyst can do it manually. However, when using a database to store network traffic, it is necessary to consider the expected size and the number of records and database properties limiting total size of the database and record size, etc. Considering that relational databases aren't scaled as at data storage in the form of files, then NoSQL databases, such as Hadoop may be used to solve of the scalability problem [15].

IV. Conclusion

Today, network traffic monitoring of the CNs is one of the main tools to ensure their proper operation and safety, and data collection and storage is the basis of monitoring. For complete information on the activities of the CNs the constant and complete collection and storage of network traffic is needed, which allows prompt and effective response to refusals and security incidents. However, this needs a constant collection and storage of large size of monitoring data that may need excessive storage size which reduces the effectiveness of the analysis of the collected data. Another reason for the problem of collecting and storing large size of network traffic, in our opinion, is improper choice of proper data collection and storage methods.

The article analyzes the existing methods for collection and storage of network traffic, as well as problems existing in this field. As a result of the analysis we can say that collection of the network flow or enhanced flow requires much less memory for storage. This analysis can help CNs administrators to select the desired method according to the monitoring task.

References

[1] Shikhaliyev R.H. On methods and means of computer network monitoring, Problems of Information Technologies, vol.2, no.4, pp. 61-70, 2011.

[2] Shikhaliyev R.H. Ob odnom metode sokrascheniya razmernosti analiziruemyih priznakov setevyih trafikov, ispolzuemyih dlya monitoringa kompyuternyih setey, Telekommunikatsii, vol. 6, s. 44-48, 2011.

[3] Bejtlich R., Why Collect Full Content Data?. http://taosecurity.blogspot.com', 2012

[4] Quittek J., Zseby T., Claise B., Zander S., RFC 3917: Requirements for IP Flow Information Export (IPFIX). Internet Engineering Task Force, 2004. http://tools.ietf.org/html/rfc3917

[5] RFC 7011, Specification of the IP Flow Information Export (IPFIX) Protocol, a standardized network flow format, provides a more technical definition of flow. http://tools.ietf.org/search/rfc7011

[6] National Information Standards Organization (NISO). Understanding Metadata. NISO, 2004.

[7] Aceto G., Botta A., Pescape A., Westphal C. Efficient Storage and Processing of High-Volume Network Monitoring Data, IEEE Transactions on Network and Service Management, vol. 10, no. 2, pp. 162-175, 2013.

[8] Aceto G., Botta A., de Donato W., Pescape A., Cloud Monitoring: A Survey, Computer Networks vol.57, no. 9, pp. 2093-2115, 2013.

[9] Deri L., Cardigliano A., Fusco F., 10 Gbit Line Rate Packet-to-Disk Using n2disk, Proceedings IEEE INFOCOM. Turin, Italy, Apr. IEEE, pp. 3399-3404, 2013.

[10] Banks D., Custom Full Packet Capture System, SANS, 2013.

[11] Francois J., State R., Engel T., Aggregated Representations and Metrics for Scalable Flow Analysis, IEEE Conference on Communications and Network Security (CNS). Washington, D.C., pp. 478-482, 2013.

[12] Sivashakthi T., Prabakaran N., A Survey on Storage Techniques in Cloud Computing, International Journal of Emerging Technology and Advanced Engineering, vol. 3, no 12, pp. 125-128, 2013.

[13] Spoorthy V., Mamatha M., Santhosh Kumar B., A Survey on Data Storage and Security in Cloud Computing, International Journal of Computer Science and Mobile Computing, vol.3 no..6, 2014, pp. 306-313.

[14] Software Engineering Institute, Carnegie Mellon University. SiLK FAQ https://tools.netsa.cert.org/silk/faq.html (2014).

[15] White T., Hadoop: The Definitive Guide, O'Reilly Media, p. 768, 2015.

New Generation Networks Performance Analysis and Optimization

Zaychenko Yuriy, Gasanov Aydin
Institute for Applied System Analysis, KPI,
Kiev, Ukraine
zaychenkoyuri@ukr.net, ayding44@rambler.ru

Galib Hamidov
Azarishiq JSC, Information technologies department
Baku, Azerbaijan
galib.hamidov@azerishiq.az

***Abstract*--Problems of new generation networks (NG-) performance analysis is considered. The main Quality of Service indicators for NGnetworks are introduced and their models are constructed. The problem of NGN performance optimization under constraints on set values of QoS is formulated, its model constructed. The algorithms for NG networks performance analysis and optimization are developed. The experimental investigations of suggested algorithms were carried out.**

***Key words*-- NG network, Quality of Service (QoS), performance analysis, optimization**

I. INTRODUCTION

In recent years, due to a sharp increase in the volume of the traffic in the global computer networks and the increasing demands on the quality of information transfer new generation of computer networks (NGN) appeared and is implemented, in particular computer networks with MPLS technology [1]. Distinctive features of these networks are: 1) the high data transfer rate; 2) the introduction of various classes of flows (users); 3) introduction of quality of service (QoS), namely, the average delay in the delivery of packets and a packets loss ratio.

The emergence of a new generation networks put on the agenda demand for the development of appropriate models and methods of performance analysis and optimization of networks based on QoS indicators. These methods are used in the optimal (rational) design of NGN networks.

The aim of this work is to develop and investigate models and methods for analysis and optimization of next-generation networks.

II. ANALYTICAL MODELS FOR QOS ESTIMATION IN NGN

To solve the problems of the analysis and optimization of MPLS network using Quality of Service (QoS) indicators must first be developed analytical models for assessing the quality indices for different classes of service, depending on the intensity of the input flows, channel capacities, flows distribution (FD) over communication channels. Earlier in [2] the dependence of average packet delay of two priority classes in ATM networks on the intensity of the incoming flows and bandwidth of communication channels were obtained.

Generalize these models of QoS for arbitrary number of priorities.

Let it be one channel in which N classes of flows are transmitted with priorities P_i. Denote flow with priority i in channel (r,s) as f_{rs}^{i}, channel capacity - μ_{rs}. For convenience set priorities in such order: $p_0 > p_1 > \ldots > p_N$.

For obtaining analytical models for average delay of packets of k-th priority introduce the following assumptions:

1. Input flows in node of all classes are Poisson with intensity $h_{ij}^{(k)}$.
2. Service time in channels (r,s) is distributed by exponential law with parameter intensity μ_{rs} (Mbit/s), where μ_{rs} is a capacity of the channel (r,s).
3. Service time of a packet in different channels are statistically independent random variables.

Under such assumptions using queue theory write down the expression for delay in the channel (r,s) for flows of different priorities (classes) [2]:

$$t_{rs}^{0} = \frac{f_{rs}^{0}}{(\mu_{rs} - f_{rs}^{0})\mu_{rs}} \quad (1)$$

$$t_{rs}^{j} = \frac{\sum_{k=0}^{j} f_{rs}^{k}}{(\mu_{rs} - \sum_{k=0}^{j-1} f_{rs}^{k})(\mu_{rs} - \sum_{k=0}^{j} f_{rs}^{k})}. \quad (2)$$

Let be given demand matrix for flow transmission of the *l*-th priority $H_l = \left\| h_{ij}^{l} \right\|$. Using these expressions in [5] the final expression was obtained for average delay of k-th priority flow in a network:

$$T_{cp,k} = \frac{1}{H_{\Sigma}^{(k)}} \sum_{(r,s)\in E} \frac{f_{rs}^{(k)} \sum_{i=1}^{k} f_{rs}^{(i)}}{\left(\mu_{rs} - \sum_{i=1}^{k-1} f_{rs}^{(i)}\right)\cdot\left(\mu_{rs} - \sum_{i=1}^{k} f_{rs}^{(i)}\right)}, \quad (3)$$

where $H_{\Sigma}^{(k)} = \sum_{j=1}^{n}\sum_{i=1}^{n} h_{ij}^{(k)}$ is total intensity of input flow of the k - th priority (class);

$f_{rs}^{(i)}$ is the flow of i-th priority in the channel (r,s).

But the drawback of this expression is that it doesn't take into account the delays in routers connected with input packets processing and their switching. Therefore, generalize this expression at the general case.

Let input flows in routers (so-called LSR_i [1]) be Poisson with intensity $\{\lambda_{rs}\}$, and its intensity of service in LSRi be μ_i.

Let's assume that a node- LSRi is described by model M/M/n/1, where n is a number of input flows.

In [5] the following expression was obtained for average delay in MPLS network for the flow of the k -th class priority taking into account delays in all the routers:

$$T_{cp}^{(k)} = \frac{1}{H_{\Sigma}^{(k)}}\left(\sum_{(r,s)\in E} \frac{f_{rs}^{(k)} \sum_{i=1}^{K} f_{rs}^{(i)}}{\left(\mu_{rs} - \sum_{i=1}^{K-1} f_{rs}^{(i)}\right)\left(\mu_{rs} - \sum_{i=1}^{K} f_{rs}^{(i)}\right)} + \sum_{i=1}^{n} \frac{\Lambda_i^{(\kappa)}}{\mu_i - \Lambda_i^{(\kappa)}}\right) \quad (4),$$

Now obtain the expression for packets loss probability of different classes.

The probability of packets loss of the k -th class (priority) is equal to the probability of the state when all the virtual channels allotted for k -th class of flow in the channel (r,s) are occupied [2]:

$$P_{nom\,r,s}^{(k)} = P_0 \cdot \left(\frac{f_{rs}^{(k)}}{\mu}\right)^{n_k} \cdot \frac{1}{n_k!} \cdot \left(\frac{f_{rs}^{(k)}}{n_k \mu}\right)^{N_k}, \quad (5)$$

where μ is capacity of the base channel, e.g., $\mu_1 = 2.048\frac{Mbit}{s}$; n_k is a number of channels in the link(r,s) allotted for transmission of the k -th flow; N_k is a buffer size in LSR assigned for queue of k -th class of packets; P_0 is a normalizing multiplier.

Then the loss ratio of the- k -th class of packets will be equal to: *1*

$$PLR_k = 1 - \prod_{(r,s)\in E} \left(1 - P_{nom.(r,s)}^{(k)}\right), \quad (6)$$

Consider the general statement of flows distribution (FD) problem with the constraint on the average delay and packets loss ratio (PLR) which differs from known one [3] by account of QoS indicator PLR.

Let be given NGN network as a graph $G(X,E)$,

where $X = \{x_j\}$ is a set of nodes (routers), $E\{(r,s)\}$ is a set of channels, capacities of channels are given $\{\mu_{rs}\}$ and matrices of demands for transmission of all the classes of flows $H(k) = \|h_{ij}(k)\|\ i,j = \overline{1,n}$, where $h_{ij}(k)$ is an intensity of information flow to be transmitted from node x_i to node x_j (Kbit/s).

It's demanded to find such transmission routes and flows distribution for all the classes $F(k) = [f_{rs}(k)]$, under which the following constraints on average packets delay

$$T_{cp,k} \le T_{k,зад}, \quad (7)$$

and constraint on the packets loss ratio will be fulfilled

$$PLR_k \le PLR_{k,зад} \quad (8)$$

III. The Algorithm of Generalized FD Problem Solution

The suggested algorithm is based on several properties of optimal flow formulated below.

Statement 1. *The optimal flow which minimizes the average packets delay T_{cp} is a flow transmited over the shortest paths with conditional metrics* [2]

$$l_{rs} = \frac{\partial T_{cp,}(F)}{\partial f_{rs}^{(1)}}.$$

Statement 2. As it follows from formula (3) *the distribution of flows of lower priority doesn't influence the average delay of higher priorities flows*. This enables to construct an algorithm of flows distribution (FD) in which the distribution is performed sequentially: first find Fd for the flow of the highest priority, then find optimal FD for second priority flow using the channels bandwidth left after transmission of the first flow and so on.

The algorithm consists of K stages (by number of classes of service) at each of which the distribution of k-th class of flow is determined $F(k)$ using constraints (8) and (9).

1 Stage

0 step. Initialize $F_1(0) = 0$; $H_1(0) = 0$.

This stage consists of $2C_n^2 = n(n-1)$ iterations at each of which flow distribution from next demand h_{ij} is searched, $i,j = \overline{1,n}$, $i \ne j$.

1–st iteration

1. Find the initial conditional metrics:

$$l_{rs}(1) = \lambda \frac{\partial T_{cp,1}}{\partial f_{rs}^{(1)}} + (1-\lambda)\frac{\partial PLR_1}{\partial f_{rs}^{(1)}},$$

where $\lambda \in [0;1]$.

As an initial value of λ it's possible to take $\lambda = 0{,}5$.

2. Determine the shortest paths in the chosen metrics between all the nodes - $\pi_{ij}^{\min}(1)$.

3. Choose the first demand in matrix $H_1 = \left\| h_{ij}^1 \right\|$, for example $h_{i_1 j_1}$. Find the shortest path $\pi_{i_1 j_1}^{\min}$ and distribute flow from the demand $h_{i_1 j_1}$ and find the initial flows distribution:

$$f_{rs}^{(1)}(1) = \begin{cases} f_{rs}^{(1)}(0) + h_{i_1 j_1} = h_{i_1 j_1}, if\ (r,s) \in \pi_{i_1 j_1}^{\min}; \\ f_{rs}^{(1)}(0) = 0, otherwise \end{cases}. \quad (8)$$

End of the first iteration. Go to the second iteration.

r -th iteration

Let $(r-1)$ iteration be executed, flows of the first $(r-1)$ demands $f_{rs}^{(1)}(r-1)$ are found.

1. Determine the conditional metrics:

$$l_{rs}^{(1)}(r) = \lambda \frac{\partial T_{cp,1}}{\partial f_{rs}^{(1)}} + (1-\lambda) \frac{\partial PLR_1}{\partial f_{rs}^{(1)}} \mid f_{rs} = f_{rs}^{(1)}(r-1) \cdot \quad (9)$$

2. Choose the next demand $h_{i_r j_r}$ from matrix $H(1)$ and find the shortest path $\pi_{i_r j_r}^{\min}$ in the metrics $l_{rs}^{(1)}(r)$.

Distribute the flow from demand $h_{i_r j_r}$ over the path $\pi_{i_r j_r}^{\min}$ and find the new flow $F_1(r)$:

$$f_{rs}^{(1)}(r) = \begin{cases} f_{rs}^{(1)}(r-1) + h_{i_r j_r}^a, if\ (r,s) \in \pi_{i_r j_r}^{\min}; \\ f_{rs}^{(1)}(r-1), otherwise \end{cases}, \quad (10)$$

where $h_{i_r j_r}^a = \min\left\{ h_{i_r j_r}; Q_{рез}\left(\pi_{i_r j_r}^{\min}\right) \right\}$ -is a portion of demand $h_{i_r j_r}$ which is transmitted by the path $\pi_{i_r j_r}^{\min}$.

End of r -th iteration.

The rest of iterations of the first stage are fulfilled similarly up to the full exhausting of demands in matrix $H(1)$. Denote the obtained flow $F_1 = \left[f_{rs}^{(1)} \right]$.

Check up the fulfillment of the following constraints:

$$T_{cp}(F_1) \le T_{1,зад}, \quad (11)$$

$$PLR(F_1) \le PLR_{1,зад}. \quad (12)$$

If the constraints (11) and (12) are fulfilled then end of the first stage and go to the next stage 2. Otherwise, perform the additional optimization of the flow F_1.

If after this step at least one of the constraints (11), (12) won't be fulfilled, then this problem is unsolvable at the given channels capacities.

K stage

Let $k-1$ stages were performed and flows distribution from the first $(k-1)$ demands $F_1, F_2, ..., F_{k-1} = \left[f_{rs}^{(k-1)} \right]$ were found. Find the distribution of the k -th class of flows. The stage consists of $n(n-1)$ iterations like the first stage.

1. Take the first demand $h_{i_1 j_1}$ out of matrix $H_k = \left\| h_{ij}^k \right\|$. Find the shortest path in the metrics (9) $\pi_{i_1 j_1}^{(k)\min}$.

2. Determine the capacity reserve of the path $\pi_{i_1 j_1}^{(k)\min}$:

$$Q_{рез}\left(\pi_{i_1 j_1}^{\min}\right) = \min_{(r,s) \in \pi_{i_1 j_1}^{\min}(k)} \left\{ \mu_{rs} - \sum_{i=1}^{k-1} f_{rs}^{(i)} \right\} - \varepsilon.$$

3. Distribute the flow of demand $h_{i_1 j_1}^{(k)}$ with value $h_{i_1 j_1}^{(a)}$, where

$$h_{i_1 j_1}^a = \min\left\{ h_{i_1 j_1}^{(k)}; Q_{рез}\left(\pi_{i_1 j_1}^{\min}\right) \right\}$$

and calculate new flow value:

$$f_{rs}^{(k)}(1) = \begin{cases} f_{rs}^{(k)}(0) + h_{i_1 j_1}^{(k)}, if\ (r,s) \in \pi_{i_1 j_1}^{(k)\min}; \\ f_{rs}^{(k)}(0) = 0, otherwise \end{cases}$$

End of the first iteration.

The following iterations are performed similarly for the rest of demands in matrix $H(k) = \left\| h_{ij}(k) \right\|\ i, j = \overline{1,n}$.

In result obtain flows distribution of the k-th class $F(k) = \left[f_{rs}(k) \right]$.

Check the fulfillment of the constraints:

$$T_{cp}(F(k)) \le T_{k,зад}, \quad (13)$$

$$PLR(F(k)) \le PLR_{k,зад}. \quad (14)$$

If both constraints are fulfilled then STOP, the end of algorithm.

In case if $T_{cp}(F(k)) > T_{k,зад}$ then the corresponding FD problem is unsolvable under given channels capacities and requirements on given values of QoS $T_{kзад.}$

IV. Problem of Optimal Capacities Choice And Flows Distribution

For ensuring of transmission of all the input flows with given values QoS under arbitrary demand matrices it's necessary to solve combined problem of traffic engineering in which the optimal channels capacities and flows distributions of all classes should be found simultaneously.

Problem statement.

Let MPLS network structure be given as an oriented graph $G = (X, E)$, where $X = \{x_j\}\ j = \overline{1,n}$ is a set of nodes (routers), $E = \{(r,s)\}$ is a set of channels, set of channels

capacities $D = \{d_1, d_2, ..., d_k\}$ and their costs of unit length $C = \{c_1, c_2, ..., c_k\}$ are also given.

Let it also be given demand matrices of input flows of corresponding classes $H = \left\| h_{ij}^{(k)} \right\|$, $i, j = \overline{1,n}$, $k = \overline{1,K}$ and constraints on average packets delay for the k-th flow $T_{cp,k}$, $k \in K_1 \subset K$, and the constraint on packets loss ratio for different classes of flows are established.

It's required to choose such channels capacities $\{\mu_{rs}^{(0)}\}$ and to find the flows distributions of all the classes $F(k) = [f_{rs}(k)]$ for which total cost of NG network would be minimal and the established constraints on given values of QoS be fulfilled completely. Mathematical model of this problem is following:

Find

$$\min C_\Sigma = \sum_{(r,s)\in E} C_{rs}(\mu_{rs}), \quad (15)$$

under constraints

$$T_{cp,\kappa}(F(k), \mu_{rs}) \le T_{зад,k} \; k = \overline{1,K}, \quad (16)$$

$$PLR(F_k) \le PLR_{k,зад}. \quad (17)$$

Describe the algorithm of solution the problem of capacities choice and flows distribution (CCFD) for NGN network. It consists of preliminary stage and finite number of iterations.

At the preliminary stage find initial channels capacities $\{\mu_{rs}(0)\}$ and flows distributions of all classes $F(k)$ Then go to the first iteration.

$(l+1)$ iteration

Let l iterations be already performed and current capacities $\{\mu_{rs}(l)\}$, flows distributions $F_k(l) = [f_{rs}^{(k)}(l)]$ and total network cost $C_\Sigma(l)$ were found.

The goal of iteration (l+1) is the optimization of channels capacities and flows distribution by criterion of total cost minimization C_Σ and check of optimality condition.

1. For given values of capacities $\mu_{rs}(l)$ solve the problem FD and find new flows distributions of all classes:

$$F_{(k)}(l+1) = [f_{rs}^{(k)}(l+1)] k = \overline{1,K}.$$

2. For new flows $F_{(k)}(l+1)$ solve the problem of optimal capacities choice (CC) [5] and find new channels capacities $\{\mu_{rs}(l+1)\}$ and total network cost

$$C_\Sigma(l+1) = \sum_{(r,s)\in E} \mu_{r0} C_{rs}(l+1).$$

3. Compare: if $|C_\Sigma(l) - C_\Sigma(l+1)| < \varepsilon$,

where ε is given accuracy, then end of the algorithm.

Found capacities $\{\mu_{rs}(l+1)\}$ and flows distribution of all classes $F_k(l+1)$ are optimal, otherwise put $l = l+1$ and go to the next iteration.

V. Experimental Investigations of Suggested Algorithms

For experimental investigation of suggested algorithms FD and CCFD the corresponding software products were developed which were included in software modeling complex NGN NETBUILDER.

All the experiments were performed at the global MPLS network in Ukraine which consists of 15 nodes, 19 channels and 3 classes of flows.

In experiments demand matrices H(k)were varied by multiplying its components on the corresponding coefficient k.

The first experiment consisted in varying of coefficient k_1 for the first class of flow. This flow has the highest priority 0. The results of experiment are presented in the table 1.

TABLE 1. Average Delay for Class 1 Flow Versus Intensity of Input Flow

Coefficient k_1	**Average delay for class_1, sec.**
10	0,0002667
20	0,0006484
30	0,0012629
40	0,0023985
50	0,0068935
54	0,023741
55	0,0831388

TABLE 2. Average Delay for Flows of Class 2 And Class 3

Coefficient k_2	**Average delay** (class_2, sec)	**Average delay** (class_3, sec)
0,4	0,0002933	0,001187
0,5	0,0003697	0,003468
0,52	0,0003888	0,0058
0,53	0,0003982	0,009874
0,535	0,0004029	0,016674
0,54	0,000407	0,152685
0,5401	0,0004071	0,198146
0,5402	0,0004072	0,285892
0,5403	0,0004073	0,527018
0,54035	0,0004073	0,929201
0,540354	0,0004073	0,990247

The next experiment consisted in varying traffic coefficient k_2 for demands of flow class 2. This class has priority 1 and it is distributed in the capacity band left after distribution of flow class1. The results of experiment are presented in the Table 2.

In the next experiments the influence of routers productivity on the total average delay of flows class 1 and 2 was explored.

The corresponding results are presented in the Fig. 1.

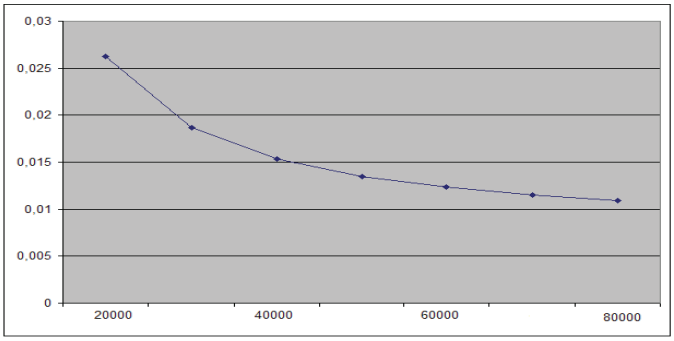


Fig.1. The average flows delay versus intensity of service (productivity) of routers

As it was expected the with the increase of routers productivity the total delay of packets of different classes falls down.

Further the comparison of the suggested CCFD algorithm and algorithm CCFD for ATM networks suggested in [3] was carried out. Initial data were the same, coefficient of traffic k was varied.

The experiments were performed at the same NGN network, number of flow classes was equal k =3. During the experiments were given matrices $H(\kappa)$ and constraints on average delays $T_{cp,1}$, $T_{cp,2}$ were introduced and intensity of input flows varied. The initial demand matrices were multiplied by coefficient k, $k \in [0{,}5 - 2.0]$. The experimental results - total network cost dependence on total intensity of input flows $C_{\Sigma} = f(k)$ for both methods is presented in the Fig. 2.

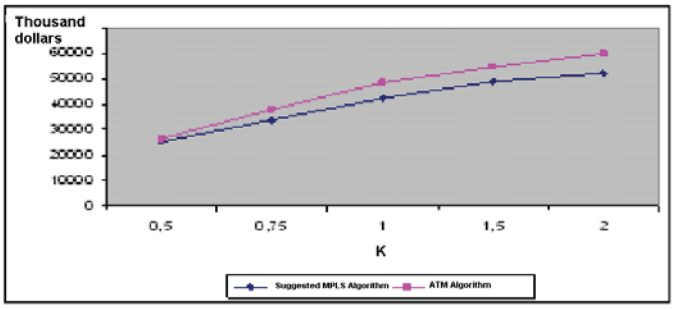


Fig.2. Network cost versus total intensity of input flow

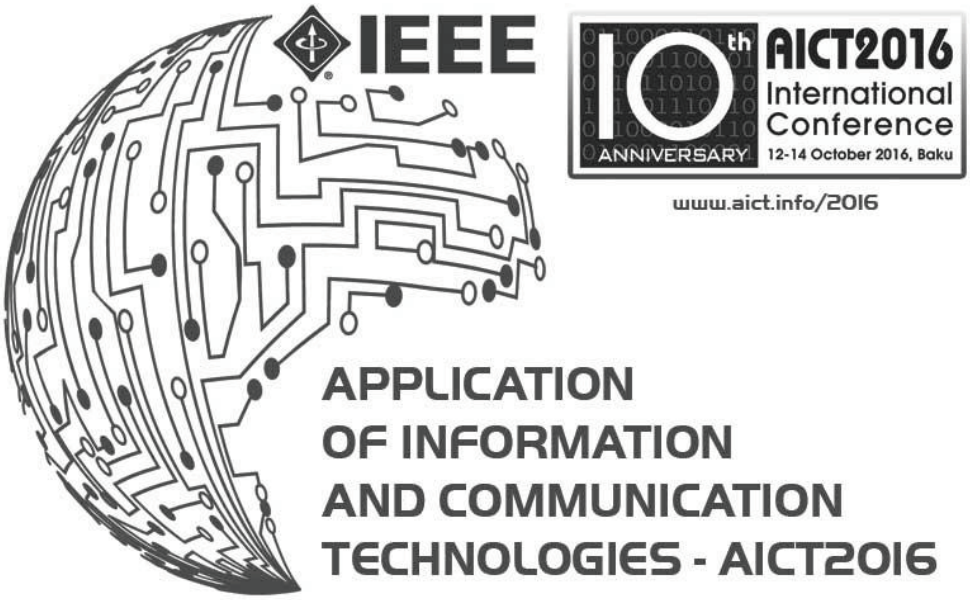
The analysis of obtained results shows of advantages of the suggested method CCFD over known one developed for ATM network [3]. The cut in network cost for the suggested algorithm increases with the growth of intensity of input flow and attain 10 ÷ 11 % of total network cost.

VI. Conclusion

1. In the paper problems of analysis and optimization of NGN networks are considered and investigated. The problem of different classes flows distribution under constraints on quality of service (QoS) indicators was formulated.
2. New algorithm FD of its solution was suggested which differs from known by accounting of delays in routers and distribution of flows by two QoS criteria- -average packets delay and packets loss ratio.
3. The combined problem of optimal channels capacities choice and flows distribution for NGN networks was considered and corresponding algorithm of its solution was suggested.
4. The experimental investigation of the suggested algorithms were carried out and comparison with known algorithm was performed.

References

[1] Vivek Alwayn. Advanced MPLS design and implementation. – Cisco Press, 2002. – 480 p.
[2] L Kleirock. Computer networks with queues. M.: Publ. house “Mir”, 1979. – 600 p.
[3] Zaychenko H. ATM networks: Modeling, analysis and optimization. – Kiev, Publ. house “VIPOL”, 2003. – 224 p (rus.).
[4] L Kleirock. Queueing systems. Volume 1. John Wiley and Sons. - 1975. – 432 p.
[5] Zaychenko H. Complex of models and algorithms of computer network optimization with MPLS technology. // System research and Information technologies, 2007. – №4.-pp. 58–71 (rus)



CONFERENCE PROCEEDINGS

SESSION 6.

ICT IN GOVERNANCE, BUSINESS ADMINISTRATION, FINANCE AND ECONOMY

12-14 October 2016, Baku, Azerbaijan

E-Services Security Assessment Model

Yadigar Imamverdiyev
Institute of Information Technology
Azerbaijan National Academy of Sciences
Baku, Azerbaijan
yadigar@lan.ab.az

Abstract— **Information security plays an extremely important role in e-services, as the e-services are vulnerable to various kinds of security and privacy threats on the Internet. The paper presents a formal model of e-services considering four different abstraction levels: application, middleware, operating system and network. Various security threats to e-services and security services to prevent these threats are described. The security level of e-services is determined by the weighted sum of security ratings of individual security services.**

Index Terms—**e-government, e-services, information security, security services, security assessment.**

I. Introduction

Electronic services (e-services) become generally accepted method of government or private sector services across Internet to wide range of users. It is known that one of the main problems of Internet is information security, and so information security is becoming a key factor in the providing of e-services [1].

E-services (and other distributed systems) are more vulnerable to various threats on the Internet than other systems because of their high heterogeneity, dynamic nature, decentralized management, remote access of users, spreading across multiple administrative domains with different security policies, use of local and global operations, etc. Modern e-services are characterized using a large volume of sensitive information that needs to be protected (personal data, billing information, health-related data etc.). Ensuring information security in such systems is not a trivial task and is essential for wide public acceptance of electronic services [2, 3].

The two main application fields of e-services are e-business (e-commerce) and e-government. Depending upon the types of interacting participants e-services can be classified into five classes [4]: G2C (Government to Citizens); G2G (Government to Government); G2B (Government to Business); B2C (Business to Consumer); B2B (Business to Business); C2C (Consumer to Consumer) e-services.

Ensuring information security of e-government services has its own specific requirements. Governmental organizations play a dual role – they are providers of e-services; at the same time they act as a consumer of e-services, by accessing the information of other governmental agencies. Other users of e-government services are commercial organizations and individuals.

Information security systems of e-services use an extensive set of security mechanisms: firewalls, user identification and authentication tools, intrusion detection and prevention systems, anti-virus tools, as well as organizational measures. However, assessment of the level of security provided by these measures is a very difficult task. Providers and users of an e-service are constantly interested in the question: how secure (or not secure) is the e-service at the moment? For this reason, methods for modeling and quantifying the security of e-services are required [5].

The paper proposes an assessment model for the security of e-services. One approach to information security evaluation is to assess compliance of information security ensuring processes with information security requirements. Information security requirements describe what should be done to guarantee the security of a system. These requirements can be determined from the best practices, standards, and regulations. The basis for this is the analysis of information security risks. Of course, it is impossible to compile a complete list of security requirements. Some e-services have their own specific security requirements (eg, PCI DSS for e-commerce, HIPPA for e-health services). Despite these comments, we are convinced that it is possible to identify some core security services common to all e-services (similar to X.800 recommendations) and create for them appropriate security requirements.

The paper is organized as follows. Section 2 presents brief review of related works. In Section 3, modified client-server architecture for e-services is proposed. Various security threats to e-services and security services to prevent these threats are described in section 4 and 5, respectively. E-service security assessment model is described in Section 6.

II. Related Works

In the past two decades, e-services security issues attracted some attention of researchers. [6-11]. The first survey of security issues in e-services appeared in 2000 [4]. This work introduced a classification framework of e-services, and proposed a model for analysis of security requirements in the e-services applications.

The problem of ensuring strong security when using e-government services was underlined in early 2000s, and [6] analyzes the last decade developments and implementations in this field, and summarizes topical security challenges for government e-services.

[7] proposes a multi-layered security architecture for e-government web sites based on the data access patterns, and evaluates the current security level of selected Jordanian e-government web sites. The author analyzes possible problems that face the expansions of e-government services and makes some recommendations on solving these problems.

A secure access to e-government systems is essential for security and trust in the e-government services. [8] studies the implementation of secure electronic identification (eID) used for access to e-government services in one medium sized Swedish municipality school. The paper also highlights the argument that secure eID systems are basis for the increased use of e-services and lead to greater legitimacy of the e-government services.

Web services are widely used as a "linking bridge" between e-services and their consumers [9]. They exchange large amounts of data more efficiently and cheaper than other technologies between applications running in different platforms. As a result, fast, convenient and productive communication is established between the providers and consumers. [10] proposes a model for studying transaction behavior of web services and analyzing the underlying transaction protocols and verifying their security properties.

[11] developed a "comprehensive framework" for integrating technical and organizational security services into "e-Government maturity models (eGMMs)". The framework integrates information security maturity model stages into eGMM. The integral model provides an approach by which governmental organizations can provide secure e- services.

Security requirements for e-services can be defined in the security policy by the service provider. However, the service customer may have specific security requirements, which are not reflected in this policy. In [12] the concept of security policy personalization to the requirements of individual consumer of e-services was introduced.

Assessment of information security is usually related to risk assessment, but it is hard to determine information security risks accurately. Moreover, information security ensuring activities cover broad issues and it is not correct to confine its assessment with risk assessment solely.

The scientific research and practical initiatives on information security assessment are intensified since 2000. The corresponding scientific-research and practical activity trends are defined as information security metrics [13]. In recent years, several research papers are written on information security metrics and measurement models. R.Savola suggests taxonomy for the security metrics [14]. Heyman and co-authors suggest using the notion of security patterns for the interpretation of information metrics and the outcomes of those [15]. In his acclaimed book [16], Andrew Jaquith introduces interesting ideas regarding the productivity of activities. The authors [17] analyze the characters of risk as a security metrics, and as alternative metrics, suggest the metrics based on compliance to criteria, detection of interventions, policy and incident.

Several national and international standards are developed on information security metrics. NIST SP 800-55 – the principles on security metrics for information systems, NIST SP 800-80 – the principles for development of performance productivity metrics for information security, ISO/IEC 21827 – maturity model of the potential for security engineering of the systems and above mentioned ISO 27004 standard can be mentioned.

III. E-SERVICE MODEL

Many definitions can be found in the literature for "e-service". In the broadest sense, e-service can be thought as the provision of any kind of services over electronic networks such as the Internet [4]. A more specific definition of e-service can be found in [18] as: "Internet-based applications that fulfill service needs by seamlessly bringing together distributed, specialized resources to enable complex, (often real-time) transactions."

E-services consist of three main components: the service provider, service customer and service delivery channel. Based on these components, e-services model can be built from three basic parts: the service provider's system, the service customer's system, and the communication network. E-services are distributed systems and can be viewed as a client-server architecture. In [5] a client-server architecture for e-services was proposed. Here, we modify this model by introducing a middleware layer between application and operating system layers. Our proposal is motivated by the need to take into account specific security requirements for various classes of e-service. In addition, the security mechanisms in distributed systems are often placed in a middleware layer. The operating system serves as a platform for deploying middleware and for running applications. The layered model is shown in figure 1.

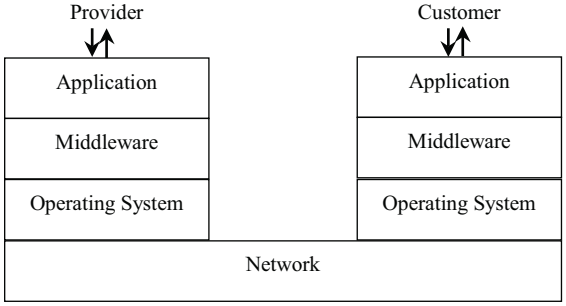


Fig. 1. The client-server model for e-service

Within the client-server architecture, the client application on the customer's system communicates with the server application on the provider's system. Each application runs on top of their respective middleware layer. Data is stored and used by applications using the mechanisms provided by the operating system. Data messages are exchanged between customer and provider applications using the facilities provided by operating systems. Operating systems transmit the message to another system via the network. At the other end, a message is sent back up through the layers to the destination application. It is possible to refine this model adding other layers and sub-layers, however, the proposed level of abstraction is sufficient for our purposes.

IV. E-Service Security Threats

This section briefly discusses general definitions for information security risks, threats and vulnerabilities in e-service systems.

According to ISO/IEC 27005:2011 standard [19], *information security risk* refers to "the potential that a given threat will exploit vulnerabilities" of an e-service, and cause harm to the provider or consumer of the e service. Information security risks could be measured in terms of a combination of the probability and consequence of a risk.

Information security threats refer to any possible actions, or events that may lead to violations of information security policy. Threats exploit vulnerabilities within e-service systems. Types of information security threats are very diverse and have a variety of classifications.

Information security vulnerability refers to flaws or weaknesses in the design or implementation of the e-service, including security measures that may be intentionally or unintentionally used to provide adverse effects on the information assets of the e-service or its operations.

A e-service, like all information systems, is subject to different types of security risks. We consider the following set of security risks for e-services:

Disclosure – release of information to a party, who does not have access rights to it;

Unauthorized access – access to e-services and information resources, by a person who does not have access rights to them;

Modification – unauthorized changes to the content of the information, or unauthorized modification of sequence or route of messages;

Misuse – use of e-service or related resources in a way that was not intended;

Abuse – abusive use of e-service or related resources by legitimate users [5];

Fraud – "misrepresentation of identity or intention in using the provided service" [5];

Repudiation – denial by legitimate users that they have used the service;

Denial of Service – legitimate users cannot get access to the provided service or the access is difficult.

Various security services are available, which can be used to counteract the aforementioned threats to information security of e-services.

V. E-Service Security Services

ITU-T X.800 Recommendations "Security Architecture for OSI" defines a security service as [20]: "a service that is provided by a protocol layer of communicating open systems and that ensures adequate security of the systems or of data transfers". X.800 classifies security services into five categories: authentication, access control, data integrity, data confidentiality, and non-repudiation [20].

A more general definition is given in RFC 2828 [20]: "a processing or communication service that is provided by a system to give a specific kind of protection to system resources." Security policy requirements are implemented by security services, in its turn, security services are implemented using one or more security mechanisms.

Authorisation or access control service provides protection against unauthorized use of e-services and resources. In deciding whether to grant or refuse the requested access this service may use access control lists, access tickets, passwords or other authentication information.

Authentication service verifies authenticity of communicating parties and data origin. Authentication can be one-way (usually the client proves its identity to a server) and mutual. An authentication service may use a simple authentication mechanism such as user passwords. Currently, commonly accepted approaches to strong authentication are based on the Kerberos system or directory services with X.509 certificates [4].

Accountability service provides evidence of who did what, and when.

Availability service ensures that the service is available at all times that is needed.

Confidentiality service protects against the unauthorized disclosure of information: allows access to sensitive data only users with the appropriate access rights, and prevents unauthorized disclosure of information to users who does not such rights. Usually, the confidentiality of information is ensured by encryption.

Integrity service prevents deliberate or accidental unauthorized changes to data. To ensure the integrity the system must detect unauthorized modification of data. The service employs cryptographic hash functions, and message authentication codes to counteract integrity violation threats.

Non-repudiation service provides proof of origin and proof of delivery of message. An assurance is provided by a trusted third party, which has enough information to its assurances could be trusted. Typically, non-repudiation mechanism relies on digital signatures.

Auditing service provides logging of events, which affect the security of e-services, and provides an analysis of accumulated data in order to detect attempts to violate the information security. A complete list of events to be analyzed is determined by the security policy. With the active audit suspicious actions are monitored in real time.

VI. E-Service Security Assessment Model

Security assessment is carried out in two stages:

1. Evaluation of the security provided by a particular security service in the e-service.

2. Evaluation of the security of e-services in general.

At the first stage security rating of a particular security service is evaluated. Ranking of security service, depending on the security level that it is capable of is defined.

Assume that the information security requirements of e-services are grouped into m directions (security services). For each i-th security service, partial information security indicators are defined. They are presented in the form of questions, answers to which make it possible to determine the assessment M_{ij}, $i=1,..,m$, $j=1,\ldots,n$. These indicators allow

forming the assessment of S_i group indicators (security rating of a particular security service).

Assessment of a partial information security indicator is performed on the following scale:

0 – requirements are not met;

0.25 – requirements are met in an insignificant scope;

0.5 – requirements are met in a modest scope;

0.75 – requirements are almost fully met;

1 – requirements are fully met.

Group indicators S_i, i=1,..,m are formed by averaging the partial indicators assuming that they are of equal importance (weight):

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} M_{ij},$$

where M_{ij} – jth partial indicator of the ith security service; n_i – number of all partial indicators of the ith security service.

When partial indicators are of various importance then group indicators are calculated as follows:

$$S_i = \sum_{j=1}^{n_i} \alpha_{ij} M_{ij},$$

where α_{ij} – importance coefficient (weight) of the jth partial indicator of the ith security service.

The current level of information security of e-service (SL) is determined by the weakest link rule:

$$SL = \min_i (S_i).$$

REFERENCES

[1] R. Roland and K.N. Lemon, "E-service and the consumer", International Journal of Electronic Commerce, vol. 5, no. 3, pp. 85-102, 2001.

[2] D. Hector and P. Puyosa, “e-Government: Security Threats,” IEEE eGovernment STC, 2011

[3] A. Jacobi, M. Lund Jensen, L. Kool, G. Munnichs and A. Weber, "Security of eGovernment Systems", Conference Report !PI AlSTOAlFWC/2008–096/LOT4/Cl/SCI0

[4] M. Mehta, S. Singh, Y. Lee “Security in e-services and applications,” Network Security: Current Status and Future Directions, 2000, pp. 157-177.

[5] R. Boutaba, B. Ishibashi, B. Shihada, “A Nnetwork management viewpoint on Ssecurity in e-services,” E. Nardelli,Maurizio Talamo (eds), Certification and Security in Inter-Organizational E-Services: IFIP 18th World Computer Congress, 2003, pp. 17-45.

[6] P. Dorogovs, A. Romanovs, “Overview of government e-service security challenges,” IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2015, pp. 1-5.

[7] I. Alsmadi, “Security challenges for expanding e-governments’ services,” International Journal of Advanced Science and Technology, vol. 37, pp. 47-61, 2011.

[8] M. Gustafsson, E. Wihlborg “Safe online e-Sservices building legitimacy for e-government. A case study of public e-services in education in Sweden,” JeDEM-eJournal of eDemocracy and Open Government, vol. 5, no. 2, pp. 155-173, 2014.

[9] S.Indrakanti, “Service Oriented Architecture Security Risks and their Mitigation,” DSTO-TN-1126, Defence Science and Technology Organisation, Commonwealth of Australia, 2012.

[10] S. Ranise, “On the verification of security-aware E-services,” Journal of Symbolic Computation, vol. 47, no. 9, pp. 1066-1088, 2012.

[11] G. Karokola, S. Kowalski and L. Yngström “Secure e-government services: Towards a framework for integrating IT security services into e-government maturity models,” Information Security for South Africa, 2011, pp. 1-9.

[12] G. Yee, “Personalized security for e-Services,” Proc. of the 1st International Conference on Availability, Reliability and Security (ARES), 2006, pp. 140-147.

[13] W. Jansen “NISTIR 7564: Directions in Security Metrics Research.” 2009, 26 p.

[14] R. M. Savola “Towards a taxonomy for information security metrics,” Proc. of the 2007 ACM workshop on Quality of protection (QoP’07), pp-28-30, 2007.

[15] T. Heyman, R. Scandariato, C. Huygens, W. Joosen “Using security patterns to combine security metrics,” Proc. of the 3rd International Conference on Availability, Reliability and Security (ARES’08), pp.1156-1163, 2008.

[16] A. Jaquith “Security metrics: replacing fear, uncertainty, and doubt.” Addison-Wesley, 2007.

[17] O. S. Saydjari “Is Risk a good security metric?,” Proc. of the 2nd ACM workshop on Quality of Protection, pp. 59-60, 2006.

[18] P. Seybold “Preparing for the e-services revolution: Designing your next-generation e-business.” 1999.

[19] ISO/IEC 27005:2011 Information technology — Security techniques — Information security risk management. 2011.

[20] W. Stallings, “Cryptography and network security principles and practice”, 5th ed., NY, Pearson Education, Inc., 2011. pp. 809-812.

A Fuzzy Logic Approach to Predict the Best Fitted Apparel Size in Online Marketing

Efendi Nasibov
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
efendi.nasibov@deu.edu.tr

Alper Vahaplar
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
alper.vahaplar@deu.edu.tr

Murat Demir
Department of Textile Engineering
Dokuz Eylul University
Izmir, Turkey
murat.demir@deu.edu.tr

Burak Okur
Department of Computer Science
Dokuz Eylul University
Izmir, Turkey
burak_okur@hotmail.com

***Abstract*— Online marketing has been showed dramatic increase related with the usage proportion of internet. Availability of wide range of products and comparing different brands and products with just a few click, makes online marketing even more desirable. Apparel is one of the product group that physical experience directly affect consumer buying behavior. On the other hand, return of online purchased apparel cause extra cost and time for suppliers and consumers. Size of apparels may differ based on brands and lead consumer in confusion. In this study, we aimed to compare measurements of the consumer and the product regardless of brands' size using with fuzzy approaching manner and concluded with a fitness ratio in terms of fuzzy numbers. By this way, online shoppers will be able to find best fitted products for their body measurements in each brand.**

***Index Terms*—Online apparel marketing, Apparel size, Fuzzy logic, Fuzzy relations.**

I. INTRODUCTION

Online marketing has showed dramatic increase within the recent years with the help of many advantages versus offline shopping. This new shopping tradition brings many advantages either online retailers or consumers. Lack of physical experience is the biggest challenge for online marketing especially for the products that buying behavior of consumers have strict relation with physical experience. Many consumers use offline and online shopping channels together in order to overcome the issues that caused from this problem [13].

Apparel online marketing is one of the product groups that consumer satisfaction is directly related with physical experience [1]. Poor fit is counted amongst the number one reason for dissatisfaction of consumers and the barrier of apparel online retailing [2-5]. When the consumer is not happy with the online purchased product, they return it and, it means different costs for online suppliers such as customer service costs, shipping costs, labor and material, disposal of written-off item and markdown on item re-sale, lost customer loyalty, potential lost revenues and margin on initial sale [5].

On the other hand, providing more information for products aid to create customer value. Because of the reason that consumers choose the online suppliers which create better value for them, creating customer value is the key strategic point for online suppliers for growth and survive [6,7]. Wodruff [7], is also claimed that there is a direct relation between customer value and customer satisfaction.

In order to reduce return costs as well as increase consumers' satisfaction, many technology and textile companies (Cyberwave, Human solutions, Browzwear, Optitex, Lectra) have been investigating on 3D virtual-try-on for offering more information to online customers. Although, using 3D modelled apparels and avatars for online apparel sales provide more information to consumers [8], this may negatively influence consumers buying behavior [15].

A part from size selection with the help of 3D modelling of user and products, many start-up businesses used different principles for right size decision. True-fit", "Virtusize", "Clotshorse" companies use body type of online shoppers and their previous size selection in order to decide right size for their future purchasing. "Fitbay", "Rent the Runway" classify online shoppers based on similar body types and make size recommendation based on similar consumers' profiles [14,15].

There are some Fuzzy Logic and other soft computing technologies in the literature that have been used for predicting vary physical properties of textile products [11,14]. In this study, fuzzy logic methods are used in order to define the most suitable size of a product for a specific user.

This study is supported under the project number 115E194 by The Scientific and Technological Research Council of Turkey (TUBITAK).

II. Fuzzy Membership

In discrete mathematics, each element is definitely member or is definitely not a member of a class A. The membership function is as follows:

$$\mu_A(x) = \begin{cases} 1, & x \in A \\ 0 & x \notin A \end{cases} \quad (1)$$

In fuzzy set theory, elements of a fuzzy set are mapped to a universe of membership values using a function-theoretic form. This function maps elements of a fuzzy set A to a real numbered value on the interval 0 to 1. Membership of an element in the universe to the fuzzy set A uses a mapping given by

$$\mu_A(x) \in [0,1] \quad (2)$$

In the study, fuzzy numbers method – which is a variation of fuzzy sets – is used for body size measurements. Fuzzy numbers can be defined in many forms as seen in Fig. 1.

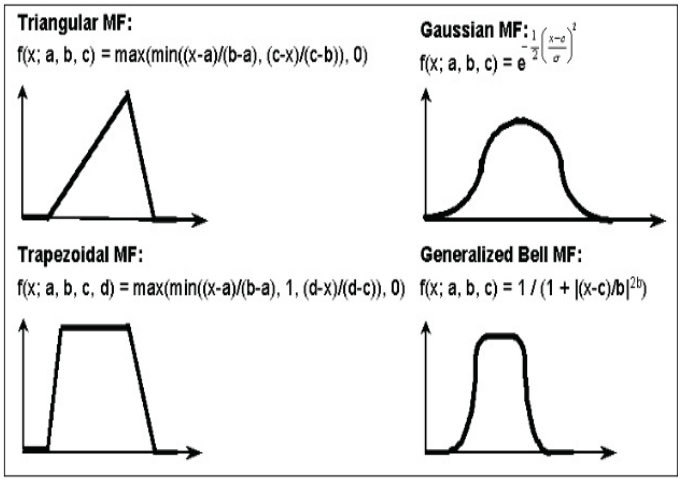


Fig. 1. Different Types of Fuzzy Numbers.

For simplicity, triangular membership functions are used in this study. The triangular fuzzy numbers is defined as T(a, b, c):

$$\mu_T(x) = \begin{cases} 0, & x < a \ \ or \ \ x > c \\ (x-a)/(b-a), & a < x \le b \\ (c-x)/(c-b) & b < x \le c \end{cases} \quad (3)$$

where a is the most left point, b is the core (with membership degree equal to 1) point and c is the rightest/right point of the fuzzy number (Fig. 1).

III. Method

Take into consideration of such a system that is designed for n clients $C = \{c_1, c_2, \dots, c_n\}$, and m kinds of products on different sizes $P = \{p_1, p_2, \dots, p_m\}$. Each product is characterized by k attributes corresponding to client's measurements $A = \{a_1, a_2, \dots, a_k\}$. So the relation between clients and related attributes: $R^1: C \times A \to (-\infty, +\infty)$, can be given as the matrix below:

$$R^1 = \begin{bmatrix} r_{11}^1, & r_{12}^1, \dots, & r_{1k}^1 \\ r_{21}^1, & r_{22}^1, \dots, & r_{2k}^1 \\ & \dots \dots \dots & \\ r_{n1}^1, & r_{n2}^1, \dots, & r_{nk}^1 \end{bmatrix} \quad (4)$$

where $r_{ij}^1 \in (-\infty, +\infty)$ is the j^{th} attribute value of i^{th} client. On the other side, the interval of a certain product to enclose a certain attribute value is given in relation matrix as: $R^2: A \times P \to [0,1]$.

$$R^2 = \begin{bmatrix} r_{11}^2, & r_{12}^2, \dots, & r_{1p}^2 \\ r_{21}^2, & r_{22}^2, \dots, & r_{2p}^2 \\ & \dots \dots \dots & \\ r_{k1}^2, & r_{k2}^2, \dots, & r_{kp}^2 \end{bmatrix} \quad (5)$$

where r_{ij}^2 is a fuzzy number specifying the fuzzy interval of the i^{th} attribute for the j^{th} product. In this case, elements of the relation according to "min-min" composition $R^3 = R^1 * R^2$ can be defined as the degree of fitness of a specific client according to their attribute values to a specific product.

$$R^3 = \begin{bmatrix} r_{11}^3, & r_{12}^3, \dots, & r_{1p}^3 \\ r_{21}^3, & r_{22}^3, \dots, & r_{2p}^3 \\ & \dots \dots \dots & \\ r_{n1}^3, & r_{n2}^3, \dots, & r_{np}^3 \end{bmatrix} \quad (6)$$

In the matrix above, r_{ij}^3 shows the degree of fitness of i^{th} client to j^{th} product where

$$r_{ij}^3 = \min_{l=1,\dots,k} (r_{il}^1 \wedge r_{lj}^2) \quad (7)$$

and

$$r_{il}^1 \wedge r_{lj}^2 = \mu_{lj}(r_{il}^1) \quad (8)$$

In other words, $\mu_{lj}(r_{il}^1)$; is the corresponding degree of attribute value r_{il}^1 for j^{th} product's l^{th} attribute. It is should be noted that each attribute of a certain product is given by the fuzzy number μ_{lj}.

In this study, each size of a brand is considered as a separate product, and all separate products are numbered as *1, 2, ... , p*. As an example, a specific brand *X* (a product) supposed to have different sizes as "Small", "Medium", "Large" and "XLarge". The measurements for these size of brand X are given in Table 1.

TABLE 1. SIZE MEASUREMENTS FOR BRAND X

	Small	Medium	Large	XLarge
Collar	35-38	39-40	41-42	43-44
Shoulder	38-42	43-44	45-46	47-48
Chest	90-98	99-104	105-110	111-114
Waist	40-46	47-52	53-54	55-56

The following rules are considered for constructing suitable fuzzy intervals:

- Measurements that smaller than client's attribute value is not suitable for the client's body size.
- Most suitable size for the client is the value of 96.5% of upper value of the related size (counting into account of ease).
- Larger sizes for a client also have some fitness degree, but smaller sizes are considered as NOT suitable for a client.

The fuzzy numbers constructed for sizes "Small", "Medium", "Large" and "XLarge" are in form of triangular fuzzy numbers defined as T(a, b, c), where T denotes one of the sizes of "Small", "Medium", "Large" or "XLarge", *a* is the bottom limit of the all sizes, *b* is the optimum measurement for the handled size and *c* is the upper limit of the related size (see Fig. 2). The following figure is an example of membership function used for collar circumference for all sizes in a brand. Functions for shoulder, chest and waist have similar shape but different lower and upper limit and core point values.

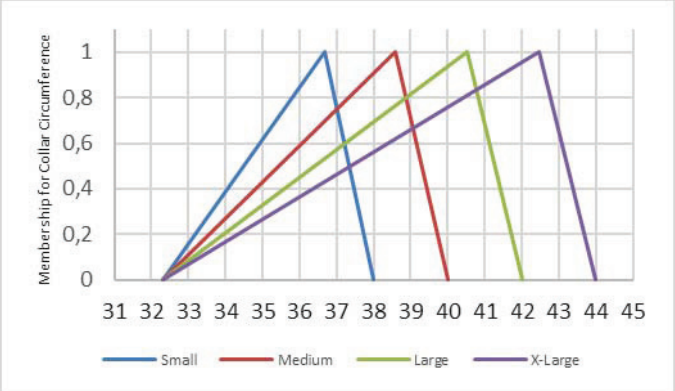


Fig. 2. Membership function example for collar circumference

IV. APPLICATION

In the study, a web interface is constructed for the users to enter their body size measurements (weight, length, collar circumference, shoulder width, chest width, waist circumference, arm length). Also, the measurements of shirts of certain brands are kept in a database. In order to find the optimum size of product for a user, a triangular membership value is computed using the measurements of the user and the size measurements of the shirts.

The system calculates all the class membership values using measurement values of the user (collar, shoulder, chest and waist measurements in cm) comparing with all sizes. For all sizes, the minimum membership value is considered. The user is assigned to the maximum valued size among these minimum membership values.

As an example, suppose that a client *C* has measurements as *C* (collar = 38cm, shoulder = 44cm, chest = 104cm, waist = 94cm). By default, the lower limit of the size is calculated as 85% of the upper limit for convenience. This lower limit value is used as the lower limit for all sizes in related measurement. The most fitting value of the size is 3.5% less than the upper limit. For example the parameters for "Small" size of "Collar" measurement is (32.3, 36.67, 38). The parameter table for "Collar" in "Small" size is shown in Table 2.

TABLE 2. PARAMETERS FOR "SMALL" SIZE IN "COLLAR" ATTRIBUTE

Size (Collar)	a	b	c
Small	32.3	36.67	38
Medium	32.3	38.6	40
Large	32.3	40.53	42
X-Large	32.3	42.46	44

Using Equation (8) for "Small" size with parameters (32.3, 36.67, 38), we get $\mu_{Small,Collar}(38) = 0$.

$$\mu_{Small,Collar}(x) = \begin{cases} \dfrac{x - 32.3}{36.67 - 32.3}, & 32.3 < x \le 36.67, \\ \dfrac{38 - x}{38 - 36.67}, & 36.67 < x \le 38, \\ 0, & otherwise. \end{cases}$$

If the membership of this client to "Medium" size class is calculated, the parameters (32.3, 38.6, 40) is used. For medium size of client *C*, we get $\mu_{Medium,Collar}(38) = 0{,}905$.

$$\mu_{Medium,Collar}(x) = \begin{cases} \dfrac{x - 32.3}{38.6 - 32.3}, & 32.3 < x \le 38.6, \\ \dfrac{40 - x}{40 - 38.6}, & 38.6 < x \le 40, \\ 0, & otherwise. \end{cases}$$

In the same way, values for "Large" and "XLarge" ($\mu_{Large,Collar}(38)$ and $\mu_{XLarge,Collar}(38)$) are also calculated. This operation is applied for the other attributes of the user. The degree of fitness for this client *C* to size class "Small" is computed as:

$$\mu_{C,Small} = \mu_{Small,Collar}(38) \wedge \mu_{Small,Shoulder}(44) \wedge \mu_{Small,Chest}(104) \wedge \mu_{Small,Waist}(94)$$

After computing class membership values for each product (i.e. “Small”, “Medium”, “Large”, “XLarge”) for client *C*, the product (size) with the highest membership value is assigned as the most suitable product (size) for the client. In other words, size that provides

$$\max(\mu_{C,Small}, \mu_{C,Medium}, \mu_{C,Large}, \mu_{C,XLarge})$$

is the best fitting size for the users within the given parameters.

Table 3 shows an example classification for a user. The user is said to be best fitted in "Large" size of this brand.

TABLE 3. MEMBERSHIP VALUES FOR A USER

Measurement	Small	Medium	Large	X-Large
Collar (38 cm)	0,00	0,42	0,32	0,26
Shoulder (44 cm)	0,00	0,00	0,55	0,45
Chest (104 cm)	0,00	0,00	0,55	0,45
Waist (94 cm)	0,00	0,39	0,33	0,29
Member to	0,00	0,00	**0,32**	0,26

In the application developed, also some graphical contents are displayed to the user in terms of size measurement classes according to the choice of brand and sizes in that brand as in Figure 3 below.

Fig. 3. Fuzzy membership values represented to user in the web site

The system users can also display and compare all the brands according to size fitness values for himself/herself (Figure 4 and Figure 5).

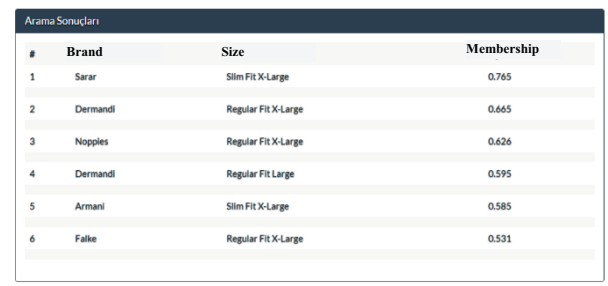


Fig. 4. List of most suitable sizes for the user in terms of selected brands.

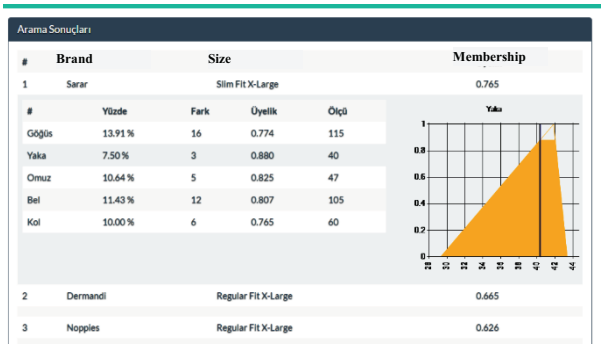


Fig. 5. Details and graphics of selected brand/size/measurement

REFERENCES

[1] M. Wagner, "A Comparative study of 3d media applications in B2C e-commerce" 2001. [Conference on Experiential E-Commerce: 3-D Visualizations, Product Simulations, Embodied Agents and Virtual Environments, September 27-29, 2001, East Lansing, Michigan, USA]

[2] F. Cordier, W. Lee, H. Seo, N. Thalmann," Virtual-Try-On on the web", 2001 [VRIC, Virtual Reality International Conference, Laval Virtual 2001, May 16-18]

[3] B. Beck, "Key Strategic Issues in Online Apparel Retailing".Available at: www.yourfit.com [Last access on December 15, 2015] 2000.

[4] J. Kim, A.M. Fiore, & H. Lee, "Infleunces of online store perception, shopping enjoyment and shoping involvement on consumer patronage behavior towards online retailer." Journal of Retailing and Consumer Services, 2007, pp. 95-107.

[5] C. Ratchcliff, "How fashion ecommerce retailers can reduce online returns." 2014.[Online] Available at: https://econsultancy.com/blog/65026-how-fashion-ecommerce-retailers-can-reduce-online-returns/ [Last access on 25 December 2015].

[6] R. Oliver, "Whence consumer loyalty". Journal of Marketing special edt. vol 63. 1999, pp. 33-44.

[7] R. Wodruff, "Customer value: the next source for competitive advantage." Journal of Academic and Marketing Science, vol.25. edt. 139. 1997, pp. 139-153.

[8] Fiore, A. & H., J., 2003. Influence of image interactivity on approach responses towards an online retailer. *Internet Research,* pp. 33-48.

[9] M. Kilic, "Determination of fabric drape using image analysis and fuzzy logic methods." Industria Textila. vol. 66(3), 2015,pp. 269-277.

[10] M. Hadizadeh, M. Tehran, A. Jeddi, "Application of an adaptive neuro-fuzzy system for prediction of initial load-extension behvior of plain woven fabrics." Textile Resarch Journal. vol. 80(10), 2010,pp. 981-990.

[11] S. Rwawiire, A. Kasedde, I. Nibikora, G. Wandera, "Prediction of polyester/cotton ring spun yarn unevenness using adaptive neuro fuzzy inference system." Journal of Textile and Apparel, Technology and Management. vol.8(4), 2014, pp. 1-12.

[12] A. Wong, Y. Li, P. Yeung, "Predicting clothing sensory comfor with artificial intelligence hybrid models. Textile Research Journal. vol.74(1), 2004. pp. 13-19.

[13] www.pwc.com, 2013. Demystifying the Online Shopper, 10 Myths of Multichannel Retailing, Available at: http://www.pwc.com/multicahnnelretailsurvey. [Last access on May 10, 2016].

[14] Consumer reports, 2014. Consumerreports.org.
Available at: Consumerreports.org, [Last access on May 11, 2016].

[15] V. Woolaston, "The end of online returns? App lets you see how clothes will fit based on items you already own" 2014. http://www.dailymail.co.uk. [Online] Available at: http://www.dailymail.co.uk/sciencetech/article-2582685/The-end-online-returns-Shopping-app-lets-clothes-fit-based-items-own.html [Last access on May 10, 2016].

Static Model of Equilibrium Prices on Mutual Debts of Subjects of Global Economy

Yadulla Hasanli
Institute of Control Systems ANAS,
Scientific Research Institute of Economic Studies Azerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
yadulla.hasanli@unec.edu.az

***Abstract*—The features of non-payment problem of global economy, including mutual debts have been indicated in the matrix form in the paper. Afterwards this problem has been converted into linear programming setting and solved by the Simplex method. Note that, LP enables us to find the market value of the existing debts in the market. The calculated equilibrium prices of mutual debts mean that potentially there are no debts to be repaid. In other words, it is required to find the price level that ensures the equality of receivables and payables on subjects. The constructed model has a static characteristic and applied using the data for the top 10 biggest world countries.**

***Index Terms*—Financial crisis, mutual debts, matrix games, Simplex method, equilibrium price index of the debts.**

I. Introduction

If we cast a glance at the world financial crisis started since the middle of the 2008, first of all it has resulted in increasing of mutual debts, problems arisen in financing of real sectors and consequently decreasing in value of debts especially in securities. Then it raises such question that in order to ensure the market equilibrium, provide with no debts and be repayable, at what level must the prices (exchange rates) be among the subjects, including, in case of current mutual debts of countries?

The economic nature of the problem depicts that prices are important in the formation of debts. Under the reciprocal relationship of demand and supply the economic system shapes prices and forces them to the equilibrium case. As a main tool of the "Invisible hand" prices play significant role. In the long-run the market equilibrium state occurs as a result the change of the level prices. In the short-run under the imperfect competition the monopoly deviates prices from the equilibrium level. Deviation of the prices from the equilibrium level causes emerging of the debts. Occurence of the debts creating multiplicative effect expands the amount of non-payments. The last global financial crisis enable us to observe the increasing of mutual debts, the problems arising in the financing of real sector and as a result we can see the searching by the system its balance with reduction the value of debts, especially securities. If we generelise the consept and look at the relationship between goods and services then it is not difficult to conclude that the overlasting market equilbrium and non-occuring of the debts simultaneously is not a case. Hence, deviation of prices from the equilibrium level can be seen as a main factor of formation of bebts. In the long-run the "temptation" of the market to the equlibrium changes the price level. To put differently, in the overall economy it forces level of prices to the equilibrium case. This in turn, causes financial crisis. For example, the monopoly in the market causes inefficiency in other sectors. This consequently causes to non-payment, with increase in expenditures for purchased goods and not repaying of credits. Under the current circumstance of mutual debts in which level the prices of appropriate goods of countries should be, in order to reach the market equilibrium without causing debts?

The answer of this problem will be investigated below.

II. Economic-Mathematical Statement Of The Problem

Suppose that, n number of subjects have mutual debts with one to another (receivables and payables of the countries) are related as depicted in the Table I:

TABLE I. Recording The Receivables And Payables Of Subjects In Matrix Form

Subjects	1	2	...	j	...	n	Payables
1	0	x_{12}	...	x_{1j}	...	x_{1n}	b_1
2	x_{21}	0	.	x_{2j}	...	x_{2n}	b_2
...	…	…	...	…	...	…	…
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in}	b_i
...	…	…	...	…	...	…	…
n	x_{n1}	x_{n2}	...	x_{nj}	...	0	b_n
Receivables	a_1	a_2	...	a_j	...	a_n	DK

Here, x_{ij}- is a quantity of payables debt of i^{th} subject due to j^{th} subject. b_i–is a sum of payables debts of i^{th} subject.

$$b_i=\sum_{j=1}^{n} x_{ij} \text{ , } i=\overline{1,n} \quad (1)$$

a_j-is a sum of receivables debt of j-th subject.

$$a_j=\sum_{i=1}^{n} x_{ij} \text{ , } j=\overline{1,n} \quad (2)$$

DK—is a sum of all receivables and payables debts.

As it seems from the Table 1 the debts of the subjects with one to another ones that is to say $x_{ii}=0$, ($i=1, 2,....,n$) is accepted. If we take a community or organization as a subject, then x_{ii} become the inner system debts and the same debts may be taken in the table form again and there are no debts of subjects due to themselves. Theoretically it may result that in initial case there is no funds in the accounts of subjects, that is to say we accept $I_k=0$ (k=1, 2,..n). Otherwise we would use the same funds in repayments of debts and it would become new table. Let`s denote the debt matrix shown in Table 1 as $\|X_{ij}\|$ and note the simple but necessary features of this matrix [1]:

- Sum of all payables of subjects are equals to all their receivables debts. That is,

$$\sum_{i=1}^{n} b_i = \sum_{j=1}^{n} a_j \quad (3)$$

- Sum of the difference between the receivables and payables of subjects equals to zero, that is

$$\sum_{i\in I} (a_i-b_i) =0 \quad (4)$$

Here, I={1, 2, ..., n}-is a set of index of subjects.

Let p_i represents the value level of the debts of i-th subject (country). Then the value of the receivables debts which j-th subject takes from 1-th subject equals to $a_{1j}p_1$ from 2-th subject $a_{2j}p_2$ and at last, from the n–th subject $a_{nj}p_n$. In order to eliminate the debts (payables debts) of j-th subject due to any another subjects the sum of receivables debts must be equal to the sum of payables debts:

$$x_{1j}p_1+x_{2j}p_2+\ldots+ x_{ij}p_i+\ldots +x_{nj}p_n\geq b_j \text{ ,}$$

$$j=1,2,...,n \quad (5)$$

The system of equations in Eq. 5 has n number of variables (p_i's) and n number of equations. The economic meaning of Eq. 5 is that the subjects having receivables (who will receive the fund) are interested in having such a price level (p_1, p_2 , …, p_i, …, p_n) for which the volume of receivables would not be less than (b_j) $p_j\geq 0$, j=1,2,...,n

Naturally, we are interested in minimizing the value of the debts (accounts payable).

$$f=\sum_{i=1}^{n} a_i*p_i \rightarrow \min \quad (6)$$

It seems that the problem in Eq. 5 and Eq.6 is a linear programming problem [2,3].

The economic sense of the problem in Eq. 5 and Eq. 6 is that the value of debts of the subjects (receivables debts) should be so minimal that subjects are able to settle its payables debts at least.

Let`s write the dual (adjoin) problem of Eq. 5 and Eq. 6 in matrix-vector form,

$$x_{i1}q_1+\ldots+ x_{in}q_n\leq a_i \text{ , } i=1,2,...,n \quad (7)$$

The economic interpretation of the system of the Eq.7 is that the subjects of payables are interested in finding that change in currency exchange rate in such level (q_1,q_2,…q_n) so that the volume of funds payable does not exceed the amount of receivables (a_i-receivables). $q_i\geq 0$, i=1,2,...,n.

In the same time subjects are interested in maximizing of receivables.

$$g=\sum_{j=1}^{n} b_j*q_j \rightarrow \max \quad (8)$$

If in an optimal solution of the problem of Eq. 5 and Eq. 6 $p^*=(p_1^*,p_2^*,\ldots,p_n^*)=(1,1,\ldots,1)$, that is $p_i^*=1$, i=1,2,…,n, this means that the shadow and objectively stipulated values (actual value) of debts equal to its nominal value an economy is in a balanced state and debts market is in equilibrium.

$$f_{min}=a_1p_1^*+a_2p_2^*+\ldots+ a_ip_i^*+\ldots +a_np_n^*=\sum_{i=1}^{n} a_i=\sum_{i=1}^{n} b_i$$

Obviously, in this case,

$$\frac{f_{min}}{\sum b_i}=1 \quad (9)$$

This case takes place if $a_i=b_i$.

Example: If we get $p_i^* = 0$ as an optimal solution, this means that the value of debts of i-th subject(country) is zero. To put it differently, due to the fact that i-th country sels it's goods over the market price it meaningless in the pure competition and should not get debit from other countries. The case $0<p_i^*<1$, means that the i-th subject has sold its products at higer prices than the market equilibrium level owed other countries in such a level so that the market values of debts (receivables) to be settled to i-th country is smaller than their nominal values.

If $p_i^*>1$ then this means that i-th country has sold its products at the lower price than the market equlibrium level. Hence, the shadow and market prices of debts (receivables) to be settled to i-th country are higher than nominal prices.

If

$$f_{min}=a_1p_1^*+a_2p_2^*+\ldots+ a_ip_i^*+\ldots +a_np_n^*<\sum_{i=1}^{n} b_i,$$

Then it means that the shadow and market prices of debits dont't allow accounts to be paid off. If we consider that the economy always "tries" to reach the equilibrium level, then we can conclude that the level of prices (inflation or devaluation) are supposed to be increased in order payables have been fallen and paid off.

Obviously, in this case

$$\frac{f_{min}}{\sum b_i}<1 \quad (10)$$

If

$$f_{min}=a_1p_1^*+a_2p_2^*+\ldots+a_ip_i^*+\ldots+a_np_n^*>\sum_{i=1}^{n} b_i,$$

Then this means that the shadow and market prices of debts allow payables not only to be fully paid off, it also creates some extra money.

In this case,

$$\frac{f_{min}}{\sum b_i}>1 \quad (11)$$

Now, money is entering to the bank and financial sectors. As it known banks create money with multiplicative effect. Then, riscs increase because real sector do not satisfy banks credit supply. Risky credits do not turn to the bank sector for a wile. This in turn, creates problem in financing of the real sector and causes to financial crisis. To put differently, we observe considerable fall in level of prices. It worth to note that the last 2008 financial crisis caused by the money excess in USA banks. The money surpluses in the banks induced risky mortgage credits which could not be paid-off later on. This, in turn, caused the problem in the financing of the real sector and caused financial crisis.The prices of oil and other goods was considerably fallen. The market price of payables are increased and the currency reserves for their repayments started to decrease.

The case $\frac{f_{min}}{\sum b_i}=1$, means that the existence of debts doesn't cause any threat for crisis. The financial market is in equlibrium state. If $\frac{f_{min}}{\sum b_i}<1$, then this means that the equlibrium case has destroyed and only increase of price level might returnt it. Again, if we consider that the economy always strives to equilibrium, it is logical to expect that the level of prices should be increased (inflation or devaluation) in order for payables have fallen and to be repayed. That is the real sector crisis is anticipated. If $\frac{f_{min}}{\sum b_i}>1$, this means that the equilibrium state has destroyed. Consider that the system always tends to the equilibrium then the falling of prices can cause equilibrium. That is, the equilibrium state necessiates financial crisis.

CI defined as below can be characterised as crisis index

$$CI=\frac{f_{min}}{\sum b_i} \quad (12)$$

III. Optimal Solution Of The Problem And Its Economic Sense

It worths to note that, the model for the equilibrium state of debit and credit debts in the case of Azerbaijan economy based on the mutual debts of firms, namely non-payments, has been investigated [4]. Information of mutual external debts for the year of 2011 for the 10 biggest developed counters (USA, Japan and European countries) given as Table 3. Using the data given in the Table III, Eq. 5 and Eq. 6-the problem of linear programming, has been solved by Simplex method and obtained results given in the Table II.

Computer realization of the model has been solved by LP Applied Software package (Look at Table V).

TABLE II. EQUILIBRIUM PRICES OF RECEIVABLES OF 10 DEVELOPED COUNTRIES IN 2011. RESULTS OF THE SOLUTION OF EQ 5 AND EQ. 6

Countries	Name	Variables	Price
1	USA	p_1	6
2	France	p_2	0
3	Spain	p_3	0
4	Portugal	p_4	0
5	Italy	p_5	3
6	Ireland	p_6	0
7	Greece	p_7	0
8	Japan	p_8	0
9	Germany	p_9	0
10	Britain	p_{10}	191
Function		(Min.)	15450

As it can be seen from the Table II, the value of the objective function of the minimum price (f=15450), namely the market value of the receivables of the countries is considerably greater than the current nominal value. Crisis index (CI = 15450 /7886.15=1.959) is greater than one.

Now let's solve the dual problem of Eq.5 and Eq.6.Eq. 7 and Eq. 8 simple is a dual linear programming problem of Eq.5 and Eq.6. The dual LP problem has been solved and the results are given in Table III. Computer realization of the model has been solved by LP Applied Software package (Look at Table VI).

As shown in Table III the value of the objective function for Eq. 7 and Eq. 8 is slightly different than that is for Eq.5 and Eq.6. The computer calculations about stemmed from a large number of computations. The value of the Crisis Index is greater than 1, which is equal to 1.76.

Thus, the solution to both issues have shown that the formation of the balance of payments (whether payment of payables, receivables, as well as to take back) the level of prices, including changes in exchange rates is necessary. This case must be take placed in the United States, Great Britain, Italy and Germany

TABLE III. EQUILIBRIUM PRICES OF CREDITOR DEBTS IN 10 MAJOR DEVELOPED COUNTRIES IN 2011. RESULTS OF THE SOLUTION OF EQ.7 AND EQ.8.

Countries	Names	Variables	Price
1	USA	p_1	2
2	France	p_2	0
3	Spain	p_3	0
4	Portugal	p_4	0
5	Italia	p_5	0
6	Ireland	p_6	0
7	Greece	p_7	1
8	Japan	p_8	0
9	Germany	p_9	3
10	Britain	p_{10}	87
Function		(max.)	13866

TABLE IV. MATRIX OF EXTERNAL DEBTS OF SOME DEVELOPED COUNTRIES IN 2011 (TABLE WAS CALCULATED BY AUTHOR) [5].

Countries		1 USA	2 France	3 Spain	4 Portugal	5 Italy	6 Ireland	7 Greece	8 Japan	9 Germany	10 Britain	Total
1	USA	0	831	413	438	830	170	0	0	0	0.15	2682.15
2	France	242	0	42	107	101	0	0	0	0	0	492
3	Spain	171	108	0	205	140	0	202	0	0	0	826
4	Portugal	201	79	165	0	225	0	38	0	0	0	708
5	Italy	576	122	377	209	0	315	0	0	113	0	1712
6	Ireland	49	20	131	111	75	0	22	20	0	0	428
7	Greece	30	33	120	307	54	29	0	0	0	0	573
8	Japan	4	0	27	19	65	3	0	0	0	0	118
9	Germany	40	15	82	24	104	0	0	0	0	0	265
10	Britain	6	0	16	41	9	0	3	7	0	0	82
Total		1319	1208	1373	1461	1603	517	265	27	113	0.15	7886.15

TABLE V. THE RESULTS OF THE COMPUTER REALIZATION OF EQ. 5 AND EQ. 6 LINEAR PROGRAMMING MODEL BASED ON INFORMATION FOR EXTERNAL DEBTS MATRIX FOR THE 10 BIGGEST DEVELOPED COUNTRIES

08-01-2016 18:55:30	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution	Reduced Cost	Basis Status
1	X1	6,0000	1 319,0000	7 914,0000	1 319,0000	at bound
2	X2	0	1 208,0000	0	1 208,0000	at bound
3	X3	0	1 373,0000	0	363,0000	at bound
4	X4	0	1 461,0000	0	1 271,0000	at bound
5	X5	3,0000	1 557,0000	4 671,0000	1 557,0000	at bound
6	X6	0	579,0000	0	469,0000	at bound
7	X7	0	268,0000	0	268,0000	at bound
8	X8	0	27,0000	0	27,0000	at bound
9	X9	0	113,0000	0	113,0000	at bound
10	X10	191,0000	15,0000	2 865,0000	0	basic
	Objective	Function	(Min.) =	15 450,0000		

TABLE VI. THE RESULTS OF THE COMPUTER REALIZATION OF EQ. 7 AND EQ. 8 LINEAR PROGRAMMING MODEL BASED ON INFORMATION FOR EXTERNAL DEBTS MATRIX FOR THE 10 BIGGEST DEVELOPED COUNTRIES

08-01-2016 18:54:33	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution	Reduced Cost	Basis Status
1	X1	2,0000	2 682,0000	5 364,0000	0	basic
2	X2	0	492,0000	0	492,0000	at bound
3	X3	0	826,0000	0	-2 230,0000	at bound
4	X4	0	708,0000	0	-7 123,0000	at bound
5	X5	0	1 712,0000	0	-7,0000	at bound
6	X6	0	428,0000	0	428,0000	at bound
7	X7	1,0000	573,0000	573,0000	0	basic
8	X8	0	137,0000	0	-1 200,0000	at bound
9	X9	3,0000	265,0000	795,0000	0	basic
10	X10	87,0000	82,0000	7 134,0000	0	basic
	Objective	Function	(Max.) =	13 866,0000		

IV. CONCLUSION

By standpoint of returning the price into equilibrium state the market price of securities should be gone down in the financial markets of USA, Spain, Italy and Great Britain. Note that, the problem has been realized on the information of 10 countries and external debts of these countries are considerably smaller than that of all countries. Therefore, obtained results may be beyond the reality. But estimated model theoretically may give contribution in forming the overall views on the way of solving the debt problem. On the other side, this model has a statistic characteristic. But in dynamic case the state of debts problem may be changed. This case has not been investigated. Note that significant change in current state of debts among the subjects occurs with severe financial crisis. One version of solving the problem goes through the financial-economic crisis which directs the market to the equilibrium.

ACKNOWLEDGMENT

The results of the study have been discussed at the scientific seminar in the Scientific Research Institute of Economic Studies and Kazakh`s University of economics, finance and international trade.

REFERENCES

[1] Hasanli Y.H. Algorithms for resolution of mutual arrears of enterprises // Az.AS News, A series on physics-mathematics, XIX edition, №3-4, Baku, "Elm", 1999, pp.119-124.

[2] David Gele, The Theory of Linear economic Models, McGraw-Hill, Book Company, Inc., New-York, Toronto, London, 1960.

[3] Ahmet Ozturk. Yoneylem Araştırması, Genişletilmiş 11. Basım, EKİN, 2007, 877s.

[4] Yadulla Hasanli, Equilibrium price model for mutual debts, ECO2014 2nd International Conference on Energy, Regional Integration and Socio-Economic Development, Baku, Azerbaijan, October 1-3, 2014, pp.42

[5] BBC Russkaya slujba, 'Kto, komu I skolko doljen v evrozone', 2011. [Online]. Available: http://www.bbc.co.uk/russian/business/2011/11/111121_eurozone_debt_stats.shtml?print. [Accessed: 09 September 2016].

Model evaluation of an Innovative Capital

Arzu Huseynova
Head department the development of innovation
Ministry of Economy of the Respublic of Azerbaijan Institute for SRER,
Baku, Azerbaijan
elm@box.az

Ophelya Mazanova
Head programmer
UNEC,
Baku, Azerbaijan
Ophelya2008@gmail.com

***Abstract*—The author analyses classification of the methods for evaluation of an enterprise's innovative potential. According to the author, the most effective model taking into account the uncertainty factor is the model based on the theory of fuzzy sets. The model has obvious advantages in comparison with the expert and statistical methods of evaluation, since it allows us to minimize the evaluation errors. The scientific-practical value of the results consists in a possibility of their application in a combination with the analysis of the official statistical data in the course of perfection of the state scientific and technical and innovative policy in the direction of a more intensive use of the scientific knowledge and achievements in the interests of modernization of the economy of Azerbaijan. The proposed approach can ensure an information integration of the** economic zone **of the scientific organizations and be used for a complex research of the industrial, innovative and economic-administrative processes within the framework of development of science**

***Index Terms*— Scientific organizations, innovations, scientific activity, evaluation criteria, innovative potential, expert evaluation methods**

I. INTRODUCTION

An effective use of the innovative potential makes possible a transition of an economic system into a qualitatively new state. Such a potential of the economic zone is transformed into a concrete form during an innovative process ensured by the economic zone activity.

One of the factors raising the scientific substantiation of the innovative activity management is evaluation of the innovative potential.

Studying and evaluation of the level and trends of development of the innovative potential in various sectors of the national innovative system allows us to single out a set of the factors and conditions necessary for a steady economic development of the economy as a whole.

Development of the techniques for evaluation of the innovative component in the new and developing sectors of the economy becomes more and more urgent. In practice, great attention is devoted to evaluation of innovations and innovative activity.

Among the existing techniques it is necessary to point out the technique of a uniform statistical investigation of the scientific research and development – Frascati Manual - (Organization of Economic Cooperation and Development - OECD), the method for evaluation of the index of the scientific-technical potential, as a component of an integrated indicator of the level of a country's competitiveness (experts of the World Economic Forum - WEF), method for evaluation of the development of the innovative activity of the European Union (EU), used by the experts of the Commission of the European Communities (CEC), methods of the national associations of automated trade, and various factor-indicative methods, which, as a rule, are based, on generalization of the statistical and analytical data, obtained from inspections of enterprises.

Integration of the estimated elements into a uniform integral indicator, as a rule, is done with the use of various mathematical probabilistic methods. In our opinion, it is possible to single out a number of common problems arising in their practical use. Among them is selection of a mathematical apparatus allowing to obtain trustworthy data and take strategic decisions in the conditions of uncertainty and insufficiency of the statistical data for the analysis of an innovative potential [3].

The usual sequence of actions in the analysis includes the following stages: problem statement; object analysis; selection of a method; elaboration process; analysis of the development results. From the point of view of the analysis of the means of evaluation, the most essential stages are selection of an evaluation method and the process of working out of an evaluation. New classifications for evaluation of an innovative potential continue to appear. The main reason behind this is a complexity of the economic zone of evaluation and indivisibility of the innovative potential into independent components. The boundaries between the components are fuzzy, and frequently it is difficult to find «a dividing line». Therefore, we believe, it would be correct to classify the methods, which are basic for evaluation of the systems of any complexity degree and acting as a basis for construction of the existing techniques.

II. OBJECTIVES

The purpose of the article is to assess scientific and technical or innovation potential in economic zone of Azerbaijan. The evaluation methods can be subdivided into classes by a number of signs concerning the specific features of the aim of an evaluation, of an investigated process and applied instruments. If we take the distinctions in the sources of the initial information as a classification basis, the evaluation methods can be divided into two classes - quantitative and qualitative ones. Division of the methods meets the basic requirements of the system analysis, consisting in a combination of formal and informal presentations, which is convenient for elaboration of the techniques and selection of methods for a gradual formalization of reflection and analysis of a situation The quantitative methods are based on a mathematical apparatus.

It is probably not realistic to have a deep knowledge of all the methods of modern mathematics, however, when selecting a method, it is important to understand the specific features of a direction and possibility of its use for evaluation of an innovative potential. Selection of an evaluation method to a great degree determines reliability of the obtained data and, hence, is a very important stage. An evaluation methodology is based on varied by their levels of scale and scientific validity methods, approaches and techniques for evaluation of an innovative potential. Thus, the methodology of research of difficult dynamic systems, to which social and economic systems belong, is rather rich and includes both the elementary methods, which do not use mathematical mechanisms, and the most complicated multifactorial computer modeling. It is obvious, that for carrying out of economic evaluations not all of the above methods are used, but only the ones, which are optimal from the point of view of accuracy and simplicity of realization and which take into account the character of the economic information.

The method of expert evaluations connected with gathering, systematization and processing of various kinds of evaluations and the statistical methods got a wide application in management of innovative activity. Such popularity of the methods is due to simplicity of their realization and minimal volume of the preparatory and auxiliary actions. In a number of cases application of the expert methods is the only possible way, if quantitative retrospective information is not available.

III. Statistical Method For Evaluation of An Innovative Potential

Such methods allow us without revealing of all the determined ties between the studied sequence of events or system elements reflected in a model, but on the basis of a selective observation to identify regularities and to extend them to the behavior of the system as a whole, to detect the character, force of mutual influence of the elements within a system structure and also of the environment components. The statistics elaborates a special methodology for studying and processing of materials: mass observations, method of groups, average values, indexes, balance method, method of graphical images, and other methods of analysis [1].

Statistics of numerical data is a basic method widely used in economic researches. The methods based on the numerical statistics have a number of drawbacks. Such inaccuracy in respect to the analysis of an innovative potential consists in impossibility to have statistical information during an indicator analysis, or in an insufficient volume of samples indicators. Obviously, for creation of an adequate and accurate model for evaluation of an innovative potential of economic zone the methods based on the numerical statistics and the methods of expert evaluations cannot be used in their pure form because of their serious drawbacks.

Therefore, usually the method of expert evaluations and the statistical method co-exist in an analysis of the economic indicators, which have digital presentations. In this case, drawbacks of one method are compensated for by the advantages of another. However, such a combined approach to evaluation of economic indicators also has its drawbacks. This is connected with the fact that an accuracy of evaluation of a probability of realization of an event depends on a number of factors, beginning from the quality of the statistical information and finishing with the expert evaluations: uncertainty is present in evaluation of this or that economic indicator.

IV. Multifactor Model of A Complex Evaluation of The Economic Zone Innovative Potential Based on The Theory of Fuzzy Sets

The above-stated order of adoption of a management decision can be presented in a form of a block-scheme for a factorial analysis of subjects. We will divide factors by n criterion. Elaboration of the system of balanced indicators for evaluation of the level of an innovative potential and determination of their interrelation within the framework of such a model was done with the use of the determined factorial analysis, and was logically predetermined by the essence of the innovative activity of the scientific-technological complex of the economic zones [5].

N criterion of factors (groups) (G) is singled out and a scale is developed for evaluation of every model's element, a correlation is done of the indicators' values with the corresponding values of the level of an innovative potential (G - G_{ij}), where i is a number of criteria i=1, n; j is a number of indicators j=1,m; (Table 1).

TABLE I. Factors For Evaluation Of The Innovative Potential Component Of The Scientific-Echnological Complex Of The Economic Zones

Numbers	Groups	Indexes	Indicators
G_1	Educational level	3	$G_{a1i}, i = \overline{1,3}$
G_2	Standard of well-being	2	$G_{a2i}, i = \overline{1,2}$
G_3	Level of infrastructure elements in a economic zone	1	$G_{a4i}, i = \overline{1,1}$
G_4	Level of economic development of a economic zone	2	$G_{a5i}, i = \overline{1,2}$

Factors for evaluation of the Innovative potential component of the scientific-technological complex of the economic zones The opinions found as a result of processing of the expert data were averaged out with the use of an arithmetic mean. Where G_i – is weight of the factor for *i*-expert, k – is the number of experts.

$$\overline{G_i} = \frac{\sum_{i=1}^{k} G_i}{k} \qquad (1)$$

V. RESULTS

On the basis of the examination of the state of the innovative potential and identification of its development problems it is possible to draw the following conclusions:

- Innovative potential of the economic zone should be understood as a system of interconnected resources, which determines real opportunities for realization of an innovative activity. Adoption of the strategic decisions based on an effective use of the innovative potential ensures additional competitive advantages for the economic zone.
- Dynamic properties of the innovative potential require, in the conditions of uncertainty, adoption of the decisions oriented on its development, which is especially important, in the scientific sphere, and demand a search for new methods of analysis and evaluation with the use of a modern mathematical apparatus.
- Diagnostics of the state of the scientific sphere and monitoring of its development have demonstrated that the major factors constraining the innovative development are a low level of innovative activity, unsatisfactory state of the technological base and unpreparedness of the personnel for an innovative activity.
- Analysis of the techniques applied for evaluation of an innovative potential has shown, that a considerable part of them leans on probabilistic methods, which demand sufficient statistical sample of data. Some of the techniques are based on the use of mainly expert evaluations. In practice, evaluation of an innovative potential of economic zone with application of such techniques often appears to be too complicated. In this connection, we should search for the methods allowing us to evaluate the innovative potential of economic zone in the conditions of uncertainty.
- The methods proposed in the work for a complex evaluation of the innovative potential of the economic zone on the basis of the theory of fuzzy sets meet the requirements for obtaining of reliable results in the conditions of uncertainty.
- The methods developed for a complex evaluation of an innovative potential allow us to apply them to different economic zone, and also to carry out monitoring of its level, which makes it possible to implement control over the economic zone activity and to improve their management system in order to ensure their effective innovative development.

ACKNOWLEDGEMENT

The research work was carried out with funds allocated according to “Expenses for economic reforms” paragraph of State Budget of the Republic of Azerbaijan for year 2015 confirmed by the number 290s order of Cabinet of Ministers of the Republic of Azerbaijan which was signed in 1st October 2015.

Topic of the article is based on the agreement signed between Ministry of Economy and Industry of the Republic of Azerbaijan and Institute for Scientific Research on Economic Reforms to ”Prepare recommendations and suggestions for increasing effectiveness of assessment of innovational activity and functionality of innovations in Azerbaijan”.

REFERENCES

[1] Gorbenko A. A. (2012) Systematic approach to analyzing the structure and nature of innovation capacity /A. A.Gorbenko / / Journal "Integral" . , 2012. Number 1 (0.5 ps).

[2] Goncharenko A.P. (2007) The role of the national system of scientific and technical information in the development of innovative economy in Ukraine / A.P.Goncharenko / / Problems of science. - 2007. - № 2. - S. 14-18.

[3] Zubov L.G., Andreeva O.N., Antropov O.A., Arzhanykh E.V. (2012) Public research organizations: the state, the dynamics and problems of development / / Information- analytical bulletin. - 2012. - № 4.

[4] Malyshev L.A., Shestakov I.V. (2012) Analysis of approaches to the evaluation of the innovation activity of Russian companies / / Herald PNIF. Socio- economic sciences, 2012, № 14 (38), p. 101-110.

[5] Nadzhafov Z.N., Gasimov F.H., Huseynova A.D. (2011) National Innovation Policy of the Republic of Azerbaijan / / magazine "Science and Innovation", September 2011, p. 36-41.

[6] Sanzhapov B.H., Kopylov A.V., Kopylov D.A. (2012) Classification methods to assess the innovation potential of the company / / The Gazette VolgGASU. Ser.: Build Informatics. ,2012. MY. 7 (21). Mode of access www.vestnik.vgasu.ru.

Implementing E-democracy in Russia: Institutional Resources and Modernization Risks

Liliia Pankratova
Department of Sociology of Culture and Communication
Saint Petersburg State University
Saint Petersburg, Russia
l.s.pankratova@spbu.ru

Sergey Pankratov
Department of Political Science
Volgograd State University
Volgograd, Russia
c.pankratov@mail.ru

***Abstract*— This article considers e-democracy as a socio-political resource and technology of the etatist model of Russian modernization. Based on the results of theoretical and empirical studies the article reveals the positive / negative factors affecting the legitimacy of public authorities and governance, Russian citizens' involvement in the political process. The analysis is focused on the risks and threats, institutional distortions that appear in the process of implementing e-democracy, as well as society and state security. The paper is based on the materials obtained during realization by the authors of the grant supported by the Russian Foundation for Humanities and Administration of the Volgograd region. The perspectives of functioning of e-democracy in the context of implementation of innovative communication technologies are defined.**

***Index Terms*— E-democracy, modernization, Russia, legitimacy of governance, technologies, risks, security.**

I. Introduction

The year 2016 marks 25 years since the collapse of the Soviet Union and the beginning of a new historical stage in the modernization of the Russian statehood. The following questions arose at that time for millions of Russians, "What is Russia today? Do I have a large homeland as a single whole, as a stable social community, and not permanently narrowing space?" The integration crisis has morphed into a crisis of fellow citizens' identity as moral and political subjects. At the same time the influence of global challenges increased, and that has called into question the possibility and necessity of the existence of the Russian Federation as a sovereign, independent state. In our opinion, the question of the Russian statehood is a crucial one for contemporary modernization process. From this perspective the efforts of Vladimir Putin to strengthen state power, as well as national and territorial integrity based on "dictatorship of law" should be analyzed.

A considerable part of citizens marks a phenomenon of dictatorship in the thesis of strong power and dictatorship of law. This contests legitimacy as a principle applicable for Russia's modernization. However, the majority of the supporters of this policy do not imply political dictatorship in the ordinary sense of the word, but something quite different: a strong state able to adjust the economic system, to protect the individual from arbitrariness and lawlessness, while maintaining political freedoms.

From our point of view, in this case it is necessary to raise the question of the possibility and necessity of the implementation of the etatist model of modernization that under current conditions should not mean a fundamental rejection of the ideas of democracy. We are talking about changing the pace of transformations, focusing on certain social groups, and understanding the nature of the existing political mechanisms. The etatist model of contemporary Russian society modernization is predetermined, on the one hand, by previous national history which produced the historical code of "strong state", and on the other hand, by the state "support of identity of a social whole in its movement from the past to the future."

II. Russian thinkers about democracy and "strong power"

We agree with the position of Ivan Ilyin who pointed out that Russia as a national and political phenomenon was created by a strong state power. "Russian man is able to keep the order and build a state; he is able to keep the exemplary discipline, sacrificially serve and die for their country. But this ability peeps and fructify [...] when it is germinated, secured, and led by strong and decent state authority" [1]. However, Ilyin warned that amongst a certain part of society there was formed "fanaticism of formal democracy", which turned the slogan of democratization of society in the "confession of faith", in the subject of "blind allegiance and oath" [2].

Pavel Novgorodtsev also pointed out that the implementation of the democratic idea is as difficult as easy it could be distorted. Naive politicians believe that it is necessary only to topple the old order and to proclaim freedom, universal right of suffrage, and the power of the people and then democracy will be realized of its own. But "the idea that with the destruction of the old foundations true freedom immediately cometh belongs to anarchist rather than democratic theory" [3]. Consequently, it is necessary to clearly understand that democracy is essentially self-government of the people. At the same time the people must be aware of their rights and respect rights of others, understand their duties and be capable of self-restricting. Pavel Novgorodtsev considered

democracy as a system of freedom and political relativism, which admit any political opportunity, any economic system, as long as it does not violate the principle of freedom. Democracy "is always a crossroad" [4].

In this article it is important to emphasize that according to Nikolai Berdyaev in Russia reception of democracy ideas is separated and disconnected from the idea of human and citizen rights. "Pathos of social equality always suppressed pathos of individual freedom. A person was not considered as a responsible creator of social life. New Life was expected as a result of changes in social environment, from external public not from creative personality, not from the spiritual rebirth of the people, its will, its mind" [5]. Therefore, now it is necessary to accept democracy in all its complexity - not as domination of the majority of the masses, but as the development of self-governing individual and the nation as a whole.

At the beginning of the last century the great Russian thinkers defined the most important methodological principles that could affect the nature of the current stage of political modernization: Russia needs a strong state power, however strong power is not the same as totalitarian power; it is important to distinguish between creative democracy from formal, mechanistic one; to take into account conditions that are required for implementation of the basic principles of democracy; not to consider democracy as a new form of religion that can painlessly solve urgent problems and contradictions.

III. E-GOVERNMENT IN THE CONTEXT OF RUSSIA'S MODERNIZATION

Consequently, the etatist model of modernization implies, on the one hand, the structural and functional aggiornamento of the Russian state as a political institution, on the other hand, elaboration of specific mechanisms of interactions between government and civil society, seeking for interests alignment, combination of many ways. The latter is the true meaning of democracy. The essence of this model is to create conditions for the formation of an open society, where the state is the political guarantee of the real alternative to underdevelopment of the country.

The adoption and implementation of "e-government" program is, firstly, a technological tool for solving administrative problems. Secondly, it is a method for creating a virtual space where issues on implementation of the national model of modernization transformations are discussed and decisions are made. Thirdly, it is a resource for democratization of relations between the society and the state. The structure of e-government contains three main component: e-democracy and e-participation, functional specificity of which involves the formation of public opinion, decision-making, and high level of the public sphere institutionalization by electronic means; electronic production networks providing cooperation between public, public and private institutions by electronic means; electronic public services designed for high-quality (with minimal time and other costs) provision of public (consumer) services through local, regional and national portals [6].

Implementation of e-government contributes to the development of e-participation as technology that engages different groups of people in public policy. Usage of information and communication technologies is crucial for improvement of effectiveness of interactions with the authorities. E-participation initially includes two-way communication, exchange of meaningful information between both - individuals and institutions of civil society, and with the authorities at various levels. As foreign and Russian practices of e-government use demonstrates there is an extension the scope of e-participation, its complexity and diversification, such as: community building, online consultations, lobbying, pre-election campaigning, online monitoring and discussions of social and regional issues, online mediation, online elections, etc.

We agree with Marina Leonova who identifies a number of organizational, methodological, and social reasons that determine the low level of e-participation in the Russian Federation. Among those reasons: the inadequate level of computer literacy of administrative staff that makes it difficult for bureaucracy to actively participate in electronic dialogue with the public and civil society institutions; insufficient attention of the heads of executive and legislative agencies to the use of e-participation, which is focused majorly on extension of the list of public services; imperfection of the instrumental (technical and software) tools used to organize and conduct e-participation; low level of trust on the part of citizens and public organizations to representatives of the authorities, that is not conducive to the expansion of the daily practices of interactivity; relatively low level of information and communication technologies usage skills among different population groups; lack of educational programs aimed at strengthening e-participation as a form of implementation of the principles and possibilities of e-democracy [7].

IV. E-DEMOCRACY IN THE SYSTEM OF SOCIAL AND POLITICAL GOVERNANCE

From our point of view, the use of e-democracy is an attempt within the etatist model of political modernization to keep up to the trend of social-political governance implementation that involves close cooperation, co-regulation, co-production and co-management between representatives of government and civil institutions at various levels. "Thus, management becomes bilateral, not unilateral any more, and even multilateral process. It changes not only the boundaries between the state and the society, but also the very nature of their interaction. It is already impossible to tell where one ends and another begins" [8].

Using the technology of e-democracy expand and complement the possibility of GR (government relations) as an institutionalized activity on building partnerships between civil society and the government. It includes collection and processing of information about government activities, preparation and dissemination of information about the

position of represented groups, influence on the process of political and administrative decisions (lobbying) [9].

V. E-DEMOCRACY IN THE SPACE OF THE INTERNET

When analyzing e-democracy in the space of the Internet, it is important to bear in mind that it serves as a citizen living territory structured based on the socio-cultural, political, legal, economic, demographic and other canons of the real world. As noted by a number of researchers there is "growing awareness that the use of the Internet technology is part of an ongoing social and political struggle within the complex power and social networks" [10]. The most popular mechanisms of e-democracy among Russian population are presented below at Fig. 1 [11].

The relation between space, place and the Internet most consistently explained in the Manuel Castells's theory of the network society. This virtual space he describes as the unity of the "space of flows" and "space of places". "The space of flows is not placeless, although its structural logic is. It is based on an electronic network, but this network links up specific places, with well-defined social, cultural, physical, and functional characteristics […] Both nodes and hubs are hierarchically organized according to their relative weight in the network. But this hierarchy may change depending upon the evolution of activities processed through the network" [12].

Daria Barinova rightly pointed that one of the Internet features is an ability to challenge existing institutions of power, which is successfully used by criminal groups and terrorists as well. The researcher states that in the information age, the role of individuals and social movements strengthens. Thus, the Internet technology is able to transform the social order in several areas, including the occurrence of isolated individuals' networks and the phenomenon of the perception of power as "flows of power in space" [13]. The information flow is central to the political structure and political behavior. Information is not only a tool and resource that is used by political actors in a strategic and psychological sense, but its performance and quality affect the identity of political actors [15].

VI. E-DEMOCRACY: RISK FOR THE SOCIETY AND THE STATE?

Consistent use of e-democracy and the expansion of the Internet as a socio-political communication raise the question of the necessity of representative democracy reproduction, if even now Internet technologies allow citizens to directly express their opinions and influence the legislative process.

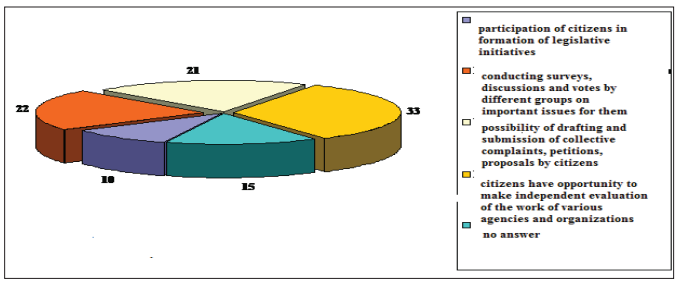


Fig. 1. The most popular mechanisms of e-democracy among Russian population (%)

In the political science community there is an active discussion about recently emerged conceptual models: direct digital democracy which involves the implementation of the direct will of citizens without the institutions of representative democracy; democracy of participation and joint action based on the organization of cooperation between the state and society on the formulation and adoption of joint solutions using internet and communication technologies; expert democracy with the involvement of self-selected experts for professional solutions to specific issues and problems that concern citizens.

In this regard it is significant to highlight the threats and risks for democratic system of governance. Cancellation or restriction of the scope of the activities of traditional institutions of representative democracy, particularly in the context of strengthening the new Russian state, is really possible outcome of these threats at the moment. These risks are:

- The lack of digital equality: different possibilities (territorial, settlement, etc.) for access to information and communication infrastructure.
- Significant inter-generational, professional and status gap in the development of competencies and skills in the use of the Internet communication technologies, which eliminates the possibility of equality in the articulation of their interests. Figure 2 presents the main obstacles of the spread of e-democracy in Russia (%) [11].
- Enhancing the use of tools of e-democracy by members of radical and extremist groups trying to destabilize the functioning of the political system, the violent change of political regime, etc.
- Not rare appear proposals to adopt populist management decisions not taking into account the multiple aspects of public and political management, quality of professionalism and competence approach to solving problematic situations.
- Progressing volume of information required for decision-making, making it difficult for ordinary citizens to fully participate in the socio-political process.
- Widespread of manipulation technologies of influence of the formation of public opinion, which is associated with the possibility of forming phantom democracy.
- Plurality of accessibility to information for online-users that creates illusory perceptions and beliefs about omniscience, rejecting the need for expert professional evaluations.
- Virtualization of expert status, the acquisition of which is linked with PR-technologies and marketing forms of promotion of "goods".
- Possibility of political preconception of an expert who acts not in the interests of society and the state, and pursue the interests of other actors in the socio-political process [16].
- Reduction of the state's ability to defend its sovereignty, national interests and to ensure the safety of social system functioning.

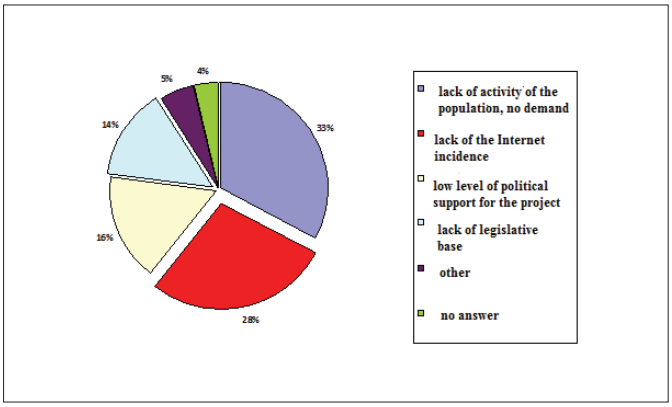


Fig. 2. Main obstacles to the spread of e-democracy in Russia

Thus, "the new technology can become a dangerous facilitator of tyranny […] There is not tyranny more dangerous than an invisible and benign tyranny, one in which subjects are complicit in their victimization, and in which enslavement is a product of circumstance rather than intention" [16].

VII. Research perspectives of e-democracy in Russia

Any risks and dangers that characterize innovative processes, especially in the political sphere, encourage researchers to optimize modern project practices of "democracy implementation". From our point of view, the least developed and therefore the most perspective aspect of analysis of the functioning and reproduction of e-democracy in the Russian Federation is its interrelation with the development of crowdsourcing, noosorsing (mind sourcing) technologies, and wiki-based collaboration (website which structure and content can be independently modified by users with the help of the tools provided by the site). As rightly pointed by some researchers "crowdsourcing technologies allow to legitimize the state policy by a new organized participation of citizens in the management that counteracts to the spread of political apathy and fatigue from democracy" [18]. Currently there is analytics that presents the generalization of such experience in the United States [19], but such political science research in Russia at federal and regional levels is only in the initial stage. In the framework of the study funded by the Russian Foundation for Humanities in the Volgograd region the authors will conduct further theoretical and practical conceptualization of e-democracy expansion with the help of crowdsourcing technological resources in the context of strengthening of the regional public space and increase of governmental authorities legitimacy.

VIII. Conclusion

It should be emphasized that social consciousness as the most important factor of development in all its forms, including political modernization, demonstrates the leading role of the public (civil) institutions in these processes. Forced modernization, whatever the circumstances caused it, divides society calling into question the achievement of the ultimate goal. In this regard you must consistently use the positive aspects of e-government and e-democracy as part of the etatist model of political modernization. The influence of the state on civil institutions is in the search for consensus and the harmonization of the most important aspects of the modernization process among its leading actors, as well as society as a whole.

Acknowledgment

This article was prepared within the framework of the study funded by the Russian Foundation for Humanities regional competition "Volga lands in the history and culture of Russia" project No15-13-34011 "The system of administrative and political management in the Volgograd region: the monitoring of the effectiveness, resource discovery and development of innovative technologies of modernization".

References

[1] I.A. Ilyin, "About strong power", in About the Future Russia. Moscow: Voenizdat, 1993, pp. 280.

[2] I.A. Ilyin, "Prerequisites of creative democracy", in About the Future Russia. Moscow: Voenizdat, 1993, pp. 29–35.

[3] P.I. Novgorodtsev, "Democracy at the Crossroads", in About social ideals. Moscow: Izdatel'stvo «Pressa», 1991, pp. 547–553.

[4] P.I. Novgorodtsev, About social ideals. Moscow: Izdatel'stvo «Pressa», 1991, pp. 553.

[5] N.A. Berdyaev, "Democracy and personality", in The fate of Russia. Moscow: Sovetskij pisatel', 1990, pp. 196.

[6] L.V. Smorgunov, "E-government, knowledge management and administrative reforms", in POLITEKS (Political Expertise), 2005, no. 2, pp. 213-225.

[7] M.V. Leonova, "Assessment and the development of e-participation in the Russian Federation", in Human. Community. Control, 2010, no. 4, pp. 57-68.

[8] A.V. Pavroz, "Government relations as an institution of social and political engagement", in POLITEKS (Political Expertise), 2005, no. 2, pp. 238-251.

[9] W.K. Agee, G.T. Cameron, P.H. Ault, and D.L. Wilcox, Public Relations: Strategies and Tactics. Allyn & Bacon, 2002.

[10] M. Zook, "The geographies of the Internet", in Annual review of information science and technology (ARIST), B. Cronin, Ed. Hoboken, NJ: John Wiley & Sons, 2006, Vol. 40, pp. 54.

[11] The results of the public opinion research in relation to the development of mechanism of e-government and e-democracy, WCIOM, Moscow, 2013, URL: http://wciom.ru/fileadmin/file/reports_conferences/2013/2013-05-28-elektrin-pravitelstvo.pdf (Accessed 12.06.2016).

[12] M. Castells, The rise of the network society. Malden, Mass: Blackwell publishers, 1996, pp. 443.

[13] D.S. Barinova, "Methodological aspects of the study of the virtual space of the Internet", in Method: Moscow Yearbook works of social science disciplines: Collection of scientific works. Vol. 1. Alternative models of the formation of nations. RAN INION, Moscow, 2010, pp. 109-122.

[14] L.A. Bershadskaya, "Institutionalization of socio-communicative structure of e-government in contemporary Russian society (example of St. Petersburg)". Author's abstract of PhD thesis, St. Petersburg State University, 2013.

[15] California Internet Voting Task Force: A Report on the Feasibility of Internet Voting. Sacramento, Office of California Secretary of State, 2000. URL:

http://www.umic.pt/images/stories/publicacoes1/final_report.pdf (Accessed 22.05.2016).

[16] S.V. Volodenkov, Internet communication in the global space of the modern political management. Moscow: Izd-vo Moskovskogo un-ta; Prospekt, 2015, pp. 272.

[17] B. Barber, "Three Scenarios for the Future of Technology and Strong Democracy", in Political Science Quarterly, 1998/1999, Vol.113, no. 4, pp. 581-582.

[18] I.N. Miroshnichenko, "Social Networks in the Russian public policy". Author's abstract of Doctor in Political Science thesis, Lomonosov Moscow State University, 2013.

[19] P. Peterson, "Government needs smart-sourcing, not crowdsourcin", March 24, 2009, URL: http://techpresident.com/blog-entry/government-needs-smart-sourcing-not-crowdsourcing (Accessed 12.06.2016)

Analyze of Results of Implementation of Fuzzy-Queries and Mathematical Structure of The Relational Database in Corporate Information Systems

Kenan Kilic Suleyman
ICT Centre Manager and Lecturer at the
Department of Computer Engineering
Qafqaz University
Baku, Azerbaijan
kkilic@qu.edu.az

Nuray Kilic Bilal
PhD Student of Qafqaz University
Qafqaz University
Baku, Azerbaijan
nuray.ermis@qu.edu.az

Tarlan Abdullayev Samad
Vice-Rector of Scientific Affairs
Odlar Yurdu University
Baku, Azerbaijan
oyu-asp@mail.ru

Anar Rustemov Murshud
Vise-Chair of ITS Department
Qafqaz University
Baku, Azerbaijan
e-mail:anrustemov@qu.edu.az

***Abstract*— Regular increase of database volume in corporate information systems (CIS) provides great opportunity to investigate relation among different type and find out hidden correlation. This is very important factor in order to develop strategic plans and dive correct decisions.**

In this article the implementation of Business Intelligence analytic information system to the reporting and query methodology to the CIS database were analyzed. Furthermore discovering of hidden information and there processing by fuzzy-logic were also investigated.

***Index Terms*—Corporate information systems, Fuzzy logic queries, Business intelligence.**

I. Introduction

Rapid economical development result in high demand in application of new information systems and their implementation. Therefore the companies with remotely allocated branches demanded unique, effective and integrated corporate information system database for easy management [1,2]. Three components are very essential in formation of any corporation: general purpose; information; and information sharing [3].

In order to continue operating an enterprise for a long time to get required data from information is an indispensable factor for organizational planning and control [4]. Additionally, in the arrangement stage of enterprise information systems employees often confront with stress and intensity [5]. Therefore, to provide the success of the company the most essential factor is that employees should not only know their work well, but also they should motivate on their work properly.

Increase in the number of user in CIS database leads to enormously expansion of volume of the database. This case create some barrier in decision making process because of big data usage. So effective usage and management of CIS database are still actual issues in modern business world. In this regards the main purpose is to identify hidden information and learn their affect on business requirement.

II. Analyze of existing methods and approaches

In early 90s there were two approaches in integration of different corporation information systems [6]:

- Enterprise Resource Planning, ERP;
- Data Warehouse [7,8].

ERP is based on the common transaction in the same system and Data Warehouse is based on the data integration according to the any requirement.

Furthermore Data Mining and Business Intelligence System (BIS) are also important fields in implementation in business sector. These area have some shortage in data selection and formation because of the duration of implementation process. BIS are mainly applied in macro economy, social marketing and medical database management and linguistic description forms. From other side these kind of systems information can be given not only in numerical form and also in different forms such as fuzzy logic and other type of description form.

Business information systems have two major problems:

1. Filter and search of data;

2. Management and sharing of data.

As we mentioned above none proper search of data has negative result in decision making. From this point of view fuzzy-logic, artificial intelligence and other type based filtering system has much effect in decision making and management process.

Although there are many researches in intellectual analyze of data, some problems are available in this direction. Considering the fact that these methods doesn't suitable for implantation fuzzy attributes. However we can find several researches in this area [9].

III. Problems and Solutions

Existing BIS doesn't cover all categorical, linguistic, artificial data and other type of illustrational version of data. Still there some un-investigated research questions [9]. Reporting and queries coming from the users in CIS database can be extended in terms of accuracy and effectiveness by implementing fuzzy-logic approach.

A. Formation of relational databases

The main object of the relational databases can be given in general form as follows:

$$P= \{O, K^o (A(x)^o,V^o), R=\{o_i \otimes o_j\} \} , o_i,o_j \in O \qquad (1)$$

,where O – represents the fields of objects; V^o – the state of the set of objects; $A(x)^o$ attributes of the objects; K^o – special attributes of the objects, such as state space and their combination, $\forall x \in X$ – the domain of the objects; R – represents the relation between object.

The queries in the relational database (Z) carry out some possibilities in some set of $C(x) = \{A(x)^o_i \otimes B(x)^o_j\}$ as a result of intersection, union and unification operation relyasiyon (join – $R=\{o_i \otimes o_j\}$) based on conditions of $K^o_i (A(x)^o_i,V^o_i) \oplus K^o_j (A(x)^o_j,V^o_j)$. These operations executed on the objects of (o_i, o_j).

B. General form of queries

$C(x) = \{A(x)^o_i \otimes B(x)^o_j\}$

$R=\{o_i \otimes o_j\}$ $K^o_i (A(x)^o_i,V^o_i) \oplus K^o_j (A(x)^o_j,V^o_j)$

Z = {Select <attributes> From o_i < join> o_j

Where <conditions> } (2)

, where $\otimes$ – all possible join (left outer join, right outer join , full outer join , cross join and etc) relations and $\oplus$ – represents "and", "or" and "no" relations in where clause of the query.

From other side the domains of the $\forall x \in X$ queries is limited within the ranges ($a \le x \le b$), where a and b are the lower and upper limit of the interval. Example given as below:

```
SELECT ASSORTIMENT.CODE KODU,
ASSORTIMENT.NAME ADI,
ASSORTIMENT.COUNTRY OLKE,
ASSORTIMENT.FIRMA FIRMA,
ASSORTIMENT.QALIQ QALIQ, ( CONVERT ( INT,
CONVERT (DATETIME, (RS.SONKULLANIM) ) ) -
CONVERT( INT, GETDATE())) / ((1.111/1.111)*30
SIT_AY,
SATIS. MIQDARI MIQDARI, SATIS.PRICEORG
PRICE
FROM APTEKMALLAR ASSORTIMENT
FULL OUTER JOIN R_BAZAS SATIS
ON ASSORTIMENT.MAL_LOGU = SATIS.MAL
LOGU
WHERE ( SROK<=3) AND ( (SATIS>=1 AND
SATIS<=10) AND (.PRICE>=50 ))
```

As an example, relational database tables of pharmacy information management system have been used in queries and they have been described as a subject area.

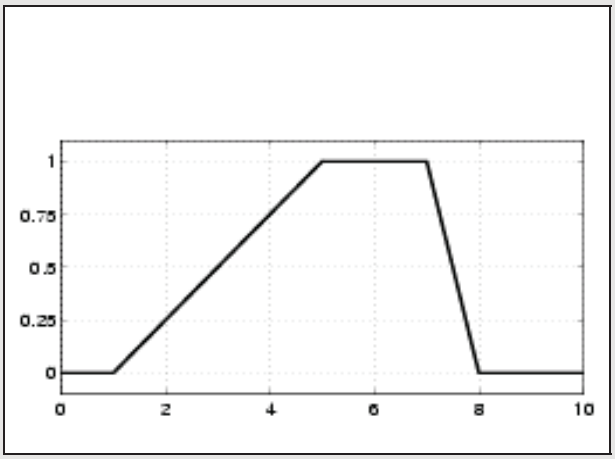
As can be seen from the above illustrated examples of SQL query, 1, 3, 10, 50 – are precisely defined border values . A relation function of fuzzy trapezium is shown in Table 1.

TABLE 1. A RELATED FUNCTİON OF TRAPEZİUM

TRAPEZIUM

$$f(x,a,b,c,d) = \begin{cases} 0, x \le a \\ \frac{x-a}{b-a}, a \le x \le b \\ 1, b \ge x \le c \\ \frac{d-x}{d-c}, c \le x \le d \\ 0, d \le x \end{cases}$$

a,d are bases of trapezium, b,c are legs



A related functions in SQL language have been realized in the form of special features of library. As an instance, trapezium relation function of SQL text has been given in Table 2.

TABLE 2. TRAPEZİUM RELATİON FUNCTİON OF SQL TEXT

```
ALTER FUNCTION [dbo].[Trapezium]
(
@TRAP_SeatA FLOAT,
@TRAP_SeatD FLOAT,
@TRAP_PeakB FLOAT,
@TRAP_PeakC FLOAT,
@X FLOAT
)
RETURNS FLOAT
AS
BEGIN
DECLARE @Y FLOAT
IF @X<=@TRAP_SeatA
BEGIN
SET @Y = 0
END
IF (@TRAP_SeatA<=@X) AND
(@TRAP_PeakB>=@X)
BEGIN
SET @Y = (@X-@TRAP_SeatA)/(@TRAP_PeakB-
@TRAP_OSeatA)
END
IF (@TRAP_PeakB<=@X) AND
(@TRAP_PeakC>=@X)
BEGIN
SET @Y = 1
END
IF (@TRAP_PeakC<=@X) AND
(@TRAP_SeatD>=@X)
BEGIN
SET @Y = (@TRAP_SeatD-@X)/(@TRAP_SeatD-
@TRAP_PeakC)
END
IF (@TRAP_SeatD<=@X)
BEGIN
SET @Y = 0
END
RETURN @Y
END
```

It is obvious that describing these functions in SQL language is very easy. Once again this proves the possibility of the use in linguistic variables.

If we think of the below given query as examples for fuzzy queries:

Query : Compile a list of the goods that have less expiration date.

Managers usually solve this type of fuzzy queries according to the form of understandable queries (2). Thus, if we consider that in drug sales firms on the 15th of each month lists of goods are prepared that have 3 months to the expiration date (or prior to 5 months) and drugs are refunded to vendors acoording to the same lists, in this case the discussed query can be described as a clear query as follows:

Select * from r_bases rs

where (convert(int,convert(datetime,(rs.expiration)))-

convert(int, getdate())) / ((1.111/1.111)*30 <=3

, here expiration expresses the expiration date of goods (SIT), getdate() expresses today's date i.e. report date.

So, if the manager prepared the list in getdate () date, as we mentioned above compiling a new report will be on the 15th of the next month. If there are 3.1 monthes (i.e. 3 months and one day) to the expiration date of product, in this case this product will not be in the manager's list and even if the same product is in the list of next month, seller will not refund that product. And as a result, if such kind of goods are not sold in time, they will be in the item disposal list and increase the cost of other goods of the same type and eventually, will reduce earnings of a company.

The solution of this type of problems can be solved on the basis of fuzzy queries. Thus, processing of fuzzy queries and their fulfilment in a relational database is a very actual issue. The following approach has been applied to he solution of the problem.

Taking into account the above mentioned information, $K^o(A(x)^o,V^o)$, a procession of relations at the same time should enable the fulfilment of fuzzy relationships and operations over relational database.

Fuzzy queries in comparison with clear queries (2) can be formalized as follows:

Select

< attributes > ,

<relation functions of fuzzy concepts >,

<indexes of belonging to a query>

From <fuzzy A plurality> join <fuzzy B plurality>

Where <conditions described with the fuzzy concepts (LV) > ⊕ < the terms of the exact interval > (3)

, join – expresses integration rules on pluralities.

Therefore, the condition of the query can be described as exact interval terms in form (2) (assortment number = 100, there are 5 months to the expiration data etc.) and at the same time as fuzzy conditions (these are called linguistic variables

(LV): for example, products which have less assortment number, products which have well performed sales quota etc.).

In general, the linguistic variables (LV) – are accepted as variables that get words and word combinations as a value and these values are illustrated as fuzzy pluralities.

For instance, the description of expiration date linguistic variable as fuzzy pluralities is shown in Fig.1 in a graphic way **Expiration Date = {little, good, much}.**

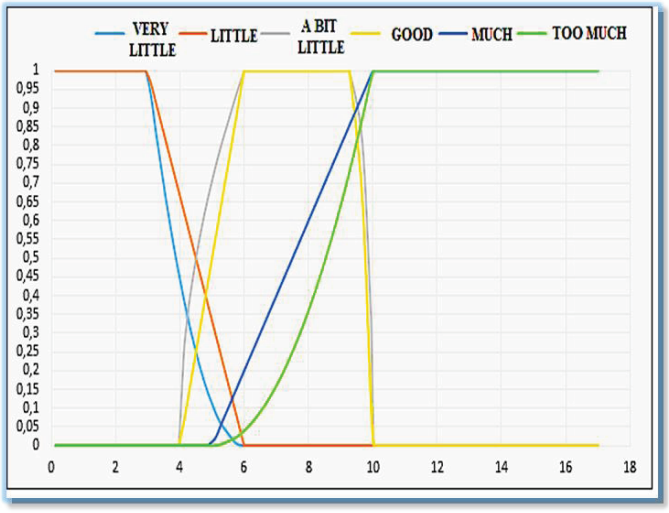


Fig.1. Graphic description of expiration date linguistic variable

IV. Conclusion

Solving the problem with this method will enable the solution of a wide spectrum of business challenges:

- Creation of efficient and comprehensive effective corporate reporting system for a short period of time as a result of fuzzy description of categorical and linguistic data,

- Non-regulated query processing – to these users: the list of young professionals whose salaries are not very high, the list of goods that has the highest sales figures and the cheapest price according to the regions and finding a more complete response to quiries of this type,

- In CIS database - solution of a different type of analysis issues of data,

- To study a different type of hidden connections and other regulations among the data,

- On the basis of fuzzy criteria, clusterization of products, customers, and other objects involved in business processes – the solution of these issues will enable to place goods grouped according to certain criteria in storehouses properly, selection of appropriate methods according to customer groups and materials management and the solution other data analysis issues,

- In order to compete in this market with other enterprises it is of great importance to fulfil necessary actions that will not allow employees to lose their motivation.

References

[1] The Trend of Development in Modern Enterprise Information Systems, http://www.management.com.ua/ims/ims010.html, URL Accessed date: 12 January 2016.

[2] Vavilov, K., S. Sherbina, Web-Integration, Electronic Journal "Open Systems", Vol. 1, 2001, p.12, http://ecm-journal.ru/post/Web-integracija-korporativnykh-sistem.aspx?from=more, URL Accessed date: 17 January 2016.

[3] Corporate Networks and Systems / Enterprise Network Concept, http://az.wikibooks.org/wiki/Korporativ_shəbəkələr_və_sistemlər/Korporativ_shəbəkə_anlayısi, URL Accessed date: 1 March 2016.

[4] Shahin A., Suleyman Demirel University, Institute of Social Sciences, Department of Public Administration, Management Information Systems, Master's Thesis, Isparta, 2006.

[5] Bayraktar E., Efe M., Enterprise Resource Planning (ERP) and Software Selection Process, Selchuk University, Social Sciences Institute Journal, Vol. 15, 2006: 689-709.

[6] Development and Implementation of Enterprise Portals, http://www.topsbi.ru/default.asp?trID=92, URL Accessed date: 15 Dekabr 2015.

[7] Alperovich M., Storage of Technology and Enterprise Data (Data Warehousing, OLAP, Data Mining), http://www.sft.ru/reviews/ DevConv97/DC2/DC2T12.htm, 2008, URL Accessed date: 26 April 2012.

[8] Inmon W.H., Building the Data Warehouse (Second Edition), New York: John Wiley & Sons, Inc.,1993.

[9] Abdullayev T.S, Alekperov R.B, Kilic K.S, Processing Of Fuzzy Logic Queries In Corporate Information System, Journals of Qafqaz University, Mathematics and Computer Science Number:34, Qafqaz University, 2012, p.105-113.

Management of Higher Education Institution Staff Activity Efficiency on the Basis of Flexible Stimulation System

Dmitriy Anufriev[1], Olga Shikulskaya[2]
[1]Rector, [2]Dept. of research activities
Astrakhan state university of civil engineering - ASUCE
Astrakhan,
Russian Federation
[1]ADP@aucu.ru, [2]shikul@mail.ru

Kanagat Dyusekeyev
Dept. of information technologies
Astrakhan state university - ASU
Astrakhan,
Russian Federation
abetovich@mail.ru

Mikhail Shikulskiy
Dept. of applied informatics in economy
Astrakhan State Technical University - ASTU
Astrakhan,
Russian Federation
shikul_m@mail.ru

Abstract — **Research object is the system of the differentiated compensation of higher education institution staff. The stimulation systems of higher education institution staff operating in Russia, Kazakhstan and other countries have been analysed by authors. The applied criteria of efficiency of higher education institutions in these countries are compared. Need of improvement of stimulation systems of higher education institution staff is proved. It is shown that the problem of staff qualification and efficiency of their work can be devided into two interconnected subtasks: definition of mechanism of the material stimulation of the higher education institution staff and formation of the flexible dynamic system of indicators adapted to the external conditions changing. For the first task solution the method of the system analysis was used, for the second task solution the methodology of the balanced system of indicators was used. As a result of the made research the strategic map, hierarchical system of criteria for evaluation of efficiency of activity of higher education institution staff are developed. Theoretical developments will be used in model of improvement of system of the differentiated compensation in higher education institution.**

Index Terms —**differentiated pay system, multiagent approach, strategy map, balanced scorecard, hierarchical system, criterion, objective tree**

I. Introduction

The number of higher education institutions around the world grows. At the same time education expands borders. Mobility of enrollees has increased. They aren't so connected to the place of residence as it was earlier. If not so long ago for many enrollees only existence of the diploma about the higher education mattered as guarantees of good employment, then now a number of prospective university students who is interested in quality of the got education is increasing. It strengthens the competition between higher education institutions [1].

Financing of higher education institutions is one of the most important factors of improvement of education quality. It provides a possibility of involvement of the competent teachers, creation of the extensive infrastructure of higher education institution including both modern material and technical resources for educational and scientific processes, and providing conditions for accommodation and leisure of students and the faculty.

In turn high rates of efficiency of higher education institution activity are a source of new budgetary and off-budget fund raising. And conditions of higher education institutions with the average and low level of these criteria values catastrophically worsens. Recognition of higher education institution inefficient in general leads to the termination of his financing by the state and closing. In this regard under the conditions intense competition the matter of rational distribution of the poor resources for the purpose of increase of efficiency of educational and scientific activity is taking up with administration of the majority of higher education institutions. Naturally, a key role in this process is played by human resources, and the system of material stimulation of the higher school employees. And than less financial resources belong to higher education institution, then the system of employees stimulation has to be especially effective and flexible.

A significant amount of works of domestic and foreign scientists is devoted to a problem of development of

differential system of compensation of higher education institution employees. However, the connecting of these systems with concrete indicators of quality of higher education institution activity which importance in modern realities quickly changes significantly reduces their efficiency. In this regard there is a need for creation of new effective models of stimulation of the higher education institution employees adapted to modern operating conditions of higher education institutions.

Research objective is development of model of system of the differentiated compensation of higher education institutions teachers on the basis of flexible system of criteria for evaluation of activity of teachers.

For achievement of this purpose it is necessary to solve the following problems:

- the analysis of system of compensation in higher education institution and the review of the existing methods of her improvement;
- analysis of indicators of an assessment of efficiency of activity of higher education institutions and their teachers;
- creation of model of the differentiated compensation of higher education institutions teachers;
- development of flexible hierarchical system of criteria as model component.

II. Research Methods

For the solution of the first two problems the method of the system analysis is used. According to the first task the existing systems of compensation of higher education institution employees in Russia, Kazakhstan, Great Britain, Germany, the Netherlands are analysed. Common features and specifics of various compensation systems are revealed, financing sources of the higher education are classified. The conclusion is drawn that the stimulation model needs to be developed taking into account the best foreign practices and specifics of domestic education [2-11].

For the second task solution the comparative analysis of systems of accreditation indicators of higher education institutions of the Russian Federation, the rating indicators of the higher education institutions considered by Independent Kazakhstan agency on ensuring quality in education (HKAOKO), rating indicators of higher education institutions, shown by the National accreditation center of the Ministry of Education and Science of the Republic of Kazakhstan, the indicators used by various foreign originators (Asia week, The Cente, CHE/Stern, Good Guides, The Guardian, Maclean's, Melbourne Index, The Times, U.S. News), indicators of various rating organizations ("Expert RA" - the Russian Federation, "QS-THES" - Great Britain, ARWU - China) is made [5-8]. Conclusions are drawn on discrepancy and instability of rating estimates, on lack of the accounting of a set of the existing indicators from various systems in the existing methods of stimulation of staff of higher education institution, on absence in these methods of the accounting of a possibility of higher education institutions of achievement of various indicators, on lack of an accurate binding of level of compensation to results of professional activity and on expediency of development of flexible system of criteria for evaluation of activity of teachers and on her basis of dynamic model of management of efficiency of work of teachers of the higher school.

The solution of the third task (development of model) is based on multiagent approach [11]. The management system of higher education institution activity is presented as the model of n–agent organizational system (OS). Each agent of system is the employee of higher education institution belonging to the faculty. The *i*-th agent of system can be characterized by the type (t_i) and by the action made by him $d_i \ge 0, i \in N = \{1,2,...,n\}$. Remuneration of the *i*-th agent (σ_i) has two components: tariff (t_i) and bonus (π_i). Bonus fund $P \geqslant 0$. The tariff component of a salary $tr(t_i)$ depends on the agent type which is defined by his qualification level and an experience. Besides the tariff component is influenced by the volume of the performed work (teaching loads). The bonus component of *i*-th agent depends on the actions made by him directed to increase of higher education institution efficiency: $\pi(d_i), i \in N$ where $d = (d_1, d_2, ..., d_n) \in D_+^n$ is a vector of agents' actions.

Thus, the general remuneration of the *i*-th agent has an appearance: $\sigma(t_i, d_i) = tr(t_i) + \pi(d_i)$.

$c(t_i, d_i)$ is cost function of the *i*-th agent.

The minimum value of criterion function of the agent which needs to be provided to him is restriction of reserve usefulness $u(\cdot)$, i.e. $u(t_i)$ is reserve usefulness of the *i*-th agent, $i \in N$.

$t = \{t_1, t_2, ..., t_n\}$ is the vector of agents types; $\pi(d) = \{\pi_1(d), \pi_2(d), ..., \pi_n(d)\}$ is vector-function of bonus stimulation

$$\sum_{i=1}^{n} \pi_i \le P$$

Criterion function of the agent:

$$f(t_i, tr(\cdot), d_i, \pi(\cdot)) = tr_i(t) + \pi_i(d) - c_i(t, d).$$

Criterion function of the center:

$$\Phi(t_i, tr(\cdot), d_i, \pi(\cdot)) = \sum_{i \in N} c_i(t, d) - \sum_{i \in N} tr_i(t) - \sum_{i \in N} \pi_i(d)$$

$RN(d_i, \pi(\cdot))$ is Nash equilibrium multitude of agent's game when the tariff and bonus system of stimulation is set;

$S(d, P) = \left\{ \pi(\cdot) \middle| \forall d \in RN(d, \pi(\cdot)) \middle| \sum_{i \in N} \pi_i(d) \le P \right\}$ is a set of bonus systems of stimulation, such that for any corresponding equilibrium vector of agents actions total bonus stimulation doesn't exceed bonus fund;

$U(t, tr(\cdot), \pi(\cdot)) = \{d \in D_+^n | tr_i(t) + \pi_i(d) - c(t, d) \ge u(t_i), i \in N\}$ — set of vectors of actions of agents at which values of their criterion functions satisfy to restrictions of reserve usefulness.

The guaranteed value of criterion function of the center on a set of agents decisions $K(t,tr(\cdot),\pi(\cdot)) = \min \Phi(t,tr(\cdot),\pi(\cdot))$ defines efficiency of tariff and bonus system:

$$K(t,tr(\cdot),\pi(\cdot)) = \min_{d \in RN(\pi(\cdot)) \cap U(t,tr(\cdot),\pi(\cdot))} \left[H(d) - \sum_{i \in N} \pi_i(d) - \sum_{i \in N} tr_i(t) \right]$$

where $H(d) = \sum_{i \in N} c_i(t,d)$ — sum of expenses of all agents of system .

The general problem definition of synthesis of optimum standard bonus system (SBS) has an appearance: $K(t,tr(\cdot),\pi(\cdot)) \to \max_{\pi(\cdot) \in S(t,P),tr(\cdot),P \geq 0}$, i.e. it is required to find optimum tariff payments $tr(\cdot)$, bonus fund P and rules of his distribution i.e. bonus system of stimulation (π_i) which would provide to all agents in balance reserve usefulness. The higher education institution can't significantly influence the amount of tariff payments, only indirectly through achievement of the required monitoring indicators, ensuring the intake of enrollees on the budgetary basis. Formation of bonus fund and definition of rules of its distribution are two big independent tasks. Within this research we will be limited only to a problem of distribution of bonus fund $\pi(\cdot)$ at the fixed tariff component. For the solution of this problem it is necessary to specify and concretize a problem definition.

The purpose of modeling of material stimulation system of scientific and pedagogical staff of higher education institution is increase in the higher education institution profit as a source of increase of efficiency of higher education institution activity. For creation of criteria system (the fourth task) it is necessary to construct a tree of the purposes. The methodology of the balanced system of indicators (BSI) is applied to the solution of this task.

III. Results of Research and Their Discussion

For creation of a tree of the purposes on the Balanced Scorecard methodology basis the authors have developed a strategic map of financing of higher education institution (fig. 1).

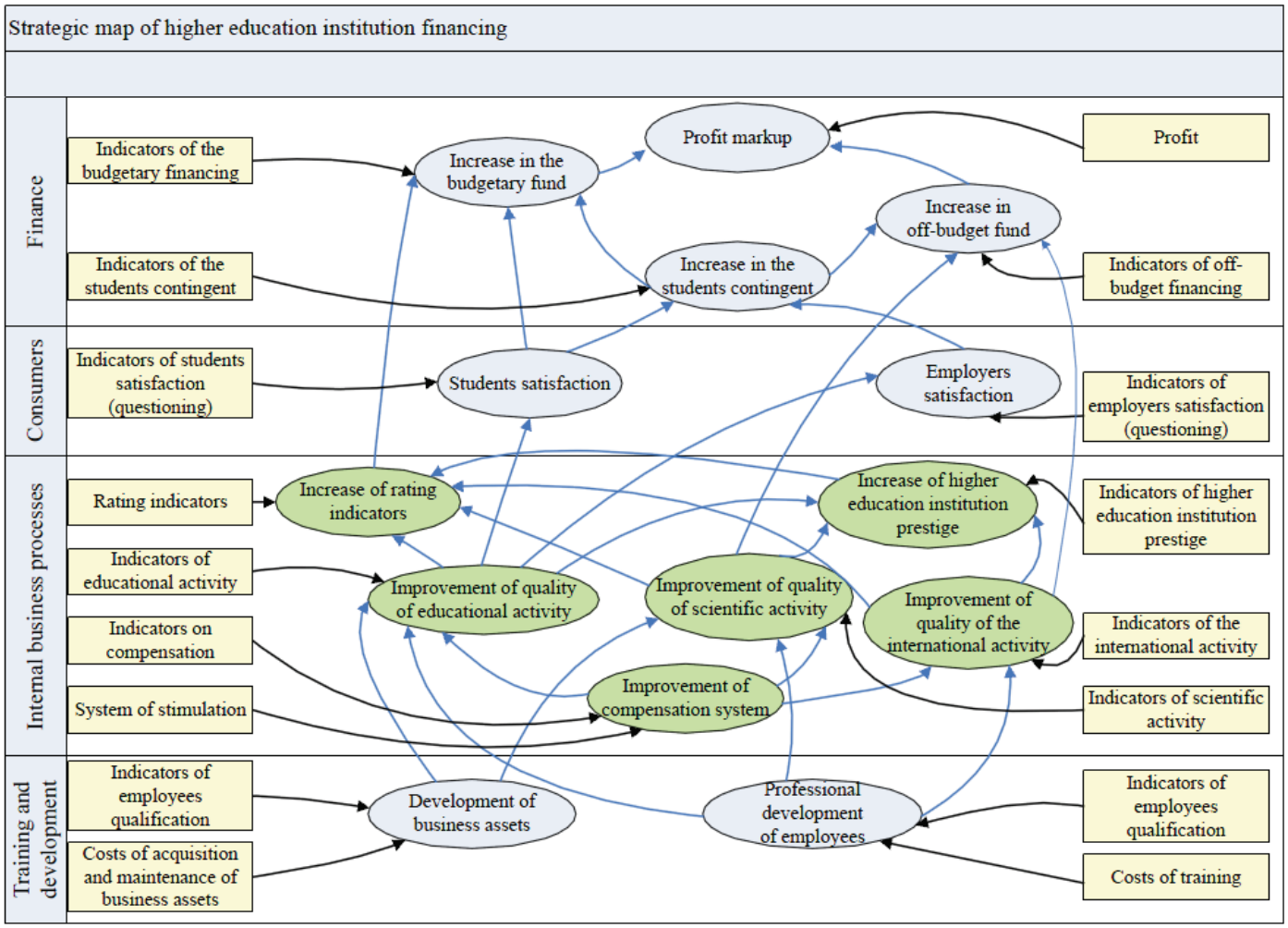


Fig. 1. Strategic map of financing of higher education institution

However, the system of indicators has to be dynamic, their standard values and trends have to change depending on a set of factors in the conditions of the competition. It is necessary to consider the importance of these criteria during various periods

of time. In this regard further improvement of indicators system is necessary.

Previously authors have analysed 44 indicators applied in various rating systems. Indicators are united in 4 groups on kinds of activity: educational, scientific, international and increase of image of university. However, the offered flexible system of criteria allows to increase quantity of indicators and to the number of groups which they treat.

For the accounting of the importance of the used indicators authors by results of the carried-out analysis have created hierarchical system of criteria. Distinctive feature of the offered system is the structured weight indicators (fig. 2). Such approach facilitates for the expert the process of an assessment of weight coefficients of criteria and increases assessment accuracy.

At each level of hierarchy, the corresponding weight coefficients on the basis of which then the generalized weight coefficients are determined by each criterion are set. Such approach allows to simplify work of experts and to increase the accuracy of their assessment at a significant amount of criteria of various type.

Amount of remuneration of the *i*-th employee is defined by the generalized indicator of quality of his work x_i which is characterized by his labor costs when performing actions by the set criteria and the importance of these criteria. The generalized indicator of quality is measured as a labor cost in the academical hours or in the astronomical ones.

Thus, the generalized indicator of quality of work of the *i*-th employee (teaching staff) will depend on performance by him of work by concrete criteria and the generalized weight coefficient of this criterion.

$$x_i = f\left(W\left(w_j\right),\ \ S\left(s_{ij}\right),\ \ K\left(k_{ij}\right)\right);\ \ i \in [1,n];\ \ j \in [1,m],$$

where k_{ij} is labor input of the performed operations by *j*-th criterion (in the academical hours or in the astronomical ones),

w_j is weight coefficient on *j*-mu to criterion,

s_{ij} is indicator of performance of action on *j*-th criterion of *i*-th the employee,

m is amount of criteria,

n is the number of the stimulated employees

$$s_{ij} = \begin{cases} 1, & \text{if meet the conditions} \quad P_j \\ 0, & \text{otherways} \end{cases}$$

The generalized weight coefficient of *j*-th criterion is equal to multiplication of the corresponding weight coefficients of each level *l*.

$$w_j = \prod_{l=1}^{3} w_{lj}$$

Labor input of the performed operations by *j*-th criterion is determined as product of standard labor input of performance of unity action by *j*-th criterion by the number of the unity operations on *j*-th criterion performed by *i*-th the employee

$$k_{ij} = k^0_{\ j} \cdot ke_{ij},$$

where k^0 is standard labor input of performance of unity action by *j*-th criterion (is defined by experts), $ke_{ij} \geq 0$ is the number of the unity operations by *j*-th criterion performed by the *i*-th employee (may be divisional).

The generalized indicator of quality of work of *i*-th employee will be determined by an equation:

$$x_i = \sum_{j=1}^{m} \left(w_j \cdot s_{ij} \cdot k_{ij}\right).$$

The price of unit of labor costs is calculated as the ratio of bonus fund to the sum of the generalized indicators of quality of all higher education institution staff.

$$cet = P \Big/ \sum_{i=1}^{n} x_i = P \Big/ \sum_{i=1}^{n} \sum_{j=1}^{m} \left(w_j \cdot s_{ij} \cdot k_{ij}\right)$$

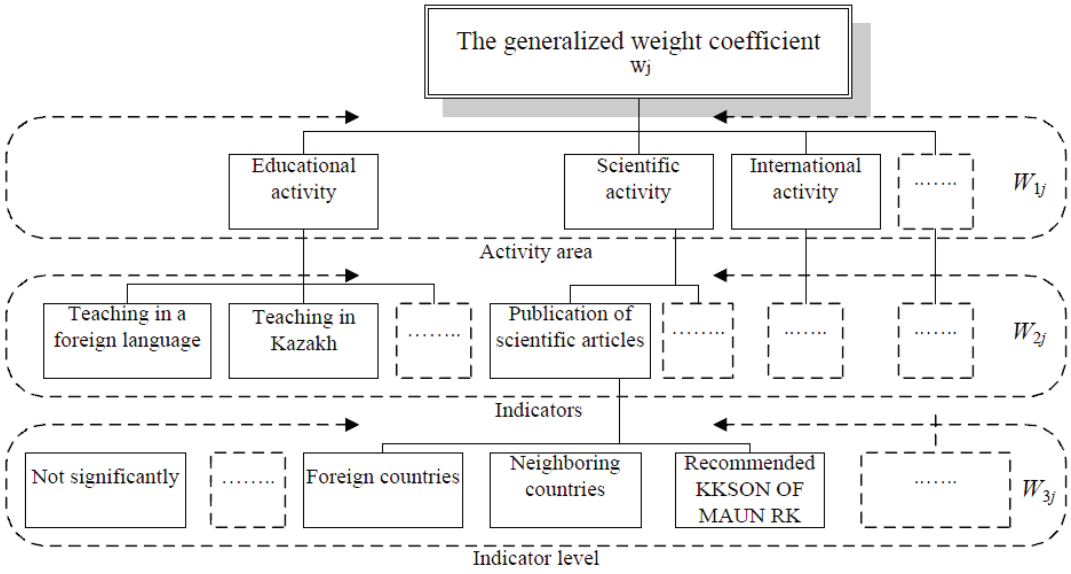


Fig. 2. Hierarchical system of criteria

Premium of *i*-th of the employee $\pi(d_i) = cet \cdot x_i$.

Settings of system is carried out by a variation of weight coefficients w_j.

We will input restrictions. Viability of system is defined by higher education institution performance of minimum admissible quantity of the accreditation indicators providing recognition of higher education institution by effective one:

$$KolVAR \geq KolVAR_{\min},$$

where $KolVAR$ is quantity of the executed accreditation indicators; $KolVAR_{\min}$ is minimum admissible quantity of the executed accreditation indicators necessary for recognition of higher education institution by effective one.

Therefore, when experts determinate weight coefficients values, they have to analyse a possibility of performance of this restrictions by higher education institution on the basis of results of self-inspection, reveal unachieved indicators, estimate an opportunity and cost intensity of their achievement on condition of material stimulation of employees, and according to it to put down weight coefficients.

To exclude imbalance of system due to the choice by employees of the most easily achievable for them indicators, it is possible to set restriction for the number of the paid unity operations on *j*-th criterion performed by one employee:

$$ke_{ij} \in \lfloor 0, ke_{j\max} \rfloor,$$

where $ke_{j\max}$ is the maximum number of the paid unity actions on *j*-th criterion.

The criteria system has to include the greatest possible amount of criteria. The exception of not actual in current time criteria is carried out by assignment of zero value to the corresponding weight coefficient.

IV. Conclusion

In the paper the need of improvement of stimulation system of higher education institution staff in the conditions of the competitive environment is proved, the analysis of the compensation system state in higher education institutions and the review of the existing methods of its improvement, the analysis of indicators of an assessment of efficiency of higher education institutions activity and their teachers in Russia, Kazakhstan and foreign countries is made. The model of the differentiated compensation system of higher education institutions teachers on the basis of flexible hierarchical criteria system is offered. Use in system of the greatest possible amount of criteria allows fast inclusion and exception necessary criteria depending on external conditions. Structure of weight coefficients considerably simplifies configuring of system.

Scientific novelty of research is the strategic map developed by authors, hierarchical criteria system for evaluation of activity efficiency of higher education institution staff with the structured weight indicators, model of the differentiated compensation of teachers of higher education institutions on the basis of flexible hierarchical system of criteria.

References

[1] Vishnevsky YU. R., “Professional choice of students and resource approach: from potential to realization”, the Bulletin of the Nizhny Novgorod university of N. I. Lobachevsky, series "Social sciences", [Professionalnyy vybor studentov i resursnyy podkhod: ot potentsiala k realizatsii // Vestnik Nizhegorodskogo universiteta im. N. I. Lobachevskogo. Ser. «Sotsialnyye nauki»], 2012, No. 1 (25), pp. 22-27.

[2] Dyusekeev K.A., Shikulskaya O. M. “The review of models of financing of state universities in various countries”, [Obzor modeley finansirovaniya gosudarstvennykh vuzov v razlichnykh stranakh], Materials of the X International scientific and practical conference, "Conduct of modern science", 2014, Volume 24, Technical sciences, Physical culture and sport, Sheffield, Science and education LTD - pp.29-31

[3] A.G. Kirillov, “Process approach in management of modern higher education institution”, Bulletin of the Chuvash state pedagogical university of I.Ya. Yakovlev, [Protsessnyy podkhod v upravlenii sovremennym vuzom, Vestnik Chuvashskogo gosudarstvennogo pedagogicheskogo universiteta im. I.Ya. Yakovleva], 2013, No. 1 (77), pp. 77 – 82.

[4] Kuzubov S. A., Ivlev A.V., “New system of compensation in higher education institution: state and Prospect”, Financial analytics: problems and prospects, [Novaya sistema oplaty truda v vuze: sostoyaniye i perspektivy. Finansovaya analitika: problemy i perspektivy], 2010, No. 8(32), pp. 31-35.

[5] Andrianov A.K., “Institutional and management structure as the factor of improving process efficiency of professional training of students in higher education institution”, Science and world (Vol.1, No. 5) , 2014, pp 259-261

[6] Richard Pircher, Attila Pausits. “Information and Knowledge Management at Higher Education Institutions”, Vol. 6 (2011), No. 2, pp. 008-016

[7] S. Sunitha, Malathy Duraisamy, “Measuring Efficiency of Technical Education Institutions in Kerala Using Data Envelopment Analysis”, Human Capital and Development, pp 129-145.

[8] Dyusekeev K.A., Shikulskaya O. M. “Analysis of systems of stimulation of employees of Higher education institution”, Technical science - from the theory to practice, [Analiz sistem stimulirovaniya sotrudnikov vuza. Tekhnicheskiye nauki - ot teorii k praktike], 2014, No. 40. p. 28-33.

[9] A. Lunev, I. Petrova V. Zaripova, “Competency-based models of learning for engineers: a comparison”, Journal: CEEE European Journal of Engineering Education Manuscript, Volume 38, Issue 5, October 2013, pp 543-555 DOI: 10.1080/03043797.2013.824410

[10] Zaripova V.M., Petrova I.Yu, “Engineering creativity education in russian universities”, Proceedings of the SEFI Annual Conference, 2014, Presented at: SEFI Annual Conference 2014 , 2014, Birmingham

[11] Tukubayev Z. B., Umarov A. A., “Model of quality management of education in Higher education institution”, Management of big systems. Release 37. M.: YIP RAHN, 2012. pp. 95-144

Management of Public Finances Through the Electronic Systems and It's Positive Sides

Tea Khorguashvili, Tamar Makasarashvili, Giuli Giguashvili
Gori State Teaching University, Professor
Gori, Georgia
tkhorguashvili@yahoo.com, tmakasarashvili@gmail.com, giuligi1@rambler.ru

Natia Khorguashvili
European Teaching University, Associate Professor
Head of Settlement Department of State Treasury
Tbilisi, Georgia
n.khorguashvili@mof.ge

***Abstract*—On the various stages of financial relations development with the state finances established and develop the local finances, corporations (firms) finances and modern household finances sphere provides financing of budgetary agencies and institutions in the form of disbursements of the employees (salaries, pensions, scholarships) in the budgetary institutions of household formation of financial resources, also financing of certain general state activities. System management of sound public finances enables the distribution of public finance effectively, which helps economic development and the eradication of poverty. Public finance management reform is closely integrated with the modern electronic services and IT technologies. Electronic systems and electrical management in many countries are growing very rapidly. Automation of exchange process of information between electronic systems of management public finance for the document processing caused a significant reduction of the time, costs, from both sides of personnel. State treasury offered convenient and transparent form of communication to the organizations. Through this system, disappeared paper documents from the circulation, it became possible to get information in real time mode, reduced operational risks minimize.**

***Index Terms*— public finances, integrated informational systems, automation of budgetary process, systems of treasury electronic service.**

I. Introduction

In any country, the most important field of the financial system presents state finances, which is committed to provide the appropriate authorities for the fulfillment of their functions with the necessary financial resources. In this field the subjects of monetary relations line up relevant state authorities, enterprises and organizations, citizens.

On the early stages mobilized monetary resources in the country were used for military purposes. It should be noted that on the first stage of the country's financial system consisted of only state budget, which had no influence on the country's economy. On the next stage of commodity-monetary relations development, with the establishment of various general state funds of monetary assets set up a financial system with its many subsystem, which actively participate in the distribution and redistribution process of national income and determines stable growth of economy.

On the various stages of financial relations development with the state finances established and develop the local finances, corporations (firms) finances and modern household finances sphere provides financing of budgetary agencies and institutions in the form of disbursements of the employees (salaries, pensions, scholarships) in the budgetary institutions of household formation of financial resources, also financing of certain general state activities.

II. Related Works

One of the main subsystem of the financial system is state (public) finances, which is the itself system and plays an important role in the development of the country's economy.

Financial system for such kind of countries as Georgia is developing concepts in Georgian practice, the sturcture of financial system used to change fundamentally during last 30 years, which is directly related to transformation of current market. The comparative stabilization of the economy, integration of Georgia in the world economy, amendments of the current financial legislation provided transformation of the financial system of Georgia and approximation with the international standards.

System management of sound public finances enables the distribution of public finance effectively, which helps economic development and the eradication of poverty.

The field of public finances includes the units of central and territorial budgets. The main goal of public finances is the formation of a centralized financial resources and the distribution for financing public needs.

Between the reforms implemented in the recent period in Georgia one of the important reform is public administration, which represents the horizontal and multi-sectoral policy and its implementation is depend on the executive authority decision-makers and high-ranking state officials involvement. The goal of public finances management reform represents provision of effective distribution of finance stability and public finance in accordance with government priorities. In Georgia the legal basis of public finances management represents. Budgetary Code of Georgia'', which regulates as state also autonomous republics and local self-government units budgetary processes. Mobilize of revenues regulate by the

Tax Code of Georgia'', state purchases by the rule of "Purchase Implementation''.

Public finance management reform is closely integrated with the modern electronic services and IT technologies. Electronic systems and electrical management in many countries are growing very rapidly. Georgia is trying not to fall behind the global tendency and according to electrical management it occupies one of the leading position in the countries of the region.

IT support of Public Financial Management is implemented by LEPL "Financial Analytical Service" of the Ministry of Finance of Georgia, the above service provides the development of information resources of the Ministry of Finance and other government agencies, business management and informational security. It is oriented for the creation and delivery of internal and external consumption of electronic products and services. Successfully functioning integrated electronic system of public finance management of the country- PFMS. [1] Within the mentioned reform was implemented electronic systems of budget management (eBudget), state treasury (eTreasury), public debt and investment projects management (eDMS), human resource management (eHRMS).

Electronic system of public finance management serves 4400 internal and half a million external customers. Through the system electronic transfer of all the budgetary salaries of organizations, pensions and other funds collecting of taxes and charges are implemented electronically.

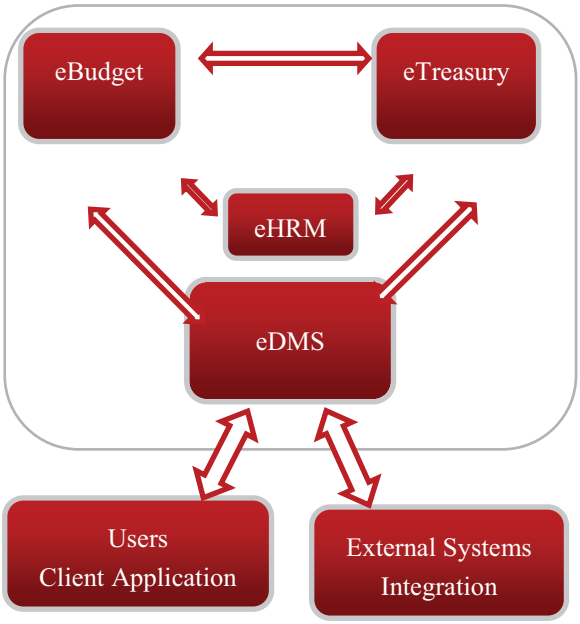


Fig. 1. PFMS architecture [5]

Implementation of electronic systems (eBudget) of budget management contributed to the public finances accounting maximally. Through the mentioned system Georgian central, autonomous republics and functioning of local management and with the performance of their duties acceptable revenues, rendered payments and management of changes in the balance of data are implemented. Through the mentioned system has been implemented automation of the budget process, which ensured the elaboration of consolidated budget management mechanism, capability of program budget management, system of dynamic reporting, and control mechanism of LEPL and N(N)LE budget implementation, integration functionality with other systems.

With the implementation of eBudget- state budget control mechanism became flexible, minimized reduced probability of mechanical errors from system user's.

One of the important component of reforms within the system of public finance management is implementation of state treasury electronic service system. Through the state treasury electronic service system are implemented accounting, reporting and calculation of revenues, payments and change of balance of state budget, budgets of the state, autonomous republics, local self-government units LEPL and N (N)LE. Until 2015 only budgetary organizations subordinated in the state budget and legal entities and non-profit (non-commercial) legal entities established by them were on the state treasury service. Within the reforms implemented in 2015 autonomous republics and budgets of local government moved on the state treasury services, but since 2016 all LEPL and N (N)LE established by local government, except 3,000 public schools involved in the system of state treasury electronic service, which involvement in the system of state treasury electronic service is planned from 2018.

In the last 10 years many fundamental reforms were implemented in the stated treasury. Since 2006 created a treasury single account (TSA), which covered all the budget revenue and payments of the state budget. Since 2012, occurred integration with the public finance management system and implemented web- system based on modern technologies, which is always available. This is an online service. Interfaces are flexible for users, simple to use and at the same time comply with legally established forms. Implementation of State Treasury Electronic Service System means not only technical tasks, but also train of consumers, signing the contracts with budgetary organizations, identification of consumer's rights and existence of one-time code generating devices. Processed and published the instructions. By 2012, about 670 budgetary organizations and 1300 consumers trained and involved in the system, and as a result of the reforms implemented in 2015 additionally involved approximately 700 organizations and 2000 new customer of local self-government budgets . Since 2016 in this system again added approximately 200 organizations LEPL and N (N)LE established by local government.

State treasury electronic system is tightly integrated with various systems: state budgeting and management system (eBudget), state debt and investment project management (eDMS) system, state electronic purchases system, revenue service (rs) [2] electronic systems, human resources management system (eHRMS), bank payment in real time mode payment (RTGS) [3] and worldwide interbank financial telecommunication SWIFT-'s system (society for worldwide interbank financial telecommunication). Through the mentioned systems constantly being data exchange. By using

of best practices and methodologies of IT process management made it possible to achieve business continuity. Electronic service system of state treasury consists of three levels: The spending side (Frontoffice), which gives the opportunity to the organization financed under the budget to manage their own funds. The intermediate layer (Middleoffice), which through the local self-government bodies have the capacity to manage revenues and payments of the district and treasury side (backoffice), which means the control for state treasury and funds, monitoring and instrument assembly required for the implementation of analysis. [4]

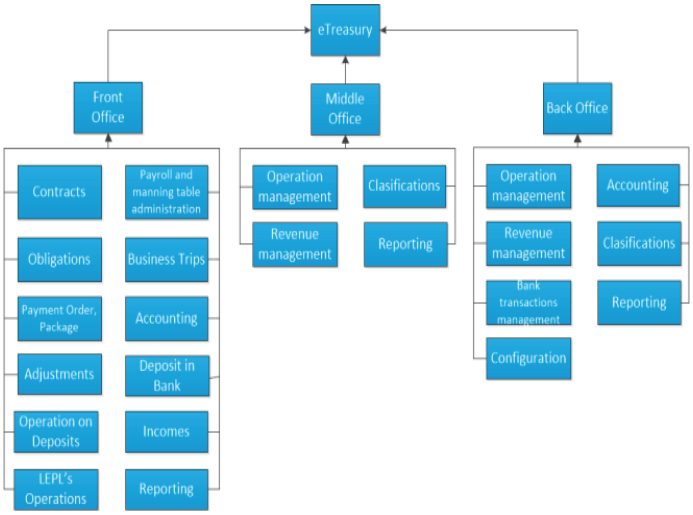


Fig. 2. eTreasuty system modules [5]

Electronic service system of state treasury became available and centralized as for all organization being on the services of the treasury, also it is possible to access from any point of the world.

Automation of exchange process of information between electronic systems of management public finance for the document processing caused a significant reduction of the time, costs, from both sides of personnel. State treasury offered convenient and transparent form of communication to the organizations. Through this system, disappeared paper documents from the circulation, it became possible to get information in real time mode, reduced operational risks minimize.

The union and gathering through the electronic systems in the common integrated system of public finances had strong support from the state government. The changes were made and improved many legislative acts. Developed and published textbooks. The state has developed a " Public Administration Reform Guide 2020". [4]

III. Conclusion

- Conducted reforms and implemented system provided complex, effective and transparent management of the state systems.
- Improvement of service, reduction of expenses, effective management of time and resources.
- In a short time provided reliable, accurate, complete and consistent collection of information about a financial event.
- Accounting data produced by the cash operations and accrual method in a common system is integrated.

The optimation is implemented and being on the main book of the treasury, which will eventually give us a complete picture of the state of public finances.

References

[1] Ministry of Finance of Georgia. [Online]. Available: http://www.mof.ge/en/

[2] Revenue Service. [Online]. Available: http://rs.ge/en/1340

[3] Ministry of Finance of Georgia State Treasury. [Online]. Available: http://treasury.ge/en/4971

[4] Public Administration Reform Roadmap 2020, Government Planning and Innovations Unit, May 25, 2015. [Online]. Available: http://gov.ge/files/425_49309_322150_15.07.21-PublicAdministrationReformRoadmap2020(Final)(1).pdf

[5] Public Expenditure Management Peer Assisted Learning, Europe and Central Asia Region, “PEMPAL TCOP thematic group meeting on Use of Information Technologies in Treasury Operations”, October 2015. [Online]. Available: https://www.pempal.org/events/pempal-tcop-thematic-group-meeting-use-information-technologies-treasury-operations

A critical analysis of Zambia's e-government adoption

Willard Munyoka
BIS Department, University of Venda,
PhD Student, University of KwaZulu-Natal
South Africa
willard.munyoka@univen.ac.za

Manoj Maharaj
School of Management, IT & Governance
University of KwaZulu-Natal
Durban, South Africa
maharajms@ukzn.ac.za

***Abstract*— E-government is the harnessing of information and communication technologies in government service delivery system aimed at enhancing government efficiency, effectiveness, accountability and empowering citizens. This paper examines the progress made in e-government adoption in Zambia using document study approach and the International Telecommunications Union's e-Government Implementation Toolkit as theoretical lens for deriving four themes for the inquiry: ICT infrastructure, policy, governance and outreach. In analyzing the progress made, comparisons are made to other SADC member-countries and world-leaders in e-government. Findings draw attention to key success areas as well as lacking areas requiring immediate attention for progressing e-government agendas in Zambia.**

***Keywords*— e-government, adoption, Zambia, SWOT analysis, e-government implementation toolkit.**

I. Introduction

Information and Communication Technologies (ICTs) underpins and facilitate high levels of business and service efficiency and transparency for both the private and public sector [1, 2]. E-government is an indispensable tool that could be used to rejuvenate a country's public service delivery system into a knowledge economy and society. It empowers both the decision making process by government officials and the general citizenry (service consumers) to actively engage and participate in government processes [3]. This is one of the major justifications for harnessing ICTs into the public-service delivery system in the domain of e-government.

E-government in Zambia took-off slowly at an institutional-level and steadily diffused to the public domain after some government interventions through enacting ICT policies [4, 5]. However, practical implementation by the government has been confronted with various challenges related to lack of skilled manpower in government to champion e-government projects, divided political will, unfavourable economic and political environment that discourage direct foreign investors into the ICT sector [6].

II. Methodology

This discussion paper adopts a document study approach [7] to critically analyse e-government adoption in Zambia. A document study approach closely analyse the investigated phenomenon using any relevant written material. Major sources of data used in this study include official documents covering government policy documents and gazettes; journal and conference publications; regional economic blocks, forums and international organizations reports; and non-governmental (NGO) reports. Additional primary data on Zambian's ICT development indicators and worldwide governance indicators were generated from [8] website for world, regional and national statistical data.

III. Data Analysis

Textual data for this inquiry is thematic analyzed. Thematic data analysis identifies and reports patterns in the studied aspects [9]. Themes capture essential patterns and meaningful phenomenon in the data to answer research question(s) for the study. In interrogating secondary data, the study aims to answer the question on what progress has been made in advancing e-government initiatives in Zambia.

In analyzing these documents, themes are derived from the data using inductive approach and from researchers' previous understanding of the investigated phenomenon, usually guided by theoretical underpinnings [10]. In the field of ICT for development (ICT4D), government efforts in e-government are designed along broader socio, political and economic frontiers to address societal needs. It is from this background that themes were discovered from textual data using the ITU e-government implementation toolkit [11].

IV. Thematic Analysis of e-Government Initiatives

The ITU e-government implementation toolkit [11] is a comprehensive framework for assessing a country's e-government readiness and adoption. The framework permits the researcher to formulate themes and do SWOT analysis along four fundamental dimensions of the e-government adoption: ICT infrastructure, policy, outreach and governance as explained in Table 1.

TABLE I. E-GOVERNMENT ADOPTION THEMES

Theme	Explanation	Locus
ICT Infrastructure	The technological infrastructure theme focuses on ICT readiness, ICT intensity, ICT impact and ICT capabilities.	Impacts/ Economic Development
Policy	The focus is on priority areas that regulate the e-government environment; which include rules to govern competition, trust and standardization; ICT policies and regulations; and government strategies on ICT and e-government.	
Governance	Focus on the traditions and institutions of exercising authority in a country. It covers process of selecting, monitoring and replacing governments; capacity of the governments to effectively formulate and implement sound policies; respect for citizens' rights and participation in national policy, economic and social issues.	
Outreach	Focus the status and efforts dedicated by the government and its agencies towards educating the general citizenry about public service delivery system (e-government).	

The locus constitutes the basis for clustering themes and in this study the major focus of the identified four themes is on sustainable economic and societal development of the e-government initiatives. Reference [12] recommends that e-government initiatives at the macro level should always be closely-linked with major globalization concerns regarding "citizen empowerment, gender empowerment, gender equity, achievement of universal access goals and more participatory governance."

V. ICT INFRASTRUCTURE THEME

To better understand and assess the status of ICT infrastructure in Zambia, this study uses the ICT Development Index's (IDI) four-stage model [13]:

- *ICT-Readiness* – investigate advances made in networked-infrastructure and citizenry's access to ICTs and e-government. Major infrastructure and access indicators are outlined in Table 1.
- *ICT-Intensity* – scrutinize the utilization of ICTs and e-government by the citizenry. An individual's capability to use e-government is directly determined by his or her skills and level of experience. Three major intensity pointers examined are persons using internet, fixed-broadband connectivity, and mobile-broadband registered users.
- *ICT-Impact* – examines the aftermaths of effective utilization of ICT and e-government by citizens.
- *ICT-Capability* – reveals the prevailing level of literacy amongst citizens essential for accepting and utilizing ICTs. Adult-literacy rate, gross-secondary-enrolment and gross-tertiary-enrolment are the three extensively used indicators. The impact of e-government is significantly affected by the prevailing levels of ICT skills and essentially, a requisite to leverage the full-potential for social-inclusion and economic growth.

Zambia's ICT development index is reported to have increased from 1.550 in 2010 to 2.040 in 2015 [8]. However, a score of 2.040 for Zambia falls far below both global-average of score of 5.03 and average-value for emerging economies of 4.12. Essentially, such low scores illustrate the existence of huge digital, access and e-service utilization divides. Just like in most developing countries, huge disparities exist in access to and utilization of e-services by citizens in Zambia and this is compounded by the fact that the level of affluence (as measured by basic infrastructure and income) drastically diminishes as one moves from urban to rural-remote areas [11]. Since access to and use of internet by individuals has strong bearing on e-government utilization, the Zambian government should pay greater attention to increase the percentage of households with internet access (6.9%), paralleled to 34.1% for developing countries and 46% for the world average [14]. Tariffs and affordability of e-services, specifically, internet in Zambia, just like across the SADC region are unaffordable to ordinary citizens and to promote e-government utilization, downward reviews are inevitable.

Zambia's mobile cellular penetration as measured by active mobile subscribers per 100 inhabitants is average – 67.3/100. With 17.3 out of 100 people using internet, Zambia legs behind many countries in e-government usage. Reference [15] posits that internet access in Zambia remains a privilege to the minority elite. With such low reported scores, Zambia is missing out to seamlessly-integrate its public-service-delivery system to become competitive in the global-economy. Internationally, with an average score of 3.71 on skills development sub-index, Zambia still trails behind Africa (4.20), developing nations (6.06) and the World (6.81) averages [8, 13].

Effective ICT and e-government operations require reliable power supply. The entire SADC region has been rocked by persistent power irregularities which threaten consistence e-government service provision. Long hours of power-cuts averaging 6-10 hours per day have been reported in the economic-hub province of Lusaka [16]; and government should do something to rescue this situation.

Despite facing huge challenges in the ICT infrastructure frontier, there are noteworthy community-based e-government projects in Zambia; mainly spearheaded by NGOs and international organizations. Notably, the Linknet ground-breaking project championed by Macha Works international organization in rural Macha community to deliver internet access to locals and assist community people to acquire ICT skills relevant to their local problems and using local language [17, 18, 19]. The organization utilizes shipping containers to install internet using satellite dishes and develop local ICT skills through training to ensure sustainability and knowledge exchange. The Zambia National Farmers Union (ZNFU) initiated an m-government service utilizing SMS to convey farm produce commodity prices, trader information and veterinary services to micro, small and medium-sized farmers across the country. Savings in transport costs and agriculture service broker charges are achieved through elimination of third parties in the agriculture-value chain.

Mwana m-health programme launched in 2010 by the Zambian Ministry of Health in conjunction with UNICEF offers 'early infant HIV test results via SMS' to parents across 40 piloted clinics. The Mwana m-health has successfully curtailed the turnaround time of receiving HIV results by half. Most of these e-government services are in the interactional and transactional phases of maturity model. The presence of digital divide and e-service utilization divide across Zambia is widespread, and it aggravates as one move away from metropolitan areas towards remote-rural communities [20].

VI. Policy Theme

The roadmap to e-government initiatives in Zambia dates back to 2001, when the national ICT policy drafting commenced in consultation with all major stakeholders focusing on thirteen core areas [21, 22]. The final ICT policy document was concluded in 2005 and then adopted in 2006. In essence, five years had elapsed since its inception, and by then technology had changed substantially - thus rendering the highly acclaimed ICT policy out-dated. The Zambian government through its Zambia Information and Communications Technology Authority (ZICTA) passed various acts and statutory instruments to amend the national ICT policy. These include The Information and Communication Technologies Act of 2009; The Electronic Communications and Transactions (ECT) Act of 2009 and Statutory Instrument no. 35, of 2010 and ICT Licensing Regulations of 2010.

Since the enactment of the ECT Act in 2009, the Zambian ICT sectors has witnessed booming investment in the mobile-voice market, broadband infrastructure, invitation of foreign direct investment into the state owned Telecoms Company by Zambia Telecommunications Company (ZAMTEL) in 2010, and increased access to and utilization of ICTs [23].

There has been a political-will and commitment to transform and rejuvenate Zambia's public service delivery system into a knowledge economy through e-government, as enacted in the National ICT policy. Nevertheless, reality checks on e-government progress made thus far, has fallen far-off expectation, as stated in the ICT policy document. Attention is drawn to the absence of a nationwide-ICT-implementation-plan for e-government [24] – thus nullifying the relevance of such ICT policies. Other cited regulatory issues requiring government intervention to encourage direct investment into the ICT sector include revising downwards licensing-fees, strengthen local-investment, lowering tariffs for broadband and internet by Zambian ICT Authority (ZICTA) to stimulate both access and utilization.

VII. Governance Theme

Governance embroils values, norms and practices by which governments are enacted and replaced through democratic processes by its citizens; its willingness to formulate citizen-centric policies and value human rights [25]. World Governance Indicators (WGI) is one of the renowned instruments for evaluating the soundness of an individual country's performance in governance using both quantitative and qualitative means. The WGI uses six main sub-indexes to measure governance: voice and accountability, political-stability and absence of violence, government efficacy, governing quality, rule-of-law and control-of-corruption.

Studies on governance [26, 27, 28] suggest that high socio-political instability characterizing sub-Saharan Africa and numerous developing nations world-over are indicators of poor governance. Moreover, widespread corruption and nepotism [24] have been voiced in the awarding of e-government and ICT project tenders as a major problem hindering the provision of high quality solutions that commands trust from the broader citizenry.

With a political stability and absence of violence governance indicator score of 0.2056, Zambia performs better than most SADC countries [8]. However, with low governance scores in the rule of law, control of corruption and government effectiveness, the Zambian government has a mammoth outreach assignment to convince citizens to trust public service delivery system and worse still e-government. Though such governance conditions prevail across most developing countries [28]; the Zambian government should institute and enforce stringent governance measures that encourage the growth and utilization of e-governance. Using Denmark as referral nation having achieved good governance (rule of law, control of corruption; censorship and accountability), Zambia should draw some lessons. Report on anti-corruption efforts [29] shows Zambia with endemic corruption of 1% (in health sector) and 29% (in police force) which is prevalent in metropolitan than rural areas.

VIII. Outreach Theme

E-government outreach concerns the status and efforts dedicated towards educating the general citizenry about public e-service delivery system. A candid account of outreach [30] maintains that e-government "services are only useful if people know about" their existence and are educated and motivated to use them. Successful e-government adoption requires governments to assess two fundamental aspects of outreach: the supply-activities and potential-demand; and service uptake and utilization by stakeholders [11].

The supply-side of e-government is measured by inspecting the complexity and functionalities of government portals. Citizen satisfaction with e-government services is one of the most well-known pointers of how well a government is satisfying the service-needs of its stakeholders. Reference [31] suggest that the failure by most governments in developing countries to ascertain citizen satisfaction with e-government services resembles the darkest-side of e-government adoption and often a causal-point of failure for most projects [37]. In the Zambian context, statistics on citizens' contentment with e-government services is unavailable in the public domain. Most studies done hitherto on e-government concentrate more on adoption [32, 24] and challenges [20] at the expense of access, utilization and satisfaction [33, 34]. Nevertheless, there are some reported cases of greater impact of small-community-based projects, often spearheaded by NGOs in rural areas of

Zambia [18]. The success of such projects is largely attributed to targeting community-based problems using language-tailored training to address ICT skills and service needs.

UN E-Government Survey's Web Measure and e-Participation indices [35] are the two widely used indicators for measuring e-government supply-outreach. Web Measure index assesses web-portal features of key government ministerial departments, in-line with their usability, completeness and how helpful they are in meeting the diverse needs of stakeholders (mainly citizens and business). Key targeted ministerial websites for assessment include the Ministries of Health, Finance, Education, Labour and Social Works [11].

In line with Web Measure index, the UN E-Government Survey [35] has identified eight core usage-enhancing website features: the 'contact us' information, 'search' facility, 'help' and 'frequently asked questions (FAQs)' sections, instructional sections to users on how to use essential features, site-map and index, advanced web-search and filter options, video contents; and current-information.

Critical analysis of Zambia's five ministerial department websites (Health, Finance, General Education, Labour and Social Security) shows some positive progress and scores as summarized in Table 2. In-depth scrutiny of the five e-government portals shows the Ministry of Health having the most usage-enhancing website feature; only falling short on advanced web-search. The rest of the ministerial departments have almost similar features and more effort is inevitably required to match high standards set by the Ministry of Health.

The E-Participation Supply Index provides essential information on the government's political-cultural consultative mechanisms with its citizenry. Internet access and utilization are a prerequisite to accessing e-government services by citizens. According to [36] Zambia's internet penetration stands at 18% of the total population and 80% of the population accessing internet are concentrated in the Lusaka Province alone. This entails that there are huge disparities in internet access and use across towns and provinces; and this worsens as one moves to rural communities were abject poverty has hit the hardest. Even for the few cosmopolitan elite, widespread technophobia cases has been reported [28] – thus, education and outreach programmes could be the solution for enhancing e-service access and utilization in Zambia.

Conversely, there are three decisive indicators for evaluating the demand-side of e-Government outreach, namely citizenry's level of education, citizens' access to ICTs and usage patterns. In all the three spheres of demand indicators Zambia remains subdued and with a mammoth task to turn around its fortunes. Internationally, with an average score of 3.71 on the skills sub-index, Zambia is trailing behind Africa (4.20), developing nations (6.06) and the World (6.81) averages [13]. Low levels of education have negative effect on citizens' awareness of e-government, capacity to use and frequency of e-service utilization. Moreover, an average score of 2.63 on citizens' access to ICTs sees Zambia trailing behind Africa (4.50) and world (6.93) average scores [13].

TABLE II. WEB MEASURE INDEX FOR ZAMBIA

Ministerial Department	Web Measure index (usage-enhancing website feature available?)						
	Contact Us	Search facility	Help & FAQs features	Advanced web-search	Site-map	Video	Current Information
Zambian e-Gov Portal	Yes	Yes	No	No	No	No	Yes
Finance	Yes	No	No	No	Yes	No	Yes
Health	Yes	Yes	Yes	No	Yes	Yes	Yes
General Education	Yes	No	No	No	Yes	No	Yes
Labour and Social Security	Yes	No	No	No	Yes	No	Yes

IX. SWOT ANALYSIS OF ZAMBIA'S E-GOVERNMENT

The SWOT analysis technique has been widely used by decision makers for ICT and e-government adoption strategic-planning. In the context of e-government, policy-makers should evaluate the local-context settings of e-government adoption in line with its strengths, weaknesses, opportunities and threats so as to increase chances of e-projects success and to identify failure points in organizational strategic planning [41]. Reference [37] pronounces that beneath the decorative-hi-tech fascination with e-government adoption lies disturbing realities of high project failure rates for most developing countries. Moreover, [37] underscores the need for public policy makers involved in e-government and ICT projects to understand 'why?' most e-government projects fail and 'how?' best to make e-government projects successful; and finally, use existing experiences in regional failure/success cases for spearheading effective e-government projects. Against this backdrop of reported high e-government failure, the SWOT analysis provides an instrument for mapping out from policy document to implementation as outlined below.

A. Strengths

Major milestone and strengths Zambia's e-government efforts lies in having sound ICT policy documents in place as a roadmap to attain various national strategies and visions. The determination by the Zambian government to enact the National ICT Policy in 2006, followed by the Information and Communication Technologies Act of 2009 indicate that ICTs constitute a pivotal catalyst for attaining national agendas and economic growth through enculturation of a twenty-first knowledge society.

Inroads in human-capacity development in ICT skills were logged in Zambia through orientation and teaching of ICTs starting from the grassroots levels (primary-to-secondary) up to tertiary institutions in which basic ICT-orientation courses are mandatory to all students. Furthermore, the liberalization of the

mobile cellular sector witnessed massive cell phone ownership by citizens; thus creating conducive environment for adopting m-government.

B. Weaknesses

Despite having enjoying some noticeable ICT and e-government progress, Zambia continue to risk suffering total or partial failure in e-government due to the absence of clear and practical implementation strategies. Against this backdrop, e-government adoption in Zambia faces severe challenges in transforming abstract statements in ICT policies into practical implementation – some trial-and-error executions are inevitable and have been observed. Such a gap-measure approach in adoption may yield results in the short run; but fail dismally in the long-run due to lack of clear implementation vision.

Weaker and often subdued economy is one of the foremost hindrances to e-government adoption in Zambia. Huge budget deficits meant that developmental and social projects bargain huge chunks of the budget at the expense of sectors like ICT and e-government.

Bad governance and stifling of citizen voices and freedom of expression has been reiterated for Zambia [3, 8]. Zambia, just like Zimbabwe, scores the lowest Governance Indicators out of the six conveniently selected countries, which include Botswana, South Africa, Seychelles and Denmark. Poor governance has dire-consequences on e-government acceptance and utilization by citizens who may regard themselves as outsiders to public policy making and e-government projects due to exclusion in decision making process.

Digital divide and e-service utilization divide remains a reality in rural remote areas of Zambia where more than 70% of the citizens resides [39, 40]. Low ICT literacy rate and massive unemployment mean the general citizenry could not enjoy the benefits of e-government initiatives due to adjective poverty.

There are subdued e-government awareness campaigns and outreach programmes in Zambia to educate the citizens on the benefits of using e-government services and how to use it. This could explain why some affluent citizens continue to shun emerging digital devices and online services like e-banking, e-government and m-government. The Zambian government engages the political elite stakeholders in ICT decision-making process at the expensive of the general public through television and radio – thus excluding the majority of the population in the rural areas without access to these communication media.

C. Opportunities

Aggressive adoption and penetration of mobile technologies in Zambia has created innovative opportunities for e-government in the area of m-government. A 75% mobile-penetration rate has been reported for Zambia [38]. Reference [36] reports an 18% internet-penetration for Zambia as at the last quarter of 2015. Taking advantage of m-government opportunities led to the growth of mobile money transactions in Zambia and successful mobile-community-based projects like Mwana m-health programme and Linknet's rural Macha community internet-connectivity.

Liberalization of the ICT sector and extensive efforts in Zambia for public-private sector and public-NGO collaborations in ICT projects have opened-up new opportunities for linking remote-rural communities with the rest of the country and the world for inclusiveness and ICT-knowledgeable-society.

D. Threats

There are insufficient highly-skilled ICT professionals in the Zambian government sector to champion e-government projects [15]. Implementation is one major failure-point of e-government initiatives for Zambia, just like for most developing nations. A project champion links-up public-strategic e-government goals with private sector and NGOs funding. For whatsoever reasons, this is missing in Zambia.

The proliferation of digital-devices with internet capabilities and delivery of ICT pedagogy in schools are inevitably concentrated in the affluent urban areas of Zambia. Sooner or later, the government is confronted with widening digital-divide challenges and e-inclusion barriers to resolve. However, with a constrained economy, this remains one of the wish-list on Zambia's national ICT policy agenda to resolve.

Endemic corruption and nepotism in awarding tenders for ICT projects led to inferior government-centric ICT project implementation and sometimes non-implementation of projects. This has significantly contributed to the huge number of reported ICT project failure in Zambia and the SADC region [24, 37].

X. CONCLUSION

The purpose of this study was to critically analyse the progress made in e-government adoption in Zambia. Through thematic analysis, findings reveal that Zambia has made some noticeable progress in formulating ICT policies to guide the e-government adoption roadmap. However, there are numerous grey-areas requiring immediate attention for e-government to achieve its set goals, typically, enacting an e-government implementation framework that reduces the existing gap between policy formulation and practical implementation; device a turnaround strategy aimed at overcoming the widespread digital divide and e-government service utilization divide; intervene and make affordable the cost of accessing internet by citizens; create awareness of e-government services through outreach programs and ensuring that e-government projects are presented in vernacular languages and addressing local problems. The major limitation of this study, however, is its overreliance on secondary data sources as its main data source. Future research should investigate e-government utilization by citizens using primary data sources. As of now, absolute assessment of user satisfaction with e-government, either qualitatively or quantitatively is still missing in literature for Zambia – thus calling for further studies.

REFERENCES

[1] World Bank, ICT for greater development impact 2012/15, 2012, Washington, DC 20433, USA

[2] B. van Niekerk, K. Pillay, and M. J. Maharaj, "Analyzing the role of ICTs in the Tunisian and Egyptian unrest," *International J. of Communication*, vol. 5, pp. 1406–1416, 2011

[3] World Bank, World development indicators, 2014, Washington, DC 20433, USA

[4] S. Habeenzu, "Zambia ICT sector performance review 2009/2010: towards evidence-based ICT policy and regulation," vol. 2, policy paper 17, 2010.

[5] M. Mzyece, "A critical analysis of e-government in Zambia," *The African J. of Information and Communication*, vol. 1, pp. 110-127, 2012

[6] K. J. Bwalya, and S. Mutula, *E-Government: implementation, adoption and synthesis in developing countries*, 2014, Walter de Gruyter: berlin, Germany

[7] E. Ruhode, "E-Government for development: a thematic analysis of Zimbabwe's information and communication technology policy documents," *The Electronic J. of Information Systems in Developing Countries*, vol. 73, pp. 1-15, 2016

[8] Knoema, Measuring the Information Society Report, 2015: ICT, Telecommunication World Rankings. Available: http://knoema.com/

[9] V. Braun, and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, pp. 77-101, 2006

[10] H. R. Bernard, *Qualitative data analysis: systematic approaches*. 2012, Newbury Park, Sage Publications.

[11] International Telecommunication Union, *E-Government: ITU e-Government Implementation toolkit*, 2009

[12] N. M. Ochara, "Emergence of the e-government artifact in an environment of social exclusion in Kenya," *The African J. of Information Systems*, vol. 1, pp. 18-43, 2009

[13] International Telecommunication Union, *ICT development index 2015: Compare countries or regions*, 2015

[14] International Telecommunication Union, *Committed to connecting the world*, 2016

[15] K. J. Bwalya, "Zambia to become an information society by 2015: a reality check," *Chinese Librarianship: an International Electronic J*, vol. 29, pp. 1-11, 2015

[16] Engineering Institution of Zambia, *Report on ZESCO load shedding, technical experts team*, 2015

[17] G, van Stam, and D. van Greunen, "Review of an African rural internet network and related academic interventions," *The J. of Community Informatics*, vol. 10, pp. 1-15, 2014

[18] Macha Works, Projects: Most popular projects, 2016, Available: http://www.machaworks.org/en/projects.html

[19] PANOS, ICTs and development in Zambia: challenges and opportunities, 2011

[20] K. J. Bwalya, *E-Government adoption and synthesis in Zambia: Context, issues and challenges*, 2011, PhD Thesis, University of Johannesburg.

[21] Zambia Information and Communications Technology Authority. The Zambia information and communications technology authority, 2015, Available: http://www.zicta.zm/

[22] Zambia Ministry of Communications and Transport, National information & communication technology policy, 2006, Lusaka, Zambia

[23] S. Habeenzu, Zambia ict sector performance review 2009/2010: towards evidence-based ict policy and regulation, vol. 2, Policy Paper 17, 2010.

[24] M. Mzyece, "A critical analysis of e-government in Zambia," *The African J. of Information and Communication*, vol. 1, pp. 110-127, 2012

[25] World Bank, Data: sub-Sahara Africa, 2015, Available: http://data.worldbank.org/country/zimbabwe

[26] International Monetary Fund, *Regional economic outlook: sub-Saharan Africa dealing with the gathering clouds*, 2015, Washington, D.C, USA

[27] R. Klitgaard, Addressing corruption together. In OECD Symposium on Anti-Corruption Development Assistance, France, 2014

[28] S. Zondi, *Governance and social policy in the SADC region*: An issues analysis. Development Planning Division Working Paper Series No.2, DBSA, Midrand, 2009

[29] NORAD. Joint evaluation of support to anti-corruption efforts Zambia country report. Report 6/2011, 2011, Norway

[30] Y. K. Majdalawi, T. Almarabeh, H. Mohammad, and W. Quteshate, "E-government strategy and plans in jordan," *J. of Software Engineering and Applications*, vol. 8, pp. 211-223, 2015

[31] A. Alawneh, H. Al-Refai, and K. Batiha, "Measuring user satisfaction from e-government services: lessons from Jordan," *Government Information Quarterly*, vol. 30, pp. 277-288, 2013

[32] K. J. Bwalya, and M. Healy, "Harnessing e-government adoption in the SADC region: A conceptual underpinning," *Electronic J. e-Government*, vol. 8, pp. 23 –32, 2010

[33] R. Danila, and A. Abdullah, "User's satisfaction on e-government services: An integrated model," *Procedia-Social and Behavioural Science*s, vol. 164, pp. 575-582, 2014

[34] N. M. Yaghoubi, A, Haghi, and S. K. Asl, "E-government and citizen satisfaction in Iran: empirical study on ict offices," *World Applied Sciences J*. vol. 12, pp. 1084-1092, 2011

[35] United Nations, *E-government Survey: E-government for the Future we want*, New York, 2014

[36] Internet World Stats, World internet users and population statistics: internet penetration in Africa, 2016, Available: http://www.internetworldstats.com/stats1.htm

[37] R. Heeks, *Most e-government-for-development projects fail: How can risks be reduced?* Institute for Development Policy and Management. University of Manchester, 2003.

[38] BuddeComm. Zambia: Telecoms, Mobile and Broadband; Statistics and Analyses, (2016), Available: http://www.budde.com.au

[39] D. M. West, Digital divide: *Improving internet access in the developing world through affordable services and diverse content,* Brookings Institution, 2015

[40] The World Bank, Digital dividends: world development report, 2016,Washington DC, USA

[41] D. P. Smith, "E-government initiatives case study: new models for success," *International J. of Applied Management and Technology*, vol. 15, pp. 21–40, 2016

Regulation of Natural Monopoly Via Competition Policy

Adalat Muradov
Azerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
rector@unec.edu.az

Nazim Hajiyev, Fariz Guliyev
UNEC Business School, Azerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
hajiyev@unec.edu.az, f.guliyev@unec.edu.az

***Abstract*—On the basis of experiences in developed and developing countries issues of regulation of competition policy and natural monopolies has been studied from both viewpoints, theoretical and practical, and recommendations and suggestions for transition economies have been made in this direction. Actually antitrust law needlessly inhibits agreements among competition designed to development and commercialize new technology.**

***Index Terms*—Regulation, natural monopoly, competition policy, institutional monopoly, concession**

I. Introduction

Natural monopoly is traditionally associated with such a strong effect that its significance allows to any firm to supply the entire market by incurring less costs per a product unit than any other competing firms. It is connected with fixed high costs, i.e. scope of which is independent on production volume. The higher the production volume, the less the standing expense per a product unit would be accordingly. That is, the same amount is distributed among growing number of products and services.

In such areas, the competition is not effective. If the market had been shared by many firms, it would be impossible to obtain a full size scale effect. As a result, expenses per product units would be high thereby requiring to be reimbursed.

II. Methodological Approach

Two methodological approach contradicting each other, one based on a theory of the natural monopoly, while another one denying its existence, to a competition policy in infrastructure areas set forth from different understanding of the nature of a competition typical on the one hand to classic and neo-classic directions, and on the other hand to Austrian and Neo-Austrian school.

Adam Smith is fairly considered a founder of the classic school. His aim was to a definite optimal variant for ensuring the people live in luxury. Thus, such a natural way according to Smith and his successors was a free competition considered as a power returning the market to a definite balance through the mutual connected mechanism of demand and offer. Such an approach is typical for the neo-classic school accepting the competition as an optimal economic environment for activity of firms.

According to Alferd Marshall, a founder of neo-classic direction, a free competition was better than other forms of arrangement of economy because of forming lower prices for balance. He justified significance and relative stability of existence of the market balance as well as the unity of a balance price. A notion about unity, sustainability and natural determinism of the balance in a perfect competition circumstance has found its ending in Leon Valras' general theory on economic balance.

Marshall did his best to enter a time factor into the economic analysis. He has chosen three periods of time – instant, short and long ones focused on assessing the balance in the market. But an analysis he referred to a mechanism of obtaining and maintaining a balance, is still static. A. Marshall, put the beginning of a technological conception on a competition connecting the concentration of production with a large saving [2, p. 189]. When explaining advantages of the large-scale production he wrote: "Expansion of its production scope rises its superiorities before rivals in no time and drops prices by which it could sell its products, and if the enterprise is able to maintain itself within hundreds of years, this enterprise and other one or two similar ones will distribute among them all the production field it operates".

Josef Schumpeter offered a quite different approach considering the competition as a dynamic process by its nature. Broadly, consideration of a competition firstly offered by Friderick von Hayeck, a representative of Neo-Austrian school, became a continuation of his conception. Hayek acted from the view-point of just using not properly the science of economy by "competition" term. To his viewpoint, it is necessary to consider the competition not just as a strategy of mutual relations of firms in the market, but more broadly as a motivating power of the economic life. Expressing these thoughts, Hayek puts a limit between the competition which is a dynamic process by its nature, and a static variant of a competition model introduced by the market balance. [4, p.50-51].

In spite of ideological closeness of Austrian school and neoclassism, neo-Austrian school representatives criticized the interpretation of the balance as an ideal state of an economic

system having no stimulus for changes and development, just because that it was favorable to do this.

In this case, as an example we can see two mutually connected directions of criticism of the neo-classic approach by neo-Austrian school. For the first direction founded by Ludwig von Mizes, a denial of a statistic approach to a balance where the competition process is changed by analysis of the state of the market in a concrete circumstance, is typical. Essentially, the matter is Smith's call to refusal from mechanistic panorama of the economy expressed in Valras' model.

If to consider the competition as a dynamic competition process among entrepreneurs, the fact that one of them attained with less expenses will not assume a great importance at present moment. The entrepreneur will not become a natural monopolist, because a permanently working mechanism of the market competition excludes formation of such a monopoly within free market circumstances. Therefore, the natural monopoly conception is a very static theoretical model. From this regard, one can conclude that a large-scale, capitalintensive production does not cause formation of the natural monopoly, just acts as a significant aspect of the competition process.

Neo-Austrian school contrasts such a dynamic approach to consideration of the competition as a constantly recoverable process assuming any domination in the market, against the neoclassic approach based upon the technological conception of the competition. In neoclassicism depending on production volume, fields are divided into groups on criterion of constant, diminishing or increasing profitability. In such an approach, the competition has been essentially considered as a technological relationship.

As possible ways for ensuring acceptable social behavior of natural monopoly, there are used two main alternatives – state adjustment and state property. In USA and Great Britain, the first alternative is preferred, while in continental Europe major part of natural monopoly subjects belongs to the government (speaking frankly, many enterprises have been privatized within past years). Nevertheless, in different countries regulatory agencies of natural monopoly use separate schemes: an establishment system, an independent regulatory agency, a competition institution (the latest for instance, in Russia, Bolivia, the Netherlands, Australia).

Members of developed countries prefer an independent regulatory commission selected for several years. Apart from an establishment scheme, members of such a commission are not responsible for current performance of the field. This allows them to resist attempts of natural monopoly subjects towards increase of prices and improvement of their indices on account of consumers. On the other hand, if the government fixes prices, (some times for instance, using political strength to satisfy electors) the prices are falsely remained at lower level. In such cases, the independent commission resumes economically significant, but politically unpopular decisions.

Key purpose for regulation of natural monopoly is fixing prices, which leads to "normal" earning of regulated firms. Thus, estimation is applied through a principle "expense plus" which implies consideration of all expenses socially significant beforehand. This creates a stimulus for ceasing the augmentation of expenses of firms. Actually, if it will be resulted in price drops but not in increase of income, it is necessary to try to diminish expenses and launch a new technique, technology or industrial union. Furthermore, diminishing expenses will result in lessening income in such estimation.

In other words, regulation on natural monopoly creates auspicious conditions for a great X-ineffectiveness. X-ineffectiveness manifests itself when actual expenses are more than possible minimum ones in a firm's any large-scale production. Usually, X-ineffectiveness can be interpreted as expansion of the firm against the desire of managers to minimize expenses, much easier life at work, attempt to avoid an entrepreneurship risk, provision of inexperienced friends with job and salary, and as other personal purposes. Another source of X-ineffectiveness can be ill-stimulated employees of a firm.

In practice, price regulation of natural monopoly is performed through two models – regulation of an income standard and regulation of tariffs level. In order to set a possible income standard, income risk generated from investments is estimated with degree of risk in competition-based fields. However, as this model incurs much expenses, it is rarely applied. This is why regulation of the tariffs level is more employed nowadays. However, it has shortages as well. In order to eliminate them, in some countries a hybrid model equalizing self-regulation, tariffs level and income standard, is applied.

Some scientists' critical attitude towards the theory of natural monopoly is explained so that, even in "competition-based" economy in America the government regulated shipments and air transportations through motor transport where scale effect is not great, and natural obstacles are enough weak against the entry of new rivals. Everybody knows that in such cases the government has institutionally limited entrance of new firms to these fields and created its own monopoly. As a consequence, both consumers and potential rivals used to lose.

Institutional monopolization of a potential competition-based industry, actually, implies formation of a legal cartel in it, which may be very profitable for regulated firms. Herein, a regulating commission can fulfil a function of a cartel – division of the market. Thus, management of American civil aviation has distributed routes among air-companies. Apart from easily-distributable simple cartel (if not to speak of a risk related with its illegality), advantages of the cartel underpinned by the government under the name of regulation is obvious. Besides, government officers render regulation "services" for firms desiring to be regulated in order to obtain a guaranteed competition-free market.

Activity of Inter-States Trade Commission in USA may be brought as a classic example for such a legal cartel. Regulation of work of railways by it in early XX century justified itself. But in 30-s a high-way network was created in United States and transportations by lorries seriously shook the railway monopoly power. It seems that, to allow competition among different types of transport would be efficient with regard to social interests. Instead of this, they followed the way as to

expand regulating activity of the commission and added shipment-motor transport and airlines to its field. This time, the Law on Inter-States Trade expanding competencies of the commission has been supported not only by railways, but also by both motor-transport companies and air-companies. With a view to avoid of the danger caused by competition and ensure stability and able to guarantee the legal cartel, they supported application of regulation in their fields [6, p 228- 229].

Liberal movement taking advantage of increasing number of extravagance and non-efficiency cases in areas governed by the legal cartel practices and its use in its support to deregulation in USA in 70- 80s proved as compelling evidence.

In the result, the legislation (deregulating) which supported air transport, motor transport, banking, railways, natural gas output and tale broadcasting to a certain extend was accepted. However, these reformers had their own critics and they confirmed that deregulation would cause to gradual monopolization of these fields thereby resulting in augmentation of prices and tariffs and diminishing of production. Another danger was the fact that deregulation might give rise to infinite competition and instability in this field. This may lead to deprivation of small territorial units from significant services for life in general. Moreover, it was presumed that, as the strengthening of competition reduces incomes of a firm, they may lessen requirements against level of safety and risk. Further events proved such cautions to be unjustified.

Success of the regulation (deregulation) offered an opportunity for dissemination of ultraliberal imaginations objecting the reality of existence of natural monopoly and absolution of methodological approach to consideration of competition as a process offered by the neo-Austrian school. Neo-Austrian theory on a competition process undoubtedly is an important contribution to the science of economy. However, conception of “legend about natural monopoly” based on that theory allows mixing of two different categories – natural monopoly and institutional monopoly.

In many cases, natural monopoly is accompanied with formation and development of elements of institutional monopoly. In this sense, natural and institutional monopoly, indeed, may be in close connection with each-other. However, this mutual relation may not serve as grounds for identification of absolutely different types of the monopoly. One of these types has been technologically determined, while another one is related with unlimited interference of the state authorities with economy. Besides this, an ultraliberal conception of “legend about natural monopoly” arises from just such identification. In fact, as it categorizes existing connection between institutional monopoly and natural monopoly in infrastructure areas, paying no attention to a quality difference between them, this approach is unilateral.

Neo-Austrian theory on competition process allows balancing the statistic approach of neo-classism to market analysis and competition. As methodological approach depends on mainly analysis item, with regard to natural monopoly, it is necessary to accept traditional neoclassic approach as a more realistic and a more suitable to their theoretical reflection, approach. Natural monopoly is sufficiently sustainable by its nature, and static in this sense. Indeed, its natural level is defined with technological necessity, and technology itself should develop. Scientific-technical progress is able to cease natural monopoly in a certain field and to make the market competition more effective. For instance, output of artificial gas from coal manifested itself as natural monopoly, but in XX century it was again destroyed by gas and electric power, which is natural monopoly. Natural monopoly in railways has been much restricted due to competition between types of transport. Analogical mobile communications have restricted natural monopoly of wireless communications.

However, these changes influence on structure of natural monopoly and on composition of fields where natural monopoly subjects operate, not countermanding by all means the existence of technologically determined natural monopoly. On the contrary, through development of technique and technology, natural monopoly attains tendencies for expanding its field of authority. Technological progress leads not to disappearance of natural monopolies, on the contrary, to their formation, for instance, as in the aviation transport.

It is naturally that, in several cases, a close connection between objectively existing natural monopoly and institutional one does not serve as grounds for identification them; a conception of ―natural monopoly myth‖ arises just from this. Principal differences between natural and institutional monopolies define necessity of different approaches to them. Effective competition policy requires elimination of all manifestations of institutional monopoly, as well as its elements formed as harmful superstructure upon natural monopoly in infrastructure areas. Therefore, a scientific analysis directed towards exactly differentiating institutional monopoly and natural monopoly in a maximum degree is necessary.

In countries with developed market economy, state regulation of activity of sea ports is considered as a factor streamlining their activity. These decisions adopted by port administration, usually, is under control of the government: annual budget, all foreign debts, annual budget over the known limit, obligations and commitments, contracts for large-scale activities (whose value exceeding the set amount), general level of main port charges, appointment of a chairman for Board of Directors and a head, sale and long-term lease of a port property, affairs related with the local safety and relationships with other countries. Even there is a competition among ports, the state coordinates their different fields of activity: estimation, investment policy etc. At this time, the state interferes with the field of estimation in the following basic forms:

- Require from all ports of a country to adopt a unique estimation structure (provided that they were independent to set up basic interest rates).
- Control over incomes from port charges in order to ensure a level fitting to national positions (for instance, a government allows a new system for port charges provided that general amount of incomes shall not grow or grow up to a certain rate).

- Control over key level of some port charges used against privileged clients of a port (for instance, some exported and imported goods assuming a great interest for a country).

The matter about whether ports should completely justify their outlays is solved differently in each country. Some ports, including those in Western Europe, receive huge subsidies allowing them to hold port charges at a lower level. So that, in France the state resumes 60% to 80% of outlays for maintaining sluices, protection devices etc. Currently, the Transport Commission of EU is not intending to elaborate a Unique All-European policy for regulation of the activity level of ports, but its policy pursues establishment of equal competition terms and conditions.

Although airports are engaged in large entrepreneurship activities and receive incomes, leases etc. from concessional charges, they are not considered a natural monopoly subject. Much of time, airports (mainly big ones) obtain more than half of profits generated from commercial operations beyond air traffic.

Generation of such high profits from the non-aviation activity acts as grounds against rise of airport charges. According to classification of the international civil aviation organization, airport charges are received for landing, services for clients, loads treatment, usage of terminals and warehouses etc. To opinion of the international organization, basic principles to be observed during collection of airport charges are users' payment abilities for airport expenses.

Activities of an airport related directly with air traffic (tariffs, investments, profitability) should be subject an entire regulation of the state. Although the role of the state is similar in separate countries, such a regulation is less or more strict everywhere.

Notwithstanding that participation level of the state in natural monopoly activity, for instance in railways, seaports, airports, it shows itself as economically balanced through tax and amortization concessions in the form of an investment, debt and subsidy in local administrations, as well as through scientific and other program and activities. Specify of such participation is; its too much elasticity manifesting itself, depending on the state of a field and the market conjuncture, in strengthening or weakening of interference.

In late XX century, an attitude towards state property in infrastructure areas across the West changed. Before, in many developed countries it was reckoned that nationalization is the best means for restricting a monopoly in these fields. However, the practice proved that the nationalization is not always feasible for solution of problems arisen due to natural monopolies. It may lead to substitution of monopolies with other more dangerous ones. It is proved especially by multiple non-effective activity facts of state enterprises in infrastructure areas.

Analysis of state of water supply is of interest as well. Water supply has been one of historically socialized services. It was an objective necessity in centralized water supply and irrigation organizations having laid down the economic foundations of Asia-like output typical for eastern despotism. Even in the United State being a traditional ―competition‖ country, private water supply firms served for 31 mln people as compared to social structures supplying water for nearly 180 mln., i.e. six times more people in 80s of XX century.

But in France, just private companies are main water suppliers. In Paris, the private water supply system has been operating since 1782. No wonder that, the French are leading worldwide in management and provision of water supply services, as well as in scientific investigations, researches and equipment production [8, p.197].

Generalization of discussions about privatization problem showed that, during analysis of comparable effectiveness of services rendered by private and national enterprises in the field of infrastructure, half of private and privatized enterprises displayed more effective level in 24 cases, while in 7 cases there were few differences or unclear. Only in five cases, investigators made such a conclusion that national enterprises are at higher level of effectiveness. [9, p. 152].

We think that, transformed and highly developed economy differs only for a scale of manifestation of the institutional monopoly. Creation of a "virtual economy" in countries with highly-developed economy is not matter of discussion at all. Only separated elements of the institutional monopoly may exist here.

Economic achievement and right of the property should be distinctive with regard to methodology. The property is a social complicated manifestation and may be divided into two aspects such as volitional and economic. The first aspect is generated from the property right, i.e. legal relationships considering to possession of property objects, their usage and management. Another branch forms essentially objective relationships of economy. The property right gets economically realized when raising an income to its owner (like any income and profit understood broadly, i.e. obtained by an owner). Without this, a property right is only a tale from legal viewpoint.

III. Competition And Innovation

The basis of competition is that technological innovation drives competition. That the sources of innovation are remarkably diverse, and the antitrust laws which we have inherited are informed by static theories of market performance and therefore are as likely to throttle innovation as to stimulate it. In short, we believe 1) that stimulating rivalry ought to be an important policy goal. 2) that the form of competition the antitrust laws should embrace is Schumpeterian dynamic competition. 3) that current antitrust is not designed to achieve this. 4) that the best guarantor of dynamic competition is a system that is open to international trade and has policies which facilitate innovation, and 5) that internal organization and interoperate links, and not government intervention, is important for successful innovation and commercialization [11].

IV. Concession Legislation

In this connection, concession may be more effective than privatization of a natural monopoly subject. Concession legislation exists in over 120 countries. According to Unique

market rules adopted by the European commission in December 1999, concession is a system of relationships where a social power passes an authority of an infrastructural object or of other social services to the third party (a private or mixed one) if it should undertake an entrepreneurship risk and responsibilities for maintenance of equipment. In other words, the state gives its property to a private or mixed company with limited time and payment terms and conditions set up in the form of a contract [10, p.14].

At this time, a property form remains unchanged. Although state authorities (local authorities) promises not to interfere with economic activity of the concessioner, they control over their fulfilment of contract-related responsibilities. When the concession term is over, the object, including industrial objects created by the concessionaire is returned back to the owner. So, a potential unity of two economic agents occurs in a concession: on the one hand, of a state property (a municipality), on the other hand, of private bodies and investments.

Concessions occur, mainly in water supply, gas economy, electric power, motor-road constructions. However, the concession does not address a problem concerning formation of a competition environment in infrastructure areas.

At the first sight, competition in such fields seems impossible. Indeed, for example, if to split a railway network monopoly exploited by a company into several areas and give them to independent companies, hardly will it rise the competition level. Moreover, such a reform can lower effectiveness. Arrangement of a competition in rendering public service is more difficult. For it is impossible to install two or three water-taps for running water in each flat, supplied by separate companies!?

Natural monopoly franchise, i.e. performance of next minimum of terms and conditions for organizing an auction according to a right of being a renderer of a relevant service is necessary (but not sufficient): existence of several educated bidders, exception of a transaction among them, as well as between any bidder and body organizing the auction. Key criterion for selection of auction winners is prices for services. Auctions for a natural monopoly franchise should be organized periodically. Supremacy of such a system, the first of all, is its being competition-natured. Apart from the state structure, when a private firm concludes a profitable contract, it does not get a guarantee for its eternity. It should never forget its rivals, because if the firm does not operate sufficiently effective, afterwards they can seize up the contract in order to revenge.

Similarly, infrastructure areas may be divided into several functions each fitting to a certain output link having its own place in unique technology chains built on differential optimization principles. Each function, each part should have its own appropriate, optimal production size. The problem is not only to define an optimal size for socialization of this function, also to find ways of consistency among separate parts.

This drew attention, for instance, during development of Internet services. It is clear that, "subject" of Internet is disseminated to a vast territory by its main providers, and then delivered to users through Internet service providers. They in turn (in any case, in current technological conditions) should use local communication or cable television networks in order to deliver it to final users. A significant question to be answered by regulatory or competition bodies is determining the extent of harm which may be inflicted by a vertical integration in this field [9, p. 157].

A practice on how the network is organizationally separated from its operation functions has been common in the West within the last decade. At this time, they try to find and distinct functions which may be organized on basis of competition within infrastructure areas.

In such cases, economic essence openly contrasts to property right. "Municipalisation" of ownership of street lighting by a formal law acts as a manifestation of nationalization of natural monopoly. However, this historical mechanism allows for competition-based operation of the network, therefore, its economic essence objectively becomes adequate to a private property. The Americans prefer to economic essence of property. Competition is more significant than a formal property right for them. That's why, for its development, they may agree even to "nationalization".

In other fields of infrastructure, there is also an opportunity for differential optimization through splitting into composite elements and using an economic mechanism optimal for each element. This time, the infrastructure field is considered not as any monolithic Leviathan, but as a complicated system of functions and operations requiring different approaches. Such an approach allows splitting it into several parts. During monopolization of functions to be fulfilled more effectively based on competition, each of them may possess institutional monopoly elements. Conducting such parts based on a contract and as a result, of competition acts as "clarification" of natural monopoly from institutional monopolism elements.

V. Deregulation And Privatization

Government in developed countries again actively implements deregulation and privatization. But when using their practice, the following cases should be taken into consideration. Privatization and competition are closely connected with each other; therefore, maintenance of state property thwarts development of competition relationships. However, privatization in infrastructure areas requires a carefully measured approach. It should not consist of just substitution of institutional monopoly with ownership one. Practice shows that, private companies having seized the natural monopoly may work non-effectively as well.

Changing the economic essence of property through maximum development of competition relationships, but not by a legal form, is duty of state policy in infrastructure areas. Selection criterion of a property form should be the first, "competitiveness" of relationships constituting its economic essence. From this viewpoint, as noted, in some cases usage of a concession may be more effective than privatization of a natural monopolist.

Privatization stimulating development of competition is unreservedly necessary. This, for instance, concerns to cleaning and asphalting streets being urban economy fields. Although they are not natural monopoly, much time are reviewed as

functions of local self-managing bodies. Municipal services, as a rule, are rendered less effectively in these fields than private firms. Fundamental investigations conducted in 20 cities of America proved that price offered by such services for road sweeping a mile of road verge was 43% more than expenses of analogical contract-based activities. For ensuring higher exactness, these calculations were made for the same cleaning and service quality (the latter was proved by a rating on check of cleanness) and the same urban conditions. Difference was interpreted with higher productivity of contract brigades, abundance of managers and shortage of executers in municipal agencies, as well as with better maintenance of equipment by private contractors. This time, to compare more exactly, expenses for control over preparation of cities, conclusion of contract and contractors' activities were taken into account as well [8, p.174-175].

Two different organizational mechanism used during asphalting streets has been exactly learnt and compared as well. It appeared that, activities undertaken by municipal agencies is almost twice expensive those implemented by contract-based private firms. Moreover, analysis of further state of asphalt cover proved quality of contract-based activities to be higher. On annual basis, contractors laid averagely 4508 tonnes of asphalt per a worker, when municipal agencies 1180 tonnes, i.e. only one fourth of this volume. This deep difference in high productivity compensated higher salaries for contractors [8, p. 175].

Explanation for existence of such non-effective municipal services is nothing other than institutional monopoly, establishment of unequal competition terms for private and municipal enterprises. In these fields of urban economy, institutional monopoly is striking especially, because here it fails to hide itself under the mask of natural monopoly. It is difficult to discover institutional monopoly in infrastructure areas where indeed natural monopoly exists and this complicates struggle against it. Elements of institutional monopoly are formed and develop not just a monopolistic "superstructure", also combine with natural monopoly creating them. That is why, there needs a measured and scientifically grounded competition policy allowing to systematically distinct natural and institutional monopoly combined with each other in such infrastructure areas.

Subject of the concession contract proves that it should block up any abuses in monopoly position. Something contradicting to institutional monopoly is observed here. Such a monopoly oppresses the competition in all places wherever it is possible. Even usage of the concession in existence of natural monopoly does not lag behind a result turned out owing to existence of competition in the market. In this context, concession system may be considered as appearance of institutional competition allowing neutralizing many negative appearances of natural monopoly.

We think that, institutional competition is a market-oriented competition falsely stimulated through institutional means thanks to usage of concession mechanisms. It may exist in an infrastructure field possessing technological characteristics impeding market-based competition. Institutional competition essentially contrasts to institutional monopoly falsely spread by government in places of the market fitting for competition. In both cases, institutes turn out to be relatively independent on technology, if institutional monopoly leads to diminishing the effectiveness; on the contrary, institutional competition helps improvement of infrastructural activity.

CONCLUSION

Therefore, necessity for their exactly differing them, the first of all, is connected with requirement of natural monopoly for a different approach by state. Competition policy offers to struggle against institutional structure combining formal and informal rules and economic behavioral standards corresponding to institutional monopoly and forming it. This policy is directed against interference of local and central authorities with competition relationships thereby creating unequal competition conditions for different participants. By this, it has been directed to deregulation and liberalization of economy.

It is necessary to regulate natural monopolies rather than destroying them. Yet, such a regulation must be carried out in a flexible condition where competition is existent and efficient.

It is also clear that the kind of competition that comes from the firms with new technology.

REFERENCES

[1] D. Lorenzo T. The Myth of "natural monopoly"//EKO. 2001- # 4- pp 81-98.

[2] Marshall A. The principles of economic science. - M.: Progress, 1993.- V.2.- 384.

[3] Hayek F. Cognition, competition and freedom. - M.: Pneuma, 1999.- 287 p.

[4] J. Clark. B. The distribution of wealth. M.: Gelios ARV, 2000. 367 p.

[5] McConnell C.R, Brue S. L Economics: Principles, Problems and policy. V 2 - Baku, "Azerbaijan", 1992. - 400 p.

[6] Bruzik F. Globalization: increased competition or create a monopoly? // Regional Conference on Competition Policy of the CIS countries, Central and Eastern Europe: The collection of materials. - K., 2000. - P.7-18.

[7] E Savas S Privatization: Key to Market: M.: Business, 1992. 411 p.

[8] World Development Report 2002. Creating the institutional framework of the market economy: Eng - M.: The world, 2002 – 250.

[9] Williamson O. Economic Institute of capitalism: firms, marketing, opinion agreements. - K.: 2001.- 472 p.

[10] Kokorev B. Institutional reform in the field of infrastructure in the natural monopoly//Problems economy. - 1998 - № 4 - S. 127-133.

[11] Strategy, technology and Public Policy. David J. Teece. 1998, p.362.

Features Electronic Payment Systems in The Information Economy

A.F.Musayev[1], A.G.Aliyev[2], E.H.Musayeva[3]
[1]Doctor of Economics, Professor, Correspondent members of ANAS (Azerbaijan National Academy of Sciences),
[2]Head of department of the Institute of Information Technology of ANAS, Ph.D. in economics,
[3]Senior scientist of Institute of Information Technology of ANAS
[1]akif.musayev@gmail.com, [2]alovsat_qaraca@mail.ru, [3]depart8@iit.ab.az

***Abstract*. The paper substantiates the relevance of non-cash payments and transactions in the formation of the modern information economy. Modern electronic payment systems are classified, and requirements to them are shown. The process of improving payment systems through POS terminals and the foreign experience in this process are analyzed. A number of proposals and recommendations are put forward basing on a real statistic as the result of the analysis of non-cash transactions in Azerbaijan.**

***Key words*: electronic payment systems, non-cash payments, non-cash transactions, POS terminals, ATM, payment cards, Internet banking.**

I. INTRODUCTION

In accordance with the development trends of emerging knowledge-based economy in the world, the modern and dynamical development strategy of the economy of Azerbaijan is based on the national development model determined by the national leader Heydar Aliyev. The source of this strategy is based on socio-economic policy of new content, which is implemented by the follower of the political and statehood legacy of the great leader, the President Ilham Aliyev. This policy has recently led to the division of the economy, to the rapid growth rate of the non-oil sector on its structure, to the development of high-tech industries, to the enhancing of international ratings of the national economy, and ultimately, to the inclusion of the country into the group of countries with middle-income [1].

Knowledge and information is a source of the development of ICT at the present stage, while an important component of economic strategy is monetary and fiscal policy, main directions of which is aimed at ensuring macroeconomic stability, stability in the banking sector, and deepening financial intermediation. To achieve these goals, monetary policy and banking supervision is improved continuously, as well as the further development of the coordination of macroeconomic policy is supported, and most importantly, comprehensive measures are taken to increase the circulation of non-cash money.

As a result of the recent globalization of the national economy and the development of ICT, the volume of non-cash payments has increased significantly, which are carried out in domestic and international financial transactions through various types of plastic cards, the Internet and other electronic payment systems. Despite extensive reforms and achievements in the context of the formation of information society and information economy, a number of problems still remain in an efficient and absolute use of infrastructure opportunities and the capacity, technological integration of the related organizations into electronic payment systems, as well as providing access to electronic financial services in the regions.

The analysis the money circulation in the country shows, that the number of cash transactions is significantly higher than of non-cash transactions. An increase in the share of non-cash amount in general amount, on one hand, is substantiated with the urgency of new management of the economy and its monitoring, on the other hand, with the expansion of transparency in taxing and economic relations [2].

II. THE FOREMOST FUNCTION OF NON-CASH PAYMENTS

The foremost functions of the activities carried out in connection with the expansion of non-cash payments in the country, include the transition of the economy to the new stage of quality, achieving transparency within the entities of the economic system, restricting the volume of the "shadow economy", which cannot be observed, minimizing the tax evasion. At present, non-controlled economic transactions, on one hand, lead to the reduction in tax revenues in the state budget, on the other hand, to the reduction of effectiveness of the economic regulation methods by the state through economic mechanisms.

Classification and requirements of modern electronic payment systems. According to defined rules the payment system includes financial institutions realizing financial transactions between economic entities using a full set of methods and tools (bank and telecommunication transfer systems, technical equipment, payment tools, etc.). All of the elements of financial institutions are participating in the payment system. These institutions have certain organizational and management mechanisms ensuring stability and further development of the economic and social system.

The payment systems are a set of internal and external regulations, technical component elements, payment methods, technology, organization, and the rules, which

enable all participants to realize financial operations and transactions with each other (Figure 1).

Electronic technology-based payment systems must comply with the following requirements [3, 4]:

•confidentiality. Only legally relevant and allowed persons are informed about financial information;

•ensuring the integrity of information;

•authentication – confirmation of identity of the system user;

•authorization - confirmation of the transaction availability;

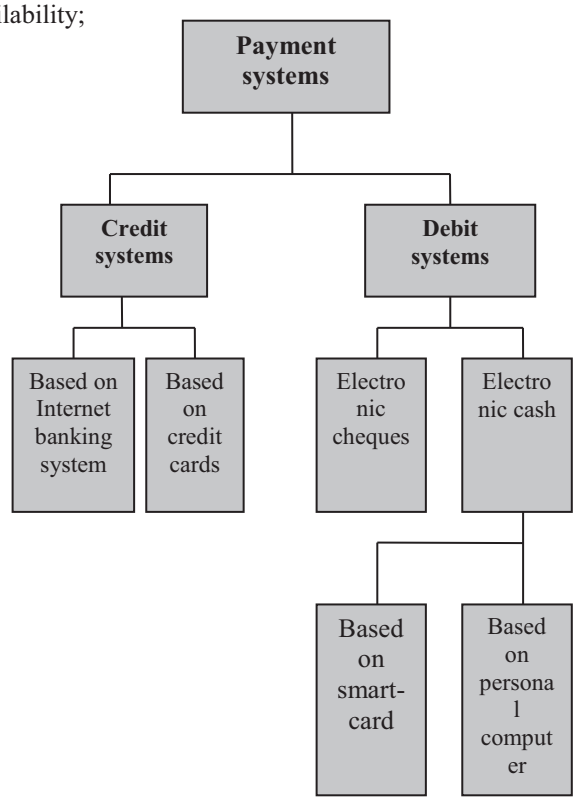


Fig.1. Elements of payment systems

•security;

•supporting a wide range of payment tools;

•reducing transaction costs and expenditures;

•verification and confirmation of the information accuracy about the payments to the participants of external arbitration and audit process during any conflict, etc.

Payment mechanisms are divided into three main groups depending on payment systems: 1) Internet payments; 2) mobile payments; 3) e-money payments.

III. Classification Of Payment Systems

Modern methods of electronic payment are divided into traditional and innovative methods. Traditional payment methods include traditional payment tools (credit cards, credit transfers, etc.) that contribute to the payment exchange through traditional means of payment (deposit money) for the transfer of payment information in electronic form. Innovative methods use the following payment means: 1) new payment tools (payment portal service, cell phones, etc.) that contribute to the payment exchange through traditional means of payment (deposit money) for the transfer of payment information in electronic form. 2)new means of payment (e-money). The first group of the innovative payment method is based on bank deposits. The main common innovation of these methods is the use of web-sites and e-mail for the communication between payment providers and consumers [5, 6].

The second group of innovative payment methods is based on new means of payment, i.e. electronic money. Electronic money represents the value of the money presented in accordance with the relevant requirements, stored in the specific device in electronic form, and used for the payment to the third party. Deposit money, that is transaction or cash stored in current accounts of credit institutions act both in traditional, and new means of payment methods. However, the use of e-money as a new means of payment is the most important recent innovation.

The most important factors for the development of new payment methods include functional abilities such as security, reliability, low transaction costs, anonymity, and so on. An extensive and intensive use of these methods is significantly dependent on these factors.

Improving modern electronic payment services through POS-terminals. The development of financial market, extended opportunities of investment in the economy, formation of unite e-payment space and improvement of modern e-payment services in regard with limited cash circulation, non-cash payment stimulation and so on are of great importance. These measures contribute to strengthening financial discipline in enterprises, improvement of transparency in the corporate sector, including the elimination of tax evasion.

Development of national payment system has led to the formation of a new service area. Thus, application of POS-terminals is the source of additional revenue of banking institutions. Banks receive appropriate service fees from enterprises for the use of POS-terminal and for the operations carried out through POS-terminals. This service is called acquiring in the international sphere. Thus, the entities are charged monthly fixed fee for the transactions carried out via POS-terminals, the monthly volume of which is 1000 AZN per month averagely. The monthly service fee is about 50 AZN. The monthly transaction service fee of the entities of more than 1000 AZN is 1-3 percent. Purchase and installation of POS-terminals, as well as communication services for equipments are covered by the bank service.

Note that the highest rate of tariffs reduces direct use of POS-terminals. Consequently, sellers do not tend to influence consumers to take advantage from plastic card payment, and ultimately it leads to the application of cash payment. The high rate of service tariffs can be regarded as one of the main factors preventing the expansion of non-cash payments.

IV. International Practices On The Use Of POS-Terminals

In Russia, acquiring amount required from the entity for the bank card payment via the same bank POS-terminal is 1-1.5 percent, acquiring amount for the card payment via other bank POS-terminal is 2 percent. These service fees are 0.5-2 and 1.5-3 percent accordingly in Kazakhstan. In Moldova, the service fee for POS-terminal transactions carried out by local banks is 1.8 percent, but by the foreign banks is 2.8 percent.

For comparison, note that in Ukraine acquiring amount required from the entity for the bank card payment via the same bank POS-terminal is 1-2 percent, acquiring amount for the card payment via other bank POS-terminal is 2-2.8 percent.

In Turkey, the situation is slightly different. Low tariffs are applied relatively for business entities in this country. So, this service fee 1.25-1.4 percent for the transaction conducted by the bank card, which installed POS-terminals, 1.7-1.85 percent for the transaction conducted by the other local bank cards, 1-1.7 percent for foreign cards. In Bulgaria, the citizens of the country are charged 0.7 percent for the card payment through POS-terminal, but juridical persons pay relatively more (1.7 percent).

Obviously, comparing to CIS member states, the rate of tariffs for POS-terminal transactions is low in other countries (Turkey, Bulgaria), which provides reduction of the bank costs of economic entities and increases their financial capacity [7, 8, 9].

V. Analysis Of Non-Cash Transactions In Azerbaijan

So far, the number of plastic cards used in electronic systems of Azerbaijan reached 5301 thousand by June 2013. 4665 thousand out of them are debit cards, while 636 thousand are credit cards. Many transactions are carried out through these cards. To analyze the characteristics of cash and non-cash payments, first of all, the structure and properties of plastic cards in circulation are of great interest.

Plastic card payments are carried out by social and salary cards. Thus, these transactions amount 70% of all transactions. Thus, credit card transactions were equal to 3 %, to 5.6% in 2012, and more than 11% in 2013. This rapid growth in early 2013 occurred as the result of the closure of credit services in the retail trade points, and due to mass distribution of small amount of cards.

Another point is the form of transactions via payment cards. In other words, the purpose of the user is cash withdrawal or non-cash payments.

Conclusions And Recommendations

Increasing the volume of domestic non-cash payment, developing card payment market, widening the scope of card payments for purchase transactions are of great importance for the country. Increasing the role of Azerbaijan in the modern globalized world, ranking 39th among 148 countries at many international ratings, including the Global Competitiveness Index, Azerbaijan raised its image in the international world. Restricting cash payments and stimulating non-cash transactions may cause further progress of Azerbaijan in this area, strengthening its position in the international economic relations.

In general, the status of the field of non-cash payments dose not coincide with the rate of economic development of the country, large-scale reforms, the goals and objectives of the development concept adopted by the president “Azerbaijan 2020: Vision to the Future”. In the globalized world economy, where non-cash transactions are dominating, the high rate of cash transactions in Azerbaijan proves to take imperative steps in this direction. The complex measures should be taken as soon as possible, in order to achieve transparency in the economic system, and the reduction of the “shadow” economy, and the strengthening of the financial system, as well.

References

[1] “Azerbaijan – 2020: A Look to the future” The Concept of Development. Baku, Dec. 29, 2012. http://www.president.az

[2] David Romer. Advanced Macroeconomics Second Edition. 2001. 651 pp

[3] A.G.Aliev. Theoretical and applied aspects of informatization of humanitarian fields. Baku, “Elm”, 2006.

[4] N.V.Baydukova. Organization of payment systems from the perspective of institutionality. J. “Problems of Modern Economics”, №1(13), 2005

[5] G.A.Nabiullina, A.G.Komarov. Key issues and trends in the development of electronic payment systems / "Problems of Modern Economics", №3(19), 2006

[6] I.S. Sirotin. Electronic money as a factor of economic development / "Problems of Modern Economics", №3(19) , 2006, www.m-economy.ru

[7] V.Yu.Kopytin. New trends in the development of retail payment services. J. "Calculations and operational work in a commercial bank", №1(95), 2010, www.gradient-alpha.ru

[8] T.Khiaonarong, J.Liebenau. Strategic Approaches to Payment Efficiency and Innovation. Banking on Innovation. Contributions to Economics 2009, pp. 29-53

[9] Q.W.Jun. Study on the Electronic Payment Technology in E-Commerce. Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012): Volume 4 Lecture Notes in Electrical Engineering. 2013, pp. 95-100

Sustainable Development of Manufacturing Enterprises Based on Structural, Infological and Dynamic Modeling

R. T. Sirazetdinov, A. V. Samodurov, A. N. Khusnutdinov, E. N. Tarchinskaya
Kazan National Research Technical University named after A.N.Tupolev - KAI (KNRTU-KAI)
Kazan, Russia

Abstract— **In the following paper we suggest the principles of structural, infological and dynamic modeling of processes within a functioning enterprise, allowing to effectively plan its strategic development and assure its sustainable functioning. We offer a new interpretation of sustainable development of an enterprise, based on separating the concepts of sustainability and development, and apply the concepts taken from reliability engineering to the functioning of an enterprise. The principles, methods and models described in the paper may serve as the basis for the automation of management and strategic planning of businesses based on cloud technologies.**

Index Terms— **Enterprise, modeling, sustainable development, business**

I. The concept of sustainable development

The concept of sustainable development has been discussed since the 1980s. At the UN Conference on Environment and Development in 1992 [1] the world community has agreed on the necessity of acquiring the model of sustainable development that would provide the balance between the solving of social and economical problems and the preservation of environment for the future generations, allowing to switch from modern approach based on utility and resources to the new biospheric-ecological approach.

V. M. Matrosov, member of RAS [2], one of the leading specialists on the problems of sustainability and one of the first scientists to acquire the methods of mathematical modelling in the paragidm of sustainable development, has more than once emphasized the ambiguity of the concept. It is usually interpreted as stable development (at least, economic growth).

We suggest a more profound understanding of sustainable development, separating the ideas of development and sustainability. According to our idea, the development of an economic system must comply with the principle of sustainability of its subsystems. In this case the strategy of managing the national economy of the state and its regions changes, including all short-term and long-term goals of the system of economic, ecological, social and industrial management.

The following scientific research is conducted within the scientific school of member of Tatarstan Academy of Sciences T. K. Sirazetdinov [3-8]. We are working with the definition of a complex system as a system containing multiple independent processes, thus the system is multi-regime, multifunctional and can operate and complete certain tasks within its smaller parts (subsystems) [9, 10]. This allows us to work with two simultaneous processes within the system – the process of development and the process of maintaining sustainability.

In this paper the principles of sustainable development stated above are applied to the problem of strategic management of an enterprise. For this purpose we have suggested several structural, infological and mathematical models of a functioning enterprise, allowing to plan the development of the enterprise, taking into consideration the degree of its sustainability.

The development of a business can take place only if new types of products are introduced and new markets are conquered. Evidently, it is connected with certain risks. On the other hand, if a business holds on to its well-developed schemes, its profitability decreases with time. The principle of sustainable development suggests the combination of old stable and new risky ways, allowing the enterprise to develop, conquering new markets and being stable at the same time.

II. Methodology of process modeling

The work is based on the research of system processes in businesses and various modeling methods of these business processes. The description of business processes is based on the *standard business unit* model (Fig. 1).

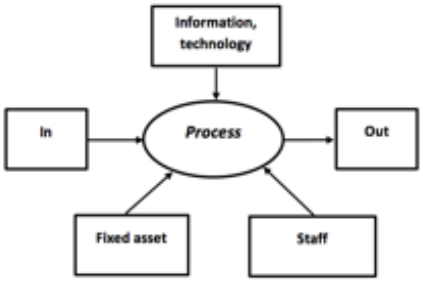


Fig. 1. Standard business unit

The idea of process is essential to many spheres of human knowledge, so there are multiple definitions of process. However, in most cases a process suggests the presence of Input receiving data or materials and Output producing the results of the process. IDEF methodology makes use of

additional inputs Controls and Mechanisms. We have expanded this model by adding a new interpretation of additional inputs: In, Out, Fixed Asset, Staff and Information-Technology.

The main *process* explored while solving our problem is manufacturing process. All secondary and conditioning processes, marked in the model as In, Out, Fixed Asset, Staff and Information-Technology, are also considered.

The "Out" block characterizes the result of the process, defining the requirements for other blocks. The block itself is defined by the manufactured products. The characteristics of manufactured products are defined, the most important being the output volume, market prices and deadlines. They make a considerable impact on the sustainability of the enterprise and define the requirements for other blocks.

The "In" block symbolizes the process of supplying raw materials, spares etc. It includes the processes of purchase, going on in supply departments. The stability of this block also has a serious impact on the sustainability of the enterprise.

The "Fixed asset" block depicts the enterprise as a single manufacturing mechanism. The modernisation of this mechanism defines the main development prospects of the enterprise.

The "Staff" block represents the staff of the enterprise. It can be subdivided into two categories – workers and other employees. The development and introduction of new systems of personnel motivation, staff recruitment and training all have considerable impact on the sustainability and development of the enterprise.

The "Information-Technology" block ties together all the processes within the enterprise based on main manufacturing technologies and management structure. This block defines the introduction of new manufacturing technologies as well as new management technologies. It suggests the optimization of the whole management system, the introduction of automated management and workflow system, the introduction of new information and infocommunication technologies, including cloud services.

The standard business unit model allows us to present the structure of a wide range of processes, from the working cycle of a machine to the functioning of a workshop or even a whole manufacturing enterprise. These structures can be linked into parallel and sequential chains and form large clusters.

The concept of *power* of such structure is introduced. It comprises all the possible results of its functioning cycle. It is defined in the space of the outputs of the unit and depends on the power of the unit's inputs. Mathematical modeling of power allows us to solve various problems dealing with the co-dependence of several such units and evaluate the required power for the units' inputs.

The standard business unit model allows us to describe not only manufacturing units, but all the departments within an enterprise.

III. INFOLOGICAL AND MATHEMATICAL MODELS OF THE ENTERPRISE

The analysis of information about an enterprise's book-keeping (including commercial papers, balance and account of loss and profit) is a sufficient way to describe its state and predict its functioning. The book-keeping system is mathematical model of an enterprise that has been known since the Middle Ages and is still being updated. It describes all aspects of the enterprise's functioning.

Structural model of a functioning enterprise based on accounting registers is depicted on Fig. 2. Principles of dynamic economico-mathematical modeling have been developed and expressed in a system of differential equations, algebraic equations and inequalities based on accounting registers [6].

Let's define the characteristics of this model:

- An enterprise is a manufacturing object. It operates as a business and produces goods.
- Manufacturing requires raw *material*, spare parts, energy bought from *suppliers* and stored in storing facilities.
- Manufacturing requires the *work* of employees receiving *salary*.
- Manufacturing requires equipment – *fixed assets*, subject to amortization and refilled through *capital investments*.
- Produced goods are stored as *ready product* and then *sold to customers* on the market.
- Money from selling goods (as well as *credits*) is remitted to the *settlement account*.
- Money from the settlement account is used for paying taxes, salaries, credit percentage, acquiring materials and other needs.

Thus, the money is converted to resources, closing the cycle of a manufacturing process.

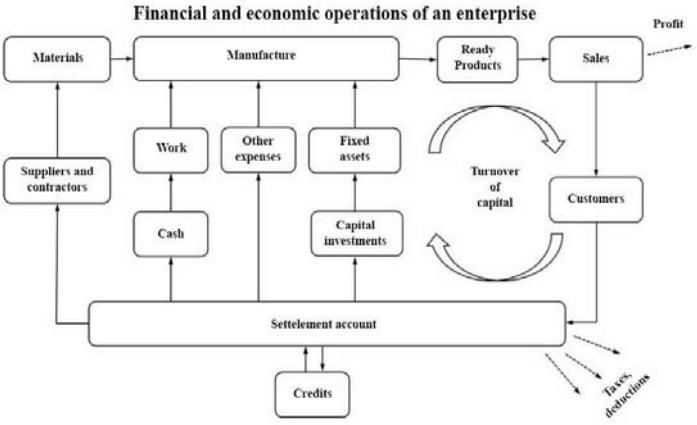


Fig. 2. Infological model of a functioning enterprise

The mathematical model of the enterprise is based on the model of an accounting book, having the following algebraic expression:

$$S(t) = S(t - \Delta t) + U_{in}(t) - U_{out}(t), \quad (1)$$

where:

U_{in} – incoming resource flow, debit turnover,

U_{out} – outgoing resource flow, credit turnover,

Δt – time span.

Each block of the model on Fig. 2 can be mathematically expressed in the way presented above. Thus the structural model presented on Fig. 2 is expressed with the system of algebraic equations. This is the discrete model of a functioning enterprise.

$\Delta t \rightarrow 0$ gives us a continuous model that can be presented as a set of algebraic and differential equations.

Some blocks within the model can, with certain assumptions, be described using retardation equation. For example, suppliers need a certain amount of time to deliver materials to the storing facility. This amount of time is often known beforehand. Thus we have a delay between ordering the materials and storing them. The suppliers often receive prepayment. This means that the delay starts from the moment of receiving this prepayment. Thus the flow of materials to the storing facilities $U_s(t)$ is linked with the flow of payments to the suppliers $U_p(t)$ via the following formula:

$$U_s(t) = U_p(t- \tau_p), \qquad (2)$$

where τ_p is the average time needed to receive the materials at the storage.

With certain simplification this equation can be substituted for inertial retardation equation. “Suppliers and contractors” and “Manufacture” blocks are described by the same type of equations with retarded argument.

IV. Modeling the Sustainable Development of an Enterprise

The infological, structural and dynamic mathematical models allow us to predict the development of an enterprise as a multiproduct manufacture. As part of the problem, let's suggest that initially the enterprises produces only one type of products. Let's suggest it's a considerably reliable business with low risks and low profitability. Let's also suggest the enterprise management is considering the introduction of a new type of product, i.e. new type of business, presumably a profitable one. However, it's connected with considerable risks, i.e. low sustainability.

The models we have developed allow us to build a dynamic model of the enterprise, manufacturing both types of products, and predict the results. The structural and infological models allow us to present the risks of each business as systems consisting of separate blocks. The sustainability of each block in the enterprise's system is determined by the risks imposed on it by each new type of business. Thus the introduction of a new risky business helps the enterprise develop and gain profit, while old businesses smooth away probable risks and assure the overall reliability and sustainability of the enterprise. The principles of sustainable development of an enterprise and dynamic models developed by our researchers allow us to predict the impact of risks of introducing a new business on the overall sustainability of an enterprise.

The models introduced in this paper have become the basis for an info-analytical system allowing to predict the functioning and development of a manufacturing enterprise and search for new opportunities to successfully manage and expand business, assuring its sustainable development.

Acknowledgment

The work is performed according to the Russian Government Program of Competitive Growth of the Kazan Federal University (KFU).

References

[1] Documents of the conference in Rio de Janeiro. Center "For our collective future". Geneve, 1993.

[2] New paradigm of the development of Russia (complex research of sustainable development). Ed. By V. A. Koptyug, V. M. Matrosov, V. K. Levashov. Second edition. “Academy”. Irkutsk, 2000. – p. 460.

[3] T. K. Sirazetdinov. Dynamic modeling of economic objects. FEN, 1996.

[4] T. K. Sirazetdinov, V. V. Rodionov, R. T. Sirazetdinov. Dynamic modeling of the economics of a region. FEN Tatarstan Academy of Sciences, 2005.

[5] R. T. Sirazetdinov, A. A. Brazhkina. Universal structural model of a basic economic cluster // Management of large systems, i. 29. IPU RAS, 2010. P. 152-166.

[6] R. T. Sirazetdinov, I. A. Enikeev. Strucutural scheme of dynamic economico-mathematical model of an enterprise based on accounting plans // Society, state, person: problems of interaction within market economics. Kazan, 2006. P. 224-226

[7] T. K. Sirazetdinov. Dynamic model of prediction and optimal management of an economical object. // Izvestiya vuzov. Aviation Technology, 1972. №4. P. 32-38.

[8] T. K. Sirazetdinov, S. K. Dzhaksybaev. Dynamic model of interacting economical objects. // Izvestiya vuzov. Aviation Technology, 1974. №1. P. 25-30.

[9] T. K. Sirazetdinov. Complex systems and the problem of analytical projecting I, II // Izvestiya vuzov. Aviation Technology, 1980. №4. P. 59-64. 1981. №2. P. 51-55.

[10] T. K. Sirazetdinov. Methods of solving multi-criterial problems of technical systems synthesis. M. Mashinostroenie. 1988. P. 160.

Hotspot Information of Public Opinion in E-Government

Ramiz Aliguliyev
Institute of Information Technology of ANAS
Baku, Azerbaijan
r.aliguliyev@gmail.com

Gunay Niftaliyeva
Institute of Information Technology of ANAS
Baku, Azerbaijan
gunayniftali@gmail.com

***Abstract*—The determination of public concern degree issues in e-government can help government agencies to improve the quality of services. For this purpose an approach for calculation of hotspot information, determination of public interests of regions with helping web log mining and text mining technologies is given in this paper.**

Keywords—e-government; hotspot information; web log mining; text mining

I. Introduction

We live in the Information and Communication Technologies' (ICT) century. Advanced technology products have become an integral part of our lives. The development of ICT is the one of the main factors to influence various organizations, companies, state bodies' activity; efficiency, better delivery of government services to citizens; and citizen participation for decision-making; and more efficient government management. ICT helps government agencies to locate service centers closed to clients. The involvement of ICT into community management process has resulted in the emergence of the concept of e-government. E-government - involves increasing efficiency and performance of government agencies and the facilitation access to e-services through the wide application of information technologies. In general, the intent of e-government is to provide new opportunities and benefits to citizens, businesses, and governments around the world [1]. Its purpose is to create citizen-centered openness in government [2] and to serve as a necessary force for better governance [3]. In other words, e-government is used to improve the effective and efficient delivery of government information and resources to citizens, more efficient government, less corruption and more transparency. E-government is not meant only for introducing or using technological tools, it's aim to change mindset and work culture to integrate government processes and functions, to serve the citizens better services.

E-government system has become one of the most important channels for public service delivery and for citizen-government interaction. This system introduces rich information resources to citizens. These resources analysis can help the government to make right decisions and improve services. Every day various information resources are shared in e-government system. It is important to discover public concern degree of this information. This information is called hotspot information. In this paper is given some information about hotspot information, their advantages, an approach for detection hotspot information on regional level and the destinations for future plans.

II. Hotspot Information in E-Government

The increasingly wide application of the Internet and the Web's led to the potential development of state-citizen relations - expansion of the use of ICT in collaboration with the state and the citizens, businesses and other government agencies. Web 2.0 technologies have opened a new stage in the relationship between citizens and the public sector. The development of ICT and information society have created a variety of potential opportunities for the public sector: transporting information over the Internet; the expansion of relations between state and the citizens, businesses and other government agencies; conducting operations and records; the development of management and democracy.

At present, everyone can use Internet, search anything and get information easily. But Internet public opinion system is limited by the key words given by the users [4]. Influenced by some subjective factors such as knowledge, information source, and concerns of the user, the system will not detect those unexpected events [5]. Therefore, computers can used for sorting out news, finding hot topic keywords automatically; updating common safety word frequency in library timely, tracking sudden events timely [6]. If the real-time hotspot information in a period is to be obtained and the internet hotspot topics in current society are to be periodically discovered, automatic solutions are becoming a valuable research orientation. When dealing with mass web pages, it is difficult to cluster the expected hotspot and clustering causes huge central bias. Therefore, the algorithms need to be improved [4, 7]. Text clustering technologies can used to improve of hotspot tracking algorithm. Other problem is to master mass of data effectively and extract hotspot information. An extensive application prospect for solving this problem is given in [8] as follows:

- It is an important means for individuals to instantly and conveniently obtain the hotspot information in current society;

- With helping it enterprises can master the most advanced information and hot technology in their fields, increase their competitiveness for enterprises;
- It can provide important clues for relevant departments of the governments to instantly know about the direction of public opinions in current society, be conductive to the governments to analyze and guide the public opinions, actively guide the healthy development of internet public opinions;
- Help the governments to grasp the problems mostly cared by the people in each period as well as the viewpoints and attitudes on these problems, so as to make scientific and correct decision, keep the society stable, and truly reach the aim that the Internet serves for the society and the people.

Public opinion information analysis is the most important part, which includes hot topic mining on the acquired news, comments and other information according to the hot topic analysis model; and warning analysis, text clustering, topic mergence etc. Public opinion information analysis can help government agencies to improve their services, increase citizens satisfactions. There are numerous and rich information resources in e-government system, which was presented by the sites of government agencies. Make the right decisions and to achieve some aims it is important to identify quickly and in a timely the hot issues of public interest on this information. It is one of the main issues of e-government. The number times a user clicks on some information and the average residence time can identify hot issues in e-government system [9]. Our goal is to discover public concern degree of regions and their interests in e-government system. Which information is very useful for them, what their problems etc. There will be mass of hotspot information in e-government and it is difficult to analyze all of them. The analysis of hotspot information can help government agencies decrease their time to analyze this information. In this paper is proposed approach to define hotspot information in e-government. The main step of this approach is given as follows:

1. Classification of government agencies (number of clicks on some information, the average residence time);
2. Determination of hotspot information;
3. Clustering regions based on their public interest;
4. Assessment of regions based on public opinion.

At first hotspot information of public opinion is calculated through the web log mining. All users' browsing information is stored in web logs: IP address, access time, the URL etc. Number of public clicks on some information and the average residence time the public spend on some information can extend hotspot information. Then a regional assessment is conducted to discover which regions and how many of them are clustering around hotspot information. It can help governments to define the main problems of regions.

Second the contents of hotspot groups are analyzed through the text mining technologies. These technologies can extend abstract of these information. With helping it government can introduce better services to citizens, decrease public concern, provide more targeted services and information timely. In our future work we will give more information and method about that.

CONCLUSION

At present one of the most important issues in e-government is to involve citizens to decision-making process and to increase citizen's satisfaction. For this purpose, researches carried out a lot of works. But the problem has not been solved completely yet. It needs to develop effective algorithms to define public opinion hotspot information classification. In this paper is given a little information about the approach for detection of hotspot information and assessment of regions with helping it. Our future work will add more information about this approach and we will work on in practice.

REFERENCES

[1] P. T. Jaeger, K. M. Thompson, "E-government around the world: Lessons, challenges, and future directions", Government Information Quarterly, vol. 20, no. 4, pp. 389–394, 2003.

[2] E. D. Daniels, Memoranda 01–28 (Citizen-centered e-government: Developing the action plan), 2001. Retrieved from. http://www.whitehouse.gov/omb/memoranda_m01-28

[3] M. P. Gupta, D. Jana, "E-government evaluation: A framework and case study", Government Information Quarterly, vol. 20, no.4, pp. 365–387, 2003.

[4] Sh. Zhang, Z. Liu, "Hotspot Detection Method of Internet Public Opinion based on Key Words Extraction", International Journal of Digital Content Technology and its Applications, vol.6, no.16, pp. 340-347, 2012

[5] D. Jie, X. Jungang, "IPOMS; an Internet Public Opinion Monitoring System", International Conference on the Applications of Digital Information and Web Technologies, pp.433-437, 2009.

[6] Z. Kui, SH. Xueming, Y. Hongyong, "Hot Spot Information Auto-detection Method of Internet Public Opinion", Computer Engineering, vol.36, no.3, pp.4-6, 2010.

[7] L. Hong, L. Xiaojun, "Internet Public Opinion Hotspot Detection Research Based on K-means Algorithm", International Conference on Swarm Intelligence (ICSI 2010), pp. 594-602, 2010.

[8] G. Wang, "Research on Hotspot Discovery in Internet Public Opinions Based on Improved K-Means", Computational Intelligence and Neuroscience, vol. 2013, pp.1-6, 2013.

[9] S. Wang, J. Zhang, F. Yang and J. Ye, "Research on Cluster Analysis Method of E-government Public Hotspot Information Based on Web Log Analysis", Journal of Computing and Information Technology, vol. 22, pp. 11-19, 2013.

Technical and Communicational Standards of E-Invoicing

A Country Example: Turkey

Ilknur Gur Nalcaci/ R&D Project Manager
Department of Research and Development
Idea Technology Solutions Inc.
Istanbul, Turkey
ilknur.gur@ideateknoloji.com.tr

***Abstract*- 21st century is completely a new era in which security and data roaming are on the upfront and borders are ceased globally by standardization. In the process of moving paper to electronic media, all these components should be taken in consider. By this way the process can be operated effectively.**

Generation and communicational standards of electronic invoice process and its components are examined in this paper. And also all components' definition and usage are explained over Turkey e-invoicing process as a country example.

***Index Terms*- UBL, XML, HTTPS, XAdES, e-Invoice, e-Signature**

I. Introduction

Main goal of e-invoicing is providing interoperability between different business processes. As known e-Invoice is interdisciplinary process and that is related with document process, communicational process, and e-Signing process. Full interoperability requires interoperating on three different levels: in terms of the content of the invoice in other word semantics, the used format and language named syntax, and the method of transmission. Semantic interoperability means including information that is required for the electronic invoice and certain meaning of invoice fields are understood in an u contains a certain amount of required information, and that the precise meaning of the exchanged information is unambiguous manner. Syntactic interoperability referred that the data elements of an electronic invoice are presented in a format which can be exchanged between the supplier and receiver. [1]

The Internet together with standardizes exchange protocols serves as an infrastructure layer. By the development of open standards like XML (Extensible Markup Language) at end of the 1990s, usage of electronic document exchange is raised into all sectors, because of lower cost of implementation and maintenance. So XML known as a milestone of e-document exchange. Then UBL (Universal Business Language) is developed by OASIS (Advancing Open Standards Information Society) based on XML technologies and approved by UN/CEFACT as an international standard.

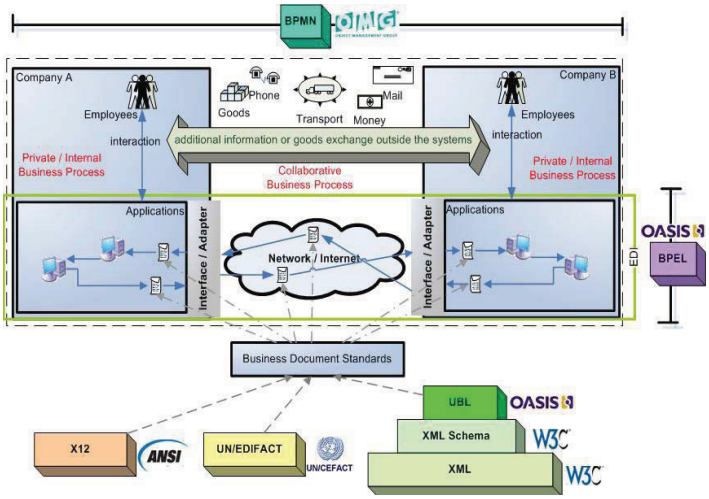


Fig. 1-Top level view showing different Business Process types, their relationships and related standards

Another subject for interoperability is transmission of the messages/document. Interoperation between different communications systems is also provided with communicational standard protocols known as OSI, which's specifications were determined by two international standardization organizations: The International Organization for Standardization (ISO) and the International Telecommunication Union-Telecommunications Standards Sector (ITU-T). This standard is used for

II. Creation Standards of E-Invoice File

For achieving in effective business process management, developing standards and specifying the semantics of messages exchanged between organizations have an important role. [3]

Many International and European organizations like ISO, OASIS, UN/CEFACT, GS1, CEN and international interoperability projects like e-PRIOR, PEPPOL, have dealt with e-invoicing matters from their respective duties and perspectives.

By enhancing coordination with regard to determined standardization activities, given interoperability adoption scenarios, could be instrumental in removing outstanding

obstacles and facilitating the deployment of e-invoicing in Europe, if carried out as part of a holistic approach to e-invoicing.

Legally accepted electronic invoice format by European Union Member States, must be provided that the integrity of their content and the authenticity of the origin are assured by one of the following methods [4]

- EDI
- Electronic Signature
- Other electronic means,

subject to acceptance by the Member States concerned.

To comply with international standards, Turkey e-invoice creation standards is examined according to these methods defined above.

A. Creation Standards of e-Invoice in Turkey

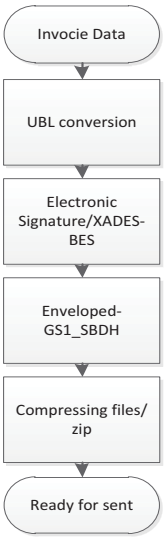


Fig. 2-Creation Workflow of e-Invoice in Turkey

Data of invoices is taken into the system, and converted to the UBL format, and added XSLT for visualization then signing with using advanced e-signature format named XAdES. E-Signed UBL invoices enveloped according to GS1 standards SBDH format. As a next step enveloped files are compressed by using Zip format for reducing file size. After generation process invoice is ready for exchanging/sending via web service.

As seen on Fig.1 international standards are used in every phase of creation e-Invoices. Explanation of these standards given below.

1) e-Invoice File Format: UBL-TR

Adopted as official nationwide e-invoice format is UBL-TR which is Turkey customized format of UBL. UBL is the Universal Business Language, defines open source library of standard electronic XML business documents and messages such as invoices, purchase orders, application response. UBL document standards developed by OASIS Technical Committee with participation from a variety of industry data standards organizations and delegates of diverse industry.

Information related to invoices are modelled in the UBL schema based upon the Semantics given by Fig.3. [5]

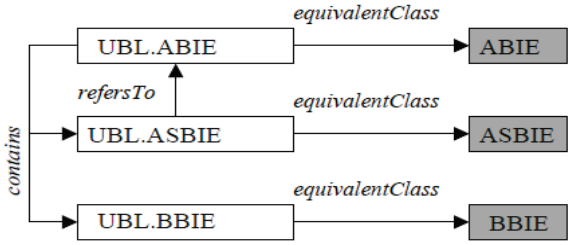


Fig. 3-UBL Schema Structure of Information Entitiy

BBIE: Basic Business Information
ABIE: Aggregated Business Information Entity
ASBIE: Associated Business Information Entity

To express the information hierarchy, ASBIE consist of ABIE, and ABIE consist of BBIE.

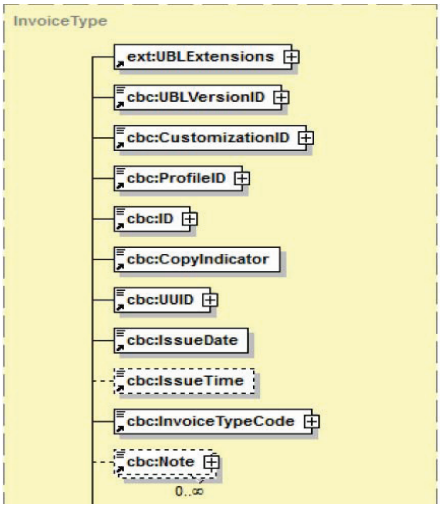


Fig. 4-UBL-TR e-Invoice Schema Staging

A part of e-Invoice fields is given at Fig.4. By the schema notations schema can be explicated; for example, solid line indicates mandatory fields and dotted line indicates facultative fields. Plus, means that element is an Aggregated Business Information Entity and contains Basic Business Information.

Numeric staging on schema means that element cardinality. For example, 0...∞ means <Note> element can be multiplied many times. Despite that "UBLVersionID" is mandatory field and it cannot be multiplied.

Used method for Customization of UBL standard to UBL-TR

- Mandatory fields of e-invoice UBL schema are applied as mandatory for UBL-TR schema. So this provides compliance between UBL and UBL-TR.
- Some of facultative fields of UBL schema which is relevant to Turkey requirement, are used in UBL-TR schema, the others are removed.
- Cardinality value of some elements are restricted
- International Code lists are set for elements that is needed to be expressed by a certain codes. (UN/EDIFACT 44611 Code List for payment means, etc.)

[1] http://www.unece.org/trade/untdid/d00a/tred/tred4461.htm

2) e-Invoice Signing Standard: XAdES-BES

In invoicing process for providing authenticiy of source, integrity of content, legibility of document, electronic signature is used. And it is determined as an e-invoice accepted method in Directive 2006/112/EC.

Electronic signature creation and verification standards given by ETSI [6] and signing process given by Fig.5 with general outlines. In addition to this definition used advanced e-signature in Turkey e-invoicing process named fiscal seal confirms the identity of company.

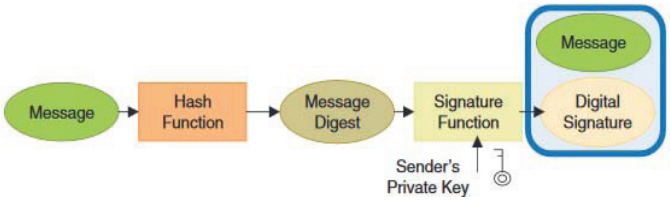


Fig. 5-Creating a Digital Signature

In Turkey used electronic signature form is a XAdES; XML Advanced Electronic Signature. XAdES used for signing XML documents. As mentioned on previous part of this article Turkey e-invoice format is UBL which is based on XML. So XAdES signing standard is appropriate for Turkey e-invoicing process. There are many different XAdES signature forms like XAdES-BES (Basic Electronic Signature), XAdES-EPES (the Explicit Policy based Electronic Signature) and XAdES-T and XAdES-C (the Electronic Signature with Validation Data)[11]

In Turkey used XAdES signature format is XAdES-BES, fields of this signing format given by Fig.6.

Fig. 6-XAdES-BES Tags

The XAdES-BES fulfilled defined legal requirements at European Directive on electronic signatures.

3) e-Invoice Enveloping Standard: GS1_SBDH

UBL formatted and e-signed electronic invoices covered with envelope which is based SBDH (Standard Business Document Header) with GS1[2] XML. By enveloping, too many invoices belong to same sender and same receiver can be sent at one time.

SBDH standard provides standard for document, message, data elements located, presentation and guidance between different applications. Software vendors can develop functionality in their applications that rely on the existence and location of the essential data used to manage the routing of business documents and also systematically determine the documents processing requirements. [7]

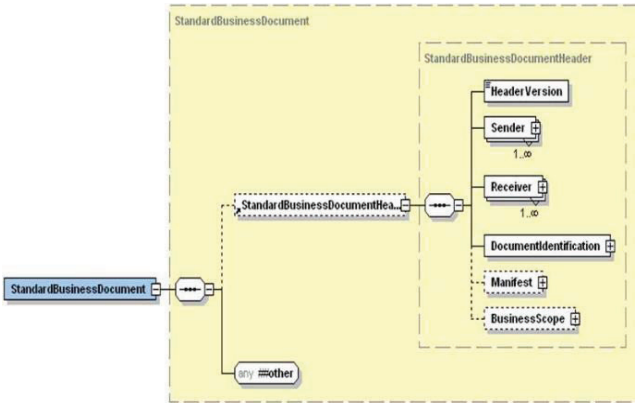


Fig. 7-SBDH Schema Staging

As it is shown at Fig.7 SBDH consist of four main fields. These are;

- "Header Version" field used for tagging used SBDH_GS1 envelope version like 1.0, 1.1, etc.
- "Sender" field used for indicating the Supplier Party of e-invoices in the envelope.
- "Received" field used for indicating the Receiver Party of e-invoices in the envelope.
- "Document Identification" field used for identify the document type which is carried in the envelope, like invoice, application response etc.

4) e-Invoice File Compression Format: ZIP

It is possible to reduce file size by using file compression formats. ZIP is a kind of this format. By ZIP format it is possible to compress one or more files without any data loss.

Prepared invoice contained envelopes compressed with .ZIP format for sending. Other formats that are used for compression is not accepted or defined in the system. Reducing file size also provides fast transaction following archiving advantages.

III. Communicational Standards of E-Invoicing

Transferring electronic data from one medium to another is carried out according to the connection and communication standards as OS. These protocols are provide equipment interoperability on data networking. OSI model consist of different layers as application, data link, physical, network, transport.

[2] Non Profit Global Standardization Company

Layer	Function	Used Standards
Application	Supporting network applications	FTP, SMTP, HTTP, **HTTPS**
Transport	Process-process data transfer	NBF, **TCP**, UDP
Network	Routing of datagrams from source to destination	AppleTalk, ICMP, IPsec, **IPv4**, IPv6
Data Link	Data transfer between neighboring network elements	ATM, MPLS
Physical	Bits "on the wire"	DOCSIS, Ethernet, DSL

Fig.8-OSI Layers' Functions and Standards [9]

Main functions of OSI layers are listed at Fig.8. General data transfer and communication is provided according to these standards on these layers.

A. Transportation Standards and Process of e-Invoicing in Turkey

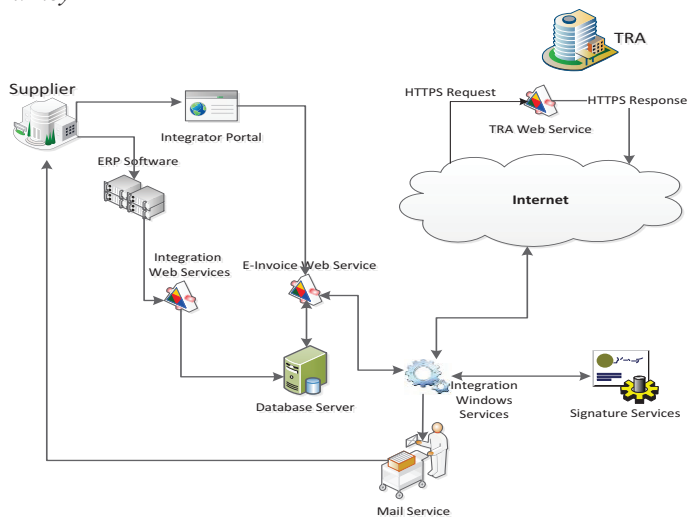


Fig. 8- Communicational Architecture of e-Invoicing System in Turkey [10]

1) E-Invoicing Transportation Standards

When invoice is issue to the transferring, safety and traceability comes to the fore.

HTTPS is used on application layer for secure transferring of e-invoice between parties over a network. HTTPS is the he most common protocol standard at application layer for secure data transfer. Hypertext Transfer Protocol (HTTP) is a protocol intended for quick-access, distributed, collaborative, hypermedia systems. This is the standard protocol of the Wide Web. HTTPS or HTTP over SSL (Secure Socket Layer) is a variant of HTTP which provides enhanced security. [8]

TCP is used at transport layer for traceability of transferring e-invoice document and also provides ordered, reliable and error-checked transport of data between applications running on hosts communicating over an IP network.

IPv4 standard at network layer is used for definition of invoice user on the TRA system. IPv4 is a connection-free protocol for use on packet-switched networks.

2) Mangement Process of e-Invoicing in Turkey

E-Invoicing system in Turkey, is managed by Turkish Revenue Administration (TRA). Taxpayers who want to send e-invoice, must be registered to the TRA system by IP and e-invoice application's web service end point as an URN[3] format. Only registered taxpayer can benefit from e-invoicing system in Turkey. Sender party prepared e-invoices in UBL format, signed with XAdES certificate, enveloped and zipped then sends them to TRA by using HTTPS specified web service. TRA transfers to the e-invoices to the Receiver's registered web service end point. In every part of the process is monitored by logging mechanism.

IV. Conclusion

By the development of trade in the world, movement of commercial documents and security of them has become one of the important issues. With the development of technology exchange of commercial document like orders, invoices, delivery notes, payment documents are moved to the electronic environment. A document which moved from paper to electronic media and the process will be subject to different operations and standards. Standards are one of the most important components enabling interoperability.

In this paper, we describe the technical components of e-invoicing system over a country example Turkey.

As it seen in the Turkey e-invoicing example international standards are using at all steps. Thus interoperability can be achieved both national and international level. Integration of interoperability and business process allows the companies save time and money, following accelerating the process as a whole. For now, using e-invoice is mandatory for some critical sectors, but in the future, companies in Turkey will use the system voluntarily because of interoperability advantages.

References

[1] Directive 2014/55/EU of The European Parliament and Of The Council, 16 April 2014J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] K.Tolle, "UBL: The DNA of Next Generation E-Business" Information Technology And Control, vol.37, 2008

[3] e-Invoicing Standardisation, European Commission, Sep. 2012

[4] Council Directive 2006/112/EC, The Council of The European Union, 28.11.2006, Section 5

[5] Y. Kabak "Semantic Interoperability of The UN/CEFACT CCTS Based Electronic Business Document Standards", Middle East Technical University, July 2009

[6] Directive 1999/93/EC of The European Parliament and Of The Council, 13 December 1999

[7] Standard Business Document Header (SBDH) Version 1.3 Technical Implementation Guide, Issue 1, July-2007

[8] J. Crichton, S. Shao, J. Staten, "Method and apparatus for tunneling TCP/IP over HTTP and HTTPS", USPTO, May 2002

[9] W. Stallings, *"Data and Computer Communications"*, Pearson-Prentice Hall, 9th Edition, 2010

[10] A. Ozcan, S. Eryigit, "e-Invoicing In Turkey And A Software-As-A-Service Solution", ICENS 2015

[11] ETSI, TS 101 903 V1.4.1, XADES Technical Specification, June 2009

[3] URN: Uniform Resource Name: Enables interaction with representations of the web resource over a network

The prediction of the chance of selling of houses as the factor of financial stability

Ceyhun Abbasov
Department of World Economy
Qafqaz University
Baku, Azerbaijan
ceyhunabbasoglu@gmail.com

***Abstract* - This paper has devoted to the estimating of the selling chance of houses in market. It is a clear that housing market is a very important for the financial stability. The database of our approach consist of the price of the houses and their quality indicators as the repair level, distance from the city centre, projects of the building, the place of the flat in a building and so on. In the base of this database, we calculated G score and found that if the computed G value is equal to 0.39 and greater than that one it is classified as the high chance of selling the house, but in the case of being less than 0.39 it is classified as none selling chance. We hope that this calculation can be useful for the banks. Because, they always face the question about what is a probability of selling of houses encumbered under loan agreement in banks. So, banks are able to estimate the probability of the realization of mortgaged houses.**

***Index Terms* - Selling chance of houses, I and II type errors, Hedonic price method, Edmister's approach, Kappa test.**

I. Introduction

The bank system is very sensitive to housing market. There are kind of loans provided under encumbrance of houses. It raises an important question in the process of lending: What is the probability of selling of houses encumbered under loan agreement. If the granted loan is not repaid by the borrower, then the bank is going to sell the same house charged against the case of non-payment. In this context it must be very interesting to the bank to know the probability of selling the house before lending money. The bank needs to have the statistical data base for assessing the probabilities. By the standpoint of the more efficiency the qualitative properties of houses should be included in this data base. From the end of the previous century till now the house market has been developing in our country, especially in Baku city. The development of house market is explained by several factors: accelerated urbanization process, increasing investment in houses, real estate direction of funds earned abroad by residents etc. As heterogeneous goods a house has three major peculiarities unlike other goods. First, it includes many other goods and meets different requirements of the family. Second, the house is an immovable. That is why the location of house is playing very important role in forming its price. Third, the demand in the house market is more stable than in the market of other goods, which could be explained by the fact that house cannot be substituted by other goods. First section of working paper has been devoted to form the quality statistical data base. The second section describes the methodology of our approach and the related literatures. Third section is about the empirical result of the statistical analysis. Also this paper has conclusion and reference sections.

II. Statistical data base

The qualitative statistics of the housing market is understood as the information about the properties of the houses. For example, the repair level, the distance from the city centre, the level ratio of the floor in the building, gas in a building, metro station etc. can be introduced as the qualitative statistic of the housing. Table 1 describes this data base. This database consists of about 200 variables. These variables are qualitative properties of the houses. Some of them have been described in table 1. In total there are 1700 observations in this database. We collected these data from the housing market in Baku, Azerbaijan. There are only 6 variables have been used in our statistical analysis. These are: to be furnished with gas, the distance from the city centre, the price of per sq m., to be near the metro station, availability of the documents of building and to be without repair. We have taken a log of the distance between the city centre and a log of the price of per sq. m. The first column of the table 1 describes the score for the quality statistical data base. This score is coded as the 0 when the house hasn't been sold during the period, otherwise is coded as 1.

III. Methodology and related literatures

Our approach consists of two steps. In the first step we have defined the independent variables of our model. In the literatures the characteristics of houses have introduced as the creator of their prices. Usually this method called hedonic price method. Hedonic price method assists in defining the influence of these properties on the price of houses. It is first cornerstone of our approach. (J)Ceyhun Abbasov (2014) had investigated the influence of these properties of houses on their price in Baku city. In this work we changed the structure of the independent variables of our model. In literature there are many research works which have been devoted to investigate the relationship between the price of houses and their

qualitative properties. For example, Ercan Baldemir (2007) noted in his work that first time Haas (1922) investigated the relationship between the price of house and its properties like "far from the city centre" and "scale of the city". Court (1941) analyzed the price of automobiles as a function of their characteristics in his research work and had defined the marginal cost of each characteristic. In this work He explained that the marginal cost of the characteristics generates the prices of automobiles. Robert J. Hill (2011), Erwin Diewert, Saeed Heravi, and Mick Silver (2007) etc. have also investigated the price of houses as a function on their characteristic properties. We also will use this major methodology. But the price of houses isn't a dependent variable in our analysis. It will be independent variable as the price of per sq. m. together other qualitative and quantitative variables. So we are able to take the qualitative properties of the houses as independent variables for statistical analyses. For to do this we refer to some empirical research works. So, the independent variables of our approach will be to be furnished with gas, the distance from the city centre, the price of per sq m., to be near the metro station, availability of the documents of building and to be without repair.

TABLE 1: A SAMPLE TO THE QUALITY STATISTICAL DATA BASE OF HOUSES

Score	Price of per sq. m (AZN)	Without repair	Parquet	To be near to metro station	Distance from the centre[1]	Level ratio of the floor[2]	Availability of the documents of building[3]	Furnished with gas
0	700.0	1	0	0	11.9	0.76	0	0
1	700.00	0	0	1	11.8	0.19	0	0
0	1192.6	0	0	1	11.8	0.88	0	0
1	1099.6	1	0	0	9.0	0.65	1	1
0	1173.9	0	1	0	17.2	0.15	0	1
1	507.63	1	0	0	13.2	0.33	0	0
1	855.27	0	0	0	12.4	0.21	1	0
0	776.39	1	0	0	12.4	0.79	0	0
0	1629.1	0	0	1	6.4	0.63	0	0
0	1300.0	0	1	0	3.4	0.38	0	0

The second cornerstone of our approach is to take the score which has been given in the first column of the table 1 as the dependent variable of our approach. In this step we refer to Edmister's (1972) approach. Edmister (1972) had used 7 financial ratios in his insolvency model. While establishing z-score values as a dependent variable in the model he coded 0 for those companies which considered as default case, otherwise it was coded as 1. In accordance with this model the companies are classified to be not default case if z-score of these ones equal to 0.530 and more. Whereas the cases less than 0.530 it is expected to exposure the default. So, the dependent variable of our approach will be the G-score (see the first column of the table 1) which is coded as the 0 when the house hasn't been sold during the period, otherwise is coded as 1.

These two approaches let us to construct our model as the following:

$$G = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

Where,
β = the parameters of the model
X_1 = to be furnished with gas
X_2 = log of the distance from the city centre
X_3 = log of the price of per sq. m.
X_4 = to be near the metro station
X_5 = availability of the documents of building
X_6 = to be without repair

G = the dependent variable of our model will be the score which is coded as the 0 when the house hasn't been sold during the period, otherwise is coded as 1

IV. EMPIRICAL RESULTS

We have 1700 observations about the price and characteristic properties of the houses. But we used 234 observations for estimating the equation (1) because to take only 3-rooms houses in new buildings which have been built after the Soviet Union time.

G-score is the dependent variable which is coded as the 0 when the house hasn't been sold during the period, otherwise is coded as 1. The results of the estimating of equation (1) have been given in table 2.

TABLE 2: THE ESTIMATING OF THE EQUATION (1)

Independent variables	Coefficients
to be furnished with gas[4]	0.18880
ln_the distance between the city centre[5]	-0.12051
ln_the price of per kv. m.[6]	-0.72534
to be near the metro station[7]	0.13618
to be documents in a building[8]	0.43440
to be without repair[9]	-0.13691
C[10]	5.61938

[1] This distance was measured by the "Google maps". We took "Icheri Sheher" metro station as the starting point in "Google maps".
[2] These ratios present the level of the flats. For example, suppose that the first flat (house) in table 1 have been observed in the 13th floor of the 17 floors building then we can calculate this ratio as 13/17=0.76. So, if the ratio is equal to 0.5 or is close to 0.5 (mining that 0.4, 0.6) it means that the flat is on the middle of the building. If the ratio is close to 1 it means that the flat is on the top of the building. If the ratio is close to 0 meaning that the flat is on the lower floors of a building.
[3] As the mentioned above all houses which have been observed are located in new buildings and some these ones don't have the important official documents which affirm the private ownership.
[4] The coefficient is significant at the 0.02 confidence level
[5] The coefficient is significant at the more than 0.01 confidence level
[6] See foot-note 5
[7] The coefficient is significant at the 0.03 confidence level
[8] See foot-note 5
[9] The coefficient is significant at the 0.1 confidence level
[10] See foot-note 5

So, if the computed G value found by using the equation (1) based on the results of the table 2 is equal to 0.39 and greater than that one it is classified as the high chance of selling the house, but in the case of being less than 0.39 it is classified as none selling chance. Note that 5 of 93 houses those had remained in none selling status, classified as the high chance of selling. It means that we have I type error. But 14 of 141 houses those had sold, classified as none selling chance. It means that we have II type of error (see the table 3).

TABLE 3: CLASSIFICATION ANALYSIS

	Selling	*Non-selling*	***Total***
Selling	127	14	**141**
Non-selling	5	88	**93**
Total	**132**	**102**	**234**
Selling %	90.1	9.9	100,0
Non-selling %	5.4	94.6	100,0

In statistical analysis I type error always is more dangerous than II type error. Because classification of houses as none selling chance house isn`t bad for the market participants. But of course in that case they may face to lose its good dealing. Here one question is it possible to make forecast the selling case in the term of existing I type error and II type error? In order to reply this question, it is necessary to define the possibility of making I type error and II type error.

I Type error = 5/234*100 = 2.14%

II Type error= 14/234*100 = 5.98%

Total occurred error = 2.14%+ 5.98% = 8.12%

We can say that the forecasting will be correct with 91.88 % probability. But with 8.12 % probability it will be resulted in error. Actually the predictability power is seen very highly (91.88%). But we are in need of to check it. For checking the predictability power Cohen's Kappa test statistics is useful. The suggested Cohen's Kappa test statistics is as follows:

The results of the table 3 will be used for this test statistics

$$k = \frac{w_1 - w_2}{N - w_2}$$

Where,

$w_1 = 127 + 88 = 215$

$w_2 = \left\{\frac{127+5}{234} * \frac{127+14}{234} * 234\right\} + \left\{\frac{88+5}{234} * \frac{88+14}{234} * 234\right\} = 79.58 + 40.5 = 120.1$

So,

$k = \frac{215 - 120.1}{234 - 120.1} = 0.833$

There are 6 significance intervals for the decision about the result of kappa test. These intervals are as follows:

$k \leq 0$ (Very small compliance)
$0 < k \leq 0.2$ (Small compliance)
$0.2 < k \leq 0.4$ (Acceptable compliance)
$0.4 < k \leq 0.6$ (Medium strong compliance)
$0.6 < k \leq 0.8$ (Significant compliance)
$0.81 < k$ (Almost perfect compliance)

In order to use of this classification in forecasting none selling chance and the high chance of selling of houses, the result of test should be more than 0.81. Otherwise there is no any significance to use of this classification in making the prognosis. Our estimating value (0.83) is greater than critical value of the kappa test statistics (0.81). It means that we have obtained perfect compliance. So, these classifications can be used in forecasting none selling chance or the high chance of selling of houses.

V. CONCLUSION

Finally, it would be usefully that the main points of the research are reviewed. The first point is about the qualitative statistical data base. This data base is a universal data base. The average price for the all properties of houses can be calculated by using of this data base. For example, you can choose any property of house say that like "to be documents in a building" and can define average price, max or min prices, number of houses which were offered to a market etc. over this property. Secondly this data base can be used for the different estimating that ours is one of them. We constructed the classification for the housing market. This classification lets us to forecast none selling chance or the chance of selling of houses. Thus we found that if G-score (which is calculated by using the equation (1) based on the results of the table 2) of any house is equal or greater than 0.39 it means that this house has the high chance of selling. But if G-score (which is calculated by using the equation (1) based on the results of the table 2) of any house is less than 0.39 than it is classified as none-selling chance house. This result can be used both the market participants and the banks. The market participants always want to know that how much selling chance their houses have. On the other hand, it is interesting the banks to know how much selling chance the houses which have been taken as the collateral by the banks if any borrower unable to give back the money. The latter one is very important for maintaining the financial stability.

REFERENCES

[1] A.T.COURT. (n.d.) "Hedonic Price Indexes with Automotive Examples" http://www.econ.wayne.edu/agoodman/research/PUBS/Court_Hedonic.pdf.

[2] Aasen M. R. (2011) Applying Altman's Z-Score to the Financial Crisis an Empirical Study of Financial Distress on Oslo Stock Exchange.

[3] Ercan Baldemir, C.Yenal Kesbich, Mustafa İnci (24-25 May 2007). Estimating hedonic demand parameters in real estate market: the case of Mugla. IIIV Turkey Econometrics and Statistics Conference. Malatya: İnonu University.

[4] Erwin Diewert, Saeed Heravi, and Mick Silver (2007). Hedonic Imputation versus Time Dummy Hedonic Indexes. IMF Working Paper- WP/07/234.

[5] Edward I. Altman, Anthony Saunders (1998). Credit risk measurement: Developments over the last 20 years. Journal of Banking & Finance

[6] Grice, J. & (2001). Tests of the generalizability of Altman's bankruptcy prediction model. Journal of Business Research, 54(1), 53-61.

[7] Hill R (2011), "Hedonic Price Indexes for Housing", OECD Statistics Working Papers, 2011/01, OECD Publishing. http://dx.doi.org/10.1787/5kghzxpt6g6f-en

[8] Robert O. Edmister (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. The Journal of Financial and Quantitative Analysis, Vol. 7, No. 2 Supplement: Outlook for the Securities Industry (Mar., 1972), pp. 1477-1493

[9] (J)Ceyhun Abbasov (2014). Hedonic price method for housing. International Journal of Business and Social Research (IJBSR), Vol 4, No 8, pp. 8-16.

Using Web-Based Geoinformation Technologies within Micro Geo-Marketing Researches

Milikhin M.M., Gritsenko Yu.B., Ekhlakov Yu.P., Zhukovsky O.I., Senchenko P.V.
Department of data processing automation
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
milikhin@gmail.com

***Abstract*—The article is dedicated to give an assessment of possibility of the applicability of web-based geographic information systems (GIS) within the micro geo-marketing researches. Description of micro geo-marketing data domain, its key features and objectives are given. Functional model of the data collection process is presented. Authors formulate requirements for micro geo-marketing web-GIS and suggests using geodata meta model to provide all necessary functionality within data store tier of GIS architecture. The paper also contains original attributive store component data model.**

***Index Terms*—web-GIS, geo-marketing, geodata, data store**

I. Introduction

The first projects in the field of geo-marketing researches in Russia were implemented at the beginning of the XXI century. During the short history of its existence, geo-marketing evolved from being a set of tools for developing business graphics to become a methodology for complex data processing and analysis, combining information monitoring, market researches and geoinformation technologies.

The main methods of the modern geo-marketing are the spatial analysis and visualization of business information by means of specialized geoinformation tools, analytical models and statistical techniques of the multiple factor spatial analysis [1]. Geo-marketing today includes three main directions with own research objectives, customers and tools [2]: the macro geo-marketing oriented on planning and implementation of international company's strategy; the geo-marketing at the mesolevel operating within the scales of the state or state's certain regions; and the micro geo-marketing used to solve problems at the level of the given local commercial organization. The work is dedicated to study the objectives of geo-marketing at the microlevel.

There are about fifty six companies providing services of micro geo-marketing researches in Russia today [3]. The goal of such researches is to maximize profitability of a distribution network. One of the most widespread methods for obtaining primary data about market status in such researches is the method of panels. In practice of micro geo-marketing a few forms of panels are used, and the panel of retail outlets of a long-term type (we will call a periodical panel research long-term if it is continued for a long enough time period – from a half a year till several years) is among them [4].

In such panels on the basis of the selection of number of outlets, that statistically reflects market's behavior, data in the given format is periodically collected. Data collecting is carried out by field workers called auditors [5]. During one visit the auditor fills inventory accounting data for the goods, selected within the research, and their accompanying equipment, estimates the volume of goods for selling, dealer support and price level. Providing audit of outlets allows to determine the general size of the market and the market share belonging to a certain brand, reveal stores with lack of goods or problems with goods placement on shelves.

II. Geoinformation technologies as geo-marketing software basic

Geoinformation technologies and systems (GIS) are usually used as an instrumental basis for geo-marketing researches. The level of development of micro geo-marketing information systems, thus, is largely defined by the level of the modern geoinformation technologies. Geoinformation systems in turn, became an independent class of information systems in the middle of the XX century and already overcame few stages in their development that corresponds to a certain level of information and network technologies and personal computers.

Among the problems of the modern GIS [6, 7] are complexity of access to already gathered materials, and difficulties of data exchange. Separation and non-uniformity of the geodata received from different sources, departmental and even personal barriers, regime restrictions and data storage disorder hinder with rational and effective use of spatial information resources. Thereof, the present stage of development of geoinformation technologies and tasks of geoinformatics are characterized with transition to new forms of the data store organization and spatio-temporal data processing.

Other important directions of GIS evolution include step-by-step displacement of desktop software for workstations with the distributed geographic information systems and wide usage of GIS for solving large number of applied problems. Improving of the Internet technologies availability considerably simplifies access to spatial data and GIS-analysis

tools for end users and causes introduction of concept of a web-based geographic information system (web-GIS). Among the directions of web-GIS development which was widely adopted recently are cloud-based GIS and geoinformation Internet services, combining advantages of web-based approach and convenience of software distribution in a cloud. Online solutions in the field of geo-marketing are provided by many leading GIS-software companies. There is common-purpose extensible online GIS software such as ArcGIS Online, CartoDB and GisCloud [7, 8], and specialized solutions for geo-marketing such as Geomatrix [9].

Besides, there is a high level of interest in implementation of geoinformation Internet services from GIS users today. According to researches of the SpatialVision [10] more than a half of the polled business executives and IT professionals in companies that are engaged in processing and analysis of spatial data are interested in implementation or trying of cloud-based GIS. We also want to mark, that 26% of respondents consider the use of cloud-based GIS within activities of the enterprise as impossible due to the issues on information security and confidentiality, connected to the needs of transferring corporate data into cloud-based storages, and requirement of maintaining a permanent Internet connection [11].

Other important tendency of geoinformation technologies evolution is implementing services for mobile devices, both within web-based systems, and as independent standalone applications that use geolocation services. Geoinformation technologies nowadays find more and more broad usage in mobile applications. 95% of smartphones owners searched information about their geographic location and its environment using their mobile device at least once [1].

III. Micro geo-marketing web-GIS

Taking into account the given traits of the modern geoinformatics, we will formulate the main functional and nonfunctional requirements for GIS intended for carrying out data within micro geo-marketing researches, subsequent analysis of retrieved data and generating analytic and thematic reports for customers.

A. Data collection process

We divide the process of data collection from the point of view of the research project manager into four main stages (fig. 1).

1) Formulate research objectives

At this stage a formulation of the research objectives which are the subjects to develop is given. The structure of the data forms for panel research corresponding to such objectives (that will further used by auditors when carrying out field operations) is defined.

2) Select retail outlets to visit

The business analyst creates a set of outlets corresponding to the research objectives using a database containing data about outlets. When the set compilation is complete, a subset of outlets, the available data about which need to be added or updated in order to provide an adequate data domain model for the research, is generated.

3) Define clusters of retail outlets

Determination of retail outlets to visit is followed by the formulation of individual tasks for auditors. The initial outlet selection is divided into the clusters corresponding to certain field workers. Within each cluster optimum start time and a route for the auditor is defined taking into account traffic statistics and outlets opening hours.

4) Collect data

The final stage of the process is collecting data about outlets by field workers and loading retrieved data into the centralized data store for the subsequent processing and analysis.

Thus, when designing the software for collection and analysis of primary data within micro geo-marketing researches, it is necessary to consider the following traits that refer to separate stages of the considered process:

- Business analyst has the ability to create unique forms for data collection within each individual research project.
- Each data form can be re-used within different researches.
- The number of auditors involved in data collecting within a single research may vary.
- Business analyst defines the set of outlets to visit, on the basis of which the project manager creates individual tasks for auditors.
- Project manager defines auditor's tasks as clusters consisted of points to visit and the data collection forms that need to be filled for each outlet.
- Auditor's route within the cluster should be generated optimally to minimize financial costs and time required to finish the tasks.
- Each data form consists of fields, and each field is intended to be filled with the data of a certain type (string, number, date, etc...) and format (e.g., address, text field with specific structure or one of several predefined values);
- When the task is finished, the auditor sends collected data to the manager for further processing and analysis.

B. Web-GIS requirements

In case when there is no common software platform for the data collection and processing, various software components and information technologies are used to fulfill users' needs. The above domain features allow us to define a number of problems arising during typical micro geo-marketing research.

1) Lack of single style when filling out form fields

Due to the lack of classifiers, fields of the same type sometimes filled in by auditors with different styles. Problems can usually occur with addresses or form fields supposed to have a qualitative assessment of different aspects of the outlet's activities.

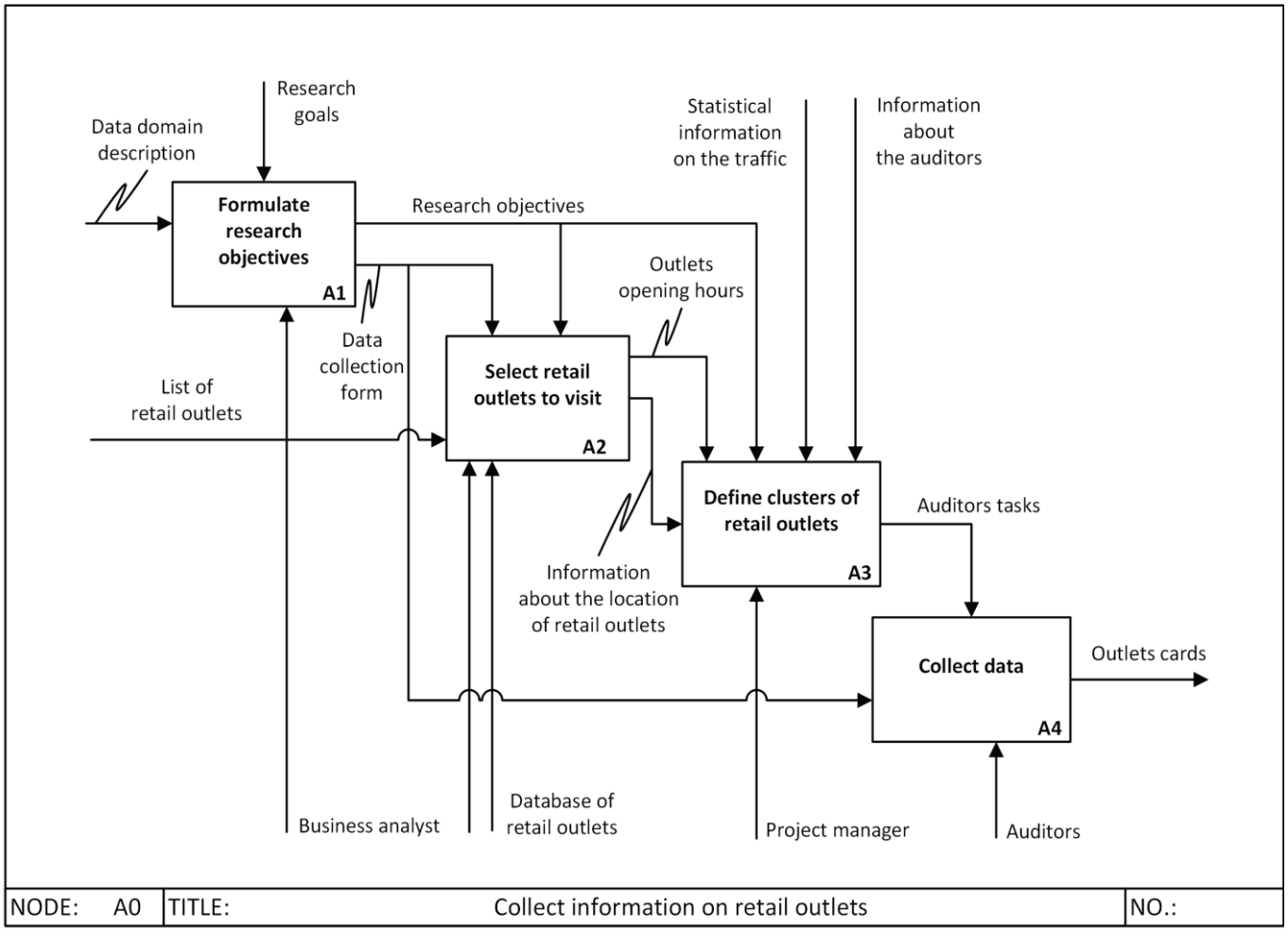


Fig. 1. Collect information on retail outlets (functional model fragment)

2) Data aggregation and storing

Data collection is usually performed with special software for mobile devices. In this case data exchange between different software systems is usually proceeded using spreadsheets. Processing of such documents and bringing the data to a single common format results in additional time spent required from the project manager.

3) Manual generation of clusters

The list of outlets to visit for each auditor must be formed manually within every research. The task increases the time spent by manager in order to organize research, and the possibility of manual optimization of generated clusters is limited significantly.

4) Need for support of different software products related to different stages of research

Performing of field work requires software products with rich data processing functionality such as pre-generating data forms, viewing tasks issued by the project manager and providing location-based navigation services. It is also necessary to mark the need for conversion of data between the software used by auditors for data collecting, and the business analyst for data processing and generating graphic reports for customers. Besides, storing large volumes of data and providing access to actual versions of the documents to company staff requires to setup separate data synchronization systems and control data access rules.

5) Requirement for customer to have specialized software to view the generated reports

Research results can be presented in the form of a spreadsheet or using file formats that are supported by common instrumental GIS. In both cases customers need special software to view generated reports.

IV. Data Model for Micro Geo-Marketing GIS

A. Web-GIS architecture

One of the main tasks within development of the GIS proposed as a solution for the revealed problems is to create the most effective system architecture considering all key features of the given business process. Such architecture should be developed taking into account the possibility of system usage at all stages of geo-marketing research and thus provide high level of productivity, scalability, reliability and minimize time-related expenses of users.

Designing GIS architecture is in a broad sense reduced to a choice of the main system's components: basic geoinformation

technology, storage for spatial data etc. In the narrow sense GIS architecture is applying the most effective architectural concepts at each level of the geographic information system. The architecture of web-based GIS usually corresponds to the three-tier [12, 13, 14] model of the client-server software and includes three main components:

1) Client tier

The user access GIS tools with a web browser or the client application that implements special API to access the GIS-server. Thus the client tier of the system must be implemented using modern web frontend technologies, such as HTML5 [12, 15, 16] or Adobe Flash [13]. This tier includes desktop and mobile GIS software and web-based GIS clients.

2) Server tier

Server tier includes application and mapping servers. Depending on the used architectural templates, data visualization, spatial analysis features and domain's business logics can also be implemented at this level [13, 14].

3) Data store

The centralized data store is used for storage of both spatial, thematic data and data of the web application, and could be implemented using relational or document-oriented databases and file storage. This tier corresponds to spatial database management systems.

Authors suggest to use the web-based geographic information system WGS4 [11, 12] as a basis of software solution to the designated tasks. Key features of the WGS4 at the client and server tiers are explicitly reviewed by authors in other articles [12, 17]. In the paper we will consider the task of implementing specified functionality of micro geo-marketing GIS at the data store level.

B. Web-GIS data store

GIS data store is intended for aggregation and storage of geodata. Geodata consists of the attributive (thematic), spatial and temporal data and reflects different object properties, the processes and the phenomena occurring on Earth [18]. The spatial component of geodata (SD) characterizes objects location and other spatial attributes, temporal component (TD) reflects change of objects properties during given time spans, and attributive component (AD) describes specificity of the data domain.

Within the data domain of micro geo-marketing researches each of three main geodata components has certain key features. Arbitrary number of data fields of different types and an inconstancy of such structure are peculiar to the attributive descriptions of outlets. The spatial component is characterized by needs of spatial data geocoding, creating sets of clusters, that defines targets to visit for each auditor, and ability to search for optimal routes for auditor within the given cluster. The temporal component reflects changes in attributive (outlets descriptions) and spatial (location of outlets) data in time and statistical information about road traffic.

Each outlet can be described from the different points of view and have number of corresponding attributive descriptions for researches of different subject. Some outlets may have different attributive descriptions within one research at the same time if such descriptions correspond to different time spans. Thus, a spatial object can correspond to one or several attributive descriptions of different types, and a separate outlet can also have different attributive descriptions of one type corresponding to different temporal data components.

Attributive descriptions of outlets are implemented in the WGS4 software as cards of spatial objects. Each card consists of a set of the characteristics divided into groups. According to the system's functional requirements the software for dealing with attributive data shall provide possibilities of adding new characteristics of any supported data type or groups of such characteristics, and creating new descriptions for the outlets, which are at the same time represent forms for data collection and provide information required for generating analytical and thematic reports.

Taking into account the stated requirements, the attributive database is implemented using the geodata meta model, according to which the storage of spatio-temporal data contains, except characteristics values, description of attributive cards, their structure and fields. Such approach allows to provide flexibility in system's setup in compliance to needs within the current research. Attributive data in represented with the following entities (fig. 2):

- Property – describes single attributive property of a certain data type which can be bound to one or several types of attributive descriptions.
- Group of properties – describes combining of properties in groups on a basis of thematic similarity.
- Attributive type – describes the certain type of attributive cards which is characterized by a similar set of form fields (properties). The attributive type characterizes structures of the data collection forms within the micro geo-marketing research.
- Attributive description – describes a separate card for a given outlet, while the set of properties of the attributive description is defined by its type. One spatial object could have several attributive descriptions corresponding to different aspects of outlet's activities.
- Relation between attributive types – describes possibility of creation of hierarchical connection (cards hierarchy) between two attributive descriptions of certain types.
- Relation between attributive descriptions – describes hierarchical connection between cards of a single outlet. Such relations are used to store data changes during the research's time span. Each child card characterizes outlet's status at a certain moment.
- Value of the property – describes value of the given property of a specific outlet at a given time.

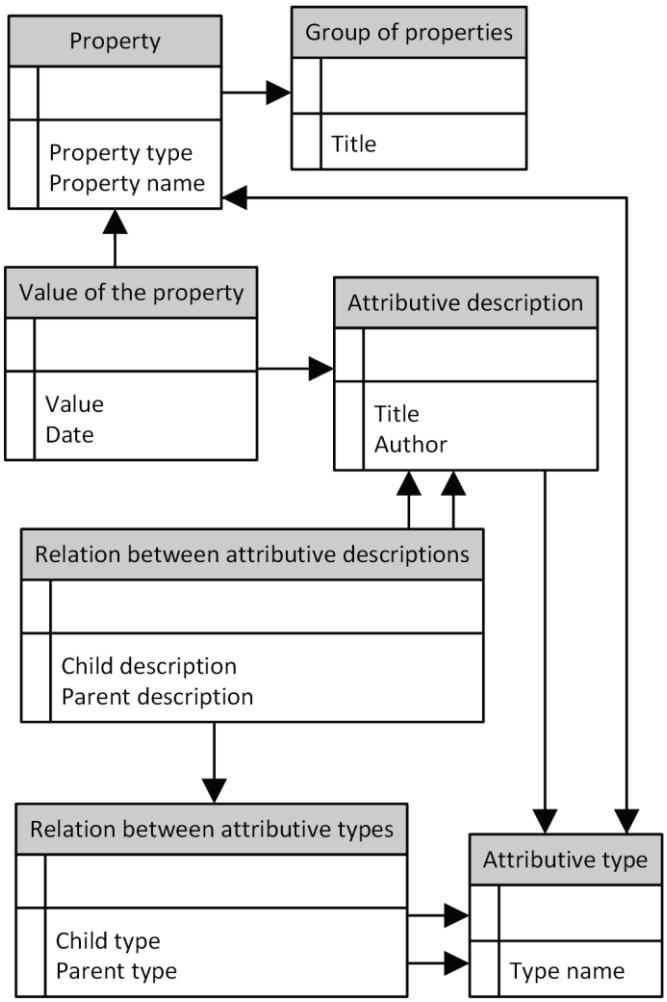


Fig. 2. AD data model fragment

The offered model supports hierarchical structure of attributive descriptions of any necessary complexity with possibility of creation of "one-to-one" or "one-to-many" nested structures corresponding to all features of micro geo-marketing researches.

V. Conclusion

Revealed data domain features allow authors to formulate functional requirements to the web-GIS for micro geo-marketing researches. Such requirements create possibility to formulate retail auditor's problem and continue further development of web-GIS as a union basis of micro geo-marketing software.

The specificity of geodata and its components in the context of data collection and processing allows developing the geodata common meta model. The data model offered by authors could be used as a basis for the centralized storing of attributive data on retail outlets taking into account the inconstancy of the data structures within research and the need for storing data changes within the research's time span.

Acknowledgment

This article is designed as part of the national project of the Ministry of Education and Science of the Russian Federation № 3653 “Models, algorithms and software to support decision-making in risk management in the socioeconomic and industrial-technological system”.

References

[1] T. Ozhereleva, “Development of Geo-Marketing,” European Researcher, 86(11-1), pp.1776-1781, 2014.

[2] A. Duplenko, E. Levina, “The development of Internet technology in geo-marketing”, 2014. Molodoy uchyonyi [Junior researcher], (61), pp.437-439.

[3] S. Shcherbina, “How to increase profits using maps,” E-xecutive. Managers Community [Online]. Available: http://www.e-xecutive.ru/management/marketing/1857639-geomarketing-kak-uvelichit-pribyl-s-pomoschu-karty.

[4] I. Belyayevsky, “Marketing research: information, analysis, forecast”, Moscow: Finansy y statistika, 2008, p.400.

[5] G.A. Churchill, T.D. Brown, “Marketing research”, 2007.

[6] U. Skjelbo, J.J. Jensen, J. J. Jorgensen, “Geographic Information in Enterprise Architecture”, Informi GIS A/S, 2009, p.22.

[7] S. Steiniger, A. J. S. Hunter, “Free and open source GIS software for building a spatial data infrastructure”, Geospatial free and open source software in the 21st century. Springer Berlin Heidelberg, 2012, pp. 247-261.

[8] M. Milikhin, M. Rychagov, “Web-GIS as Internet service”, Geoinformation Technologies within the problems of rational natural resources usage, Khanty-Mansyisk, 2015. – pp. 18-20.

[9] A. Trousse, “The main problem for geo-marketing company is to create data store”, GisGeo. Geo-marketins and spatial analysis. http://gisgeo.org/news/rilos_interview.html.

[10] “Spatial Vision hosts "Mapping the Cloud" webinar”, Spatial Vision | Innovative geospatial solutions [Online]. Available: http://www.spatialvision.com.au/index.php/about-spatial-vision/our-company/663.

[11] Yu. Gritsenko, M. Milikhin, P. Senchenko and O. Zhukovsky, “Model of role-based access to spatial data of electronic master plan”. In Application of Information and Communication Technologies (AICT), 9th International Conference, 2015, Ocober, pp. 196-200. IEEE.

[12] M. Milikhin, Yu. Gritsenko, M. Rychagov, “Combined method for geodata visualization within web-based GIS”. Doklady TUSUR [Reports of TUSUR], 2015, №. 1 (35), vol. 2, pp. 112-115.

[13] M. Adnan, A. Singleton, P. Longley, “Developing efficient web-based GIS applications”, UCL working papers series, Centre for Advanced Spatial Analysis UCL, 2010, paper 153, p. 16.

[14] M. W. Xie, et al. "Research on Web-GIS-based online visual analysis of landslide monitoring data", unpublished.

[15] S. Rao, S. Vinay, “Choosing the right GIS framework for an informed Enterprise Web GIS Solution,” CIESIN, Columbia University & NASA New York, USA, 2009, T. 17.

[16] M. N. K. Boulos et al, “Web GIS in practice VIII: HTML5 and the canvas element for interactive online mapping,” International journal of health geographics, 2010, vol. 9, №. 1, p. 14.

[17] Yu. Gritsenko, “Typical architecture solutions for masterplan geographic information systems,” Doklady TUSUR [Reports of TUSUR], 2011, vol. 2 (24), i. 2, pp. 171-175.

[18] V. Maslennikov, V. Tsvetkov, “Dynamic geodata model. Geoinformatics,” Sciences about Earth, 2012, vol. 1, pp.49-53

2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)

Comparative Analysis of Strategies and Trends Shaping The Information Economy of Azerbaijan with Leading Countries

Ph.D. in economical sciences A.G.Aliyev[1], R.O.Shahverdiyeva[2], V.A.Abbasova[3]
Institute of Information Technology of ANAS
Baku, Azerbaijan
[1]alovsat_qaraca@mail.ru, [2]shahverdiyevar@gmail.com [3]vusaleabbaszade@gmail.com

Abstract: **Here is information economy development trends in leading foreign countries are analyzed. Formation of information economy experience has been summarized in the case of the U.S., South Korea, China, Ireland and so on. Has been reviewed comparative analysis of strategies and trends shaping the information economy of Azerbaijan with developed countries. Given a number of recommendations put forward for future development.**

Keywords: information economy, ICT and e-services, "E-state, high-tech park, e-commerce

I. Introduction

In recent years, developed countries make rapid transition into the new type of economy based on modern ICT technologies, information and knowledge. Transition into the new economy is characterized with changing in the nature of traditional productions areas, shifting the information into the major production resource, and economic system to be the main development factor.

It has been globally recognized that the economic health of a nation is increasingly tied to its development of a sustainable information economy. And the information economy is characterized with changing in the nature of traditional productions areas, shifting the information into the major production resource, and economic system to be the main development factor. During the final decades of the twentieth century, developed countries invested heavily in information and communications technology (ICT) to maintain their competitiveness in the global information economy and emerging economies are pursuing a similar path of economic development through enhancement of their ICT sectors and promotion of ICT adoption in other sectors.

A strong linkage between the production and use of ICT and the development of a sustainable information economy, ICT contributes to economic growth of emerging economies. Hence, a sustainable information economy depends upon both the production of ICT goods and services and the diffusion and use of ICT in other sectors such as agriculture, manufacturing, education and public sectors. At present, a new economy is being formed in Azerbaijan, which is based on information, knowledge, and modern technologies. The development information and communication technologies, and transition to Information society, expanded use of ICT and e-services at the state and local governmental bodies, meeting information the needs of society for information products and services, and the training of scientific human resources and other qualified professionals, and etc. are of prior tasks of the Development Concept of the country by 2020 [1]. Depending on the development characteristics of the countries there exist some definite differences and problems in the formation and the development of the economy. Therefore, the USA, the European Union countries, as well as developed countries such as Japan, and the developing countries such as China, India, South Korea and etc. study current development trends of the formation of information society and economy and conduct researches for its successful implementation.

II. Formation of Information Economy in Azerbaijan

Taking into account the characteristics of the country for current period, the national development model is adapted to analyzed development trends of advanced countries.

Azerbaijan is one of the emerging economies in Europe and Central Asia, a region that has experienced substantial economic growth during the last decade. İn its effort to develop a sustainable information economy, Azerbaijan has been actively seeking ways to develop its ICT sector, attract foreign investment, foster domestic entrepreneurship, and raise its competitive position. ICT sector in the country was declared a priority in recent years, and many large-scale projects are implemented in this regard. That is why, according to some development indicators, Azerbaijan left behind some leading countries. "The Global Information Technology Report 2014", developed by the World Economic Forum, rated Azerbaijan 49th among 148 countries for its "Networked Readiness Index", which reflects the development of the countries in the field of information technology [3]. Azerbaijan was rated 56th last year. According to this table, Azerbaijan left behind even some leading countries. It maintains its leading position in the CIS, being rated 59th for the number of Internet users per 100 people out of 148 countries, and 46th for the number of broadband Internet subscribers. In addition, due to the innovative ability of the companies, Azerbaijan rated 35th in the world, and 1st among CIS.

Currently, there are about 1 million Internet users in Azerbaijan. 50% of which is broadband users.

Recently, ICT sector has become leading and dynamically developing area of the economy. Over the years, investment to the industry amounted approximately $ 2.5 billion.

Mobile network has been expanded. 4G technology was introduced in the country since 2012. There are 110 mobile subscribers per 100 people in the country. Internet channel capacity increased by 12.9 times in last five years, and reached 200 Gbit/s, and the size of the Internet services market increased about 4 times [4].

On February 8, 2013, the first telecommunication satellite of the Republic of Azerbaijan "Azerspace-1" was launched on the orbit, which is the highest technical achievement of the country since its independence. Commercial exploitation of the satellite started, which is capable to provide Europe, Middle East, Central Asia and African countries with telecommunication, Internet, television and radio broadcasting services.

"Electronic-State" project is carried out in the country successfully, the use of electronic signatures is expanded. "E-state" portal (www.e-gov.az) was launched in order to provide information exchange among information systems of government agencies, and to ensure e- services through "single-window" principle. 179 electronic services submitted by 30 state agencies are available for the users. At the same time, service centers "ASAN Xidmat" (EASY service) are organized to provide comfortable and higher quality services to citizens from single location, with the use of modern innovations, which is highly appreciated by many international organizations and universities, and became the brand of Azerbaijan.

High-specialized human resources are trained in ICT to expand national capacity. State Fund for Development of Information Technology and "High-Tech Park" have been established in the country to provide developed and competitive innovative ICT industry with high export potential [2]. These new bodies will contribute to strengthening economy of the republic of Azerbaijan, attracting foreign investment and expanding ICT products, and so on.

Never use letter spacing and never use more than one space after each other.

III. The Information Economy and ICT Development Tendencies of Various Countries

The U.S. is a leader in the field of formation of information economy. The U.S. developed comprehensive, multi-level system of national innovation in this area, which can be even regarded as the world standard. The U.S. carried out its implementations in the fields of science and innovation, education, the development of ICT, risk financing, the development of the information economy, which meets requirements of the legal framework, as well as a variety of related activities [5].

Due to the tax regulation mechanisms of American tax policy, the U.S. investment in informatics tools increased by 4 times each decade, and the labor productivity averagely by 2.1 %.

28 % of all research and experimental-design work in the U.S. (R&D) are conducted with financial support of the state. Significant role of the state in the formation of information market is associated with a high capital-intensive of the whole software and innovation activity.

The state support for newly established enterprises includes reducing state barriers, ensuring their access to credit resources, preferential tax for small and medium businesses. First of all, American experience in the information and telecommunication industry should be especially noted in the field of antitrust regulation. Investment of billions of dollars in R&D sphere is strongly influenced by the state monopoly policy.

According to the scale of costs of R&D, logistics and human resources in recent decades, United States left behind all other countries. In 2011, the total costs incurred by the US to R&D amounted $ 499 billion (3.7% of GDP). This amounts approximately 42% of total cost of research and development of Organization for Economic Cooperation and Development (OECD) member countries [10].

Scientific and technical potential of the U.S. is concentrated higher in some states - California, Michigan, Massachusetts, Pennsylvania.

Israeli high-tech revolutions is the basis of the country's economic development. Breakthrough in the field of high-tech in the country was due to the massive import of "human capital". An impact on the world high-tech products market is the characteristic of the Israeli economy

The Israeli government adopted a package of new programs to promote technological innovation. Under the programs, the government-owned insurance company provided investment funds at risk with investment guarantees up to 70% of their initial capital. Thus, the high-tech exports increased by $ 21 billion in 2011. At the same time, the share of high-tech products in total exports (up to 57% in 2011) increased significantly [10].

Japan Basic Law on the information society came into force in January 2001. According to this law, the state program "e-Japan", which provided conversion of the country into the world's leading information state, was implemented [8]. The program particularly focused on investing funds in the creation of networks with high bandwidth applications, R&D, the use of the Internet at the state enterprises, training of more professional human resources, and etc.

In early XXI century, international competitiveness of Japan is associated with the development of the Internet and overall economic liberalization. The new methods were adopted for GDP measurement, which benefits from software, new technology, services, healthcare, infrastructure and other advantages more precisely.

The Republic of Korea established its national policy for science and technology according to the "general innovation" program. The national innovation system was expected to be the most important factor for the development of national economy of the country. The development of information areas, such as production of special displays for digital information transmission, intelligent robots, digital contents and etc. have become priority of the national economy.

Success of South Korea in this field is due to the publicity of telecommunications market. Long-term development strategy of the telecommunications market is to provide high-speed Internet access. Antitrust policy of South Korea contributes to the development of video-information transmission technologies via Internet [10]. This, in its turn, contributes to the competitive advantages of South Korean manufacturers in the field of mobile devices provided with Internet access.

In 2006, the share of the R&D expenses amounted 1.8% of GDP of People's Republic of China. 40% of investment to R&D was provided by the state, and 60 % by domestic and foreign private firms. In 2003, China rated 3rd country in the world for the amount allocated to research and experimental-design projects. According to this indicator Chinese lagged behind the U.S. and Japan. China distinguishes for its highly developed Internet, and it is supported and controlled by the

state directly. Including the Internet, information technology, first of all, is developed to present China's export products to the world market [8].

The Chinese government supports distribution of Internet technologies in most advanced areas of industry, in order to improve the management quality and competitiveness efficiency.

India has made significant progress in the development of the information economy. Approximately 4 million people in the country are engaged in the spheres of service, and 340 thousand in computer companies. IT industry of India is specialized in the field of outsourcing and offshore programming [5]. Offshore programming development is fully supported by the government; India increases its budget for training human resources, R&D and e-commerce to develop information economy. India is considered to be sufficiently serious competitor for Central and Eastern European countries specialized in the field of offshore programming, as well as Russia.

In the EU countries (Sweden, Finland, Norway, Ireland, UK) implemented a successful model that combines a high level of human capital, innovation and in information and telecommunications infrastructure. Market economies are characterized by a significant influence of the state on the market mechanism and a strong system of social provision.

India, South Korea, China, Thailand had some success in improving the educational potential of the country, qualification of the officials that was the impetus for the development of knowledge-based industries in these countries. Moreover, these countries have been able to go with their goods and services to the global information technology market. At the same time the standard of living here is pretty low, on the order below the standard of living in Singapore or Taiwan [6].

In these countries the low domestic demand for high-tech products, and hence the level of development of information and telecommunications infrastructure. High qualification and low wages specialists form concurrently advantages of these countries.

The analysis of development tendencies of various countries in the formation and development of information economy showed, that ICT infrastructure, national innovation system, the development of scientific and service industries, the export of high-tech products, higher potential of education, high-level specialized human resources, and etc. are the key factors in the formation and development of information economy (Table 1).

IV. The Comparative Analysis of Development Tendencies of Various Countries

Although there is no one-size-fits-all path of information economy development, we argue that there are some general themes that are fundamentally important for developing a sustainable information economy in any context.

Next, we compare the case of Azerbaijan with that of South Korea, Ireland, India and China in order to gain perspectives on the opportunities and challenges facing Azerbaijan as it competes in the global market. Ireland's success in leapfrogging from a traditional agrarian economy to an information economy has been largely attributed to the development and maturation of the Irish ICT sector Hence, drawing on the experience of Ireland provides rich insights for analyzing Azerbaijan's path. China, India, South Korea, Singapore, Thailand and etc. are among the fastest growing economies in Asia and are at similar stages of economic development yet show clear divergence with respect to their ICT sectors [7]. Thus, these countries provide additional perspectives regarding viable paths available to nations for building their information economies.

TABLE 1. Information Economy and ICT Development Trends

Country or country groups	Development directions	Development trends of ICT industtry
The USA	Computer hardware and software industry NIS	ICT infrastructure, export of hi-tech products and services, human resources and financial support for R&D, tax incentives, antitrust policy, etc.
Asia-Pacific countries	Computer hardware and software industry	Human resources, R & D investment, and government regulation
Finland, Israel, Singapore	IT industry	Technology infrastructure, R&D, firm strategies, and capital availability
China, India, Ireland, Israel, Russia	Software industry	International and domestic demands, national policies, a skilled workforce, technological infrastructure, financial incentives, and R&D investment
South Korea	NIS Electronic industry	export of hi-tech products and services, higher education capacity, human resources, development industries
Japan	NİS IT industry	higher education capacity, ICT infrastructure, development of the industry, investment for the R & D and human resources, etc.
Azerbaijan	NİS IT industry	ICT infrastructure, investment for the R & D and human resources, national policies

The results suggest that although Azerbaijan apears to be on a positive trajectory toward a sustainable information economy, areas remain that require improvement if economic development is to continue. Azerbaijan needs coherent economic, taxation, trade, ICT, and human capital policies in order to facilitate the development of a sustainable information economy.

Infrastructural challenges include uneven physical infrastructure development across regions, and human infrastructure issues in the form of a skilled workforce that is too small, and low R&D expenditures in the ICT sector. Future economic growth of Azerbaijan will depend on an increase in investments and improvement in technology and innovation.

A well developed infrastructure is very important to facilitating the work in the ICT sector. Ireland is in the leading po-

sition with respect to infrastructure development because of its early start, the strong support from the Irish government, and the active involvement of the private sector. The indicators also show that China is far more advanced than India with respect to ICT diffusion and infrastructure development. However, China and India are at a similar developmental stage and carry similar developmental burdens such as a large population and regionally- skewed development. The advancement of China in terms of infrastructure development can be attributed to a considerable amount of investment, nationwide developmental strategy, strong interventions and controls of the central government [7].

In India, on the other hand, despite its domination in the global ICT services export market, the service industry is heavily concentrated around major clusters including Bangalore, Mumbai, Chennai, Delhi and Hyderabad.

In these clusters, the infrastructure has been developed to facilitate ICT sector work. However, for the rest of India, the infrastructure in many rural areas is inadequately developed, thereby lowering the national average indicators and posing as a barrier to future growth.

As discussed in the previous section, Azerbaijan has made steady progress towards infrastructure development. As a latecomer participant in the global ICT market, Azerbaijan can learn some valuable lessons: continuously investing in ICT infrastructure, involving both public and private sectors in development activities, focusing greater effort on establishing a strong technical and human infrastructure (education and R&D expenditure), and promoting balanced nationwide ICT development. One common lesson learned from the ICT sector and information economy development in many different countries is the criticality of ensuring the supply of a qualified knowledge workforce. The education was also made essentially free in Ireland, something that ensures the supply of a skilled workforce.

In addition to the government support, another unique aspect of the Irish software sector development is the close linkage between the academic institutions and industry. There are many similarities between China and India in terms of the supply of skilled workers. Building on the large population base in both nations, both countries are facing the challenge of skill gaps. The lack of English-speaking proficiency in the ICT workforce is considered one of the biggest disadvantages of China in competing with India in the global ICT offshore outsourcing market.

In addition to reforming the education curriculum and improving the education quality, one of the strategies that both nations can leverage is the reverse brain-drain. Each year, a large number of Chinese and Indian college students come to the U.S. and other developed countries to pursue advanced degrees and careers. Therefore, it is important for India and China to turn the brain-drain into a brain-gain, by providing incentives and opportunities that encourage returning, and that cultivate the great potential of such mobile human capital.

The challenges China and India are currently facing in workforce supply are valuable lessons for Azerbaijan. And Azerbaijan needs to improve the quality and quantity of IT workers to support the growth and value-added contribution of its IT industry.

Therefore, it is very important for Azerbaijan to carefully evaluate the global market trend and its distinctive characteristics to define its own developmental path.

CONCLUSION

The studies prove, that establishing efficient information infrastructure based on free and healthy competition of telecommunication market, the development of ICT sector, including the high-tech sector and the services, investing to R&D and high-tech institutions, expansion of the export of high-tech products, computer literacy of the population, availability of efficient and secure information exchange, training of human resources, government support and the legal and regulatory framework are of great importance for the formation and development of information economy.

Research has also shown that there is no “one best model” for information economy development. Further, globalization requires that the development of a nation’s information economy must be closely connected to the global market, as the cases of China, India, Ireland, Singapore, and Taiwan show.

Current development rate of the country, taking into account its features, benefiting from the experience of the advanced countries, information society and economy can be certainly formed and developed in Azerbaijan.

REFERENCES

[1] National Strategy for the development of information society in the Republic of Azerbaijan for 2014-2020. [Online] Available: www.e-gov.az/news/

[2] Development Concept of "Azerbaijan 2020: vision to future", 2012. [Online] Available: www.president.az/files/future_az.pdf

[3] World Economic Forum (2014) The Global Information Technology Report 2014. www3.weforum.org/docs/WEF_Global Information Technology_Report_2014.pdf

[4] A.G.Aliyev. 2013. Analysis of the current state of the Internet economy and its development prospects // Problems of Information Society, 2013, № 2.

[5] A.A.Enzhaev. Features of state regulation of information economy in the leading countries of the world / Vestnik. 2008. № 5 (24) / Economic Theory

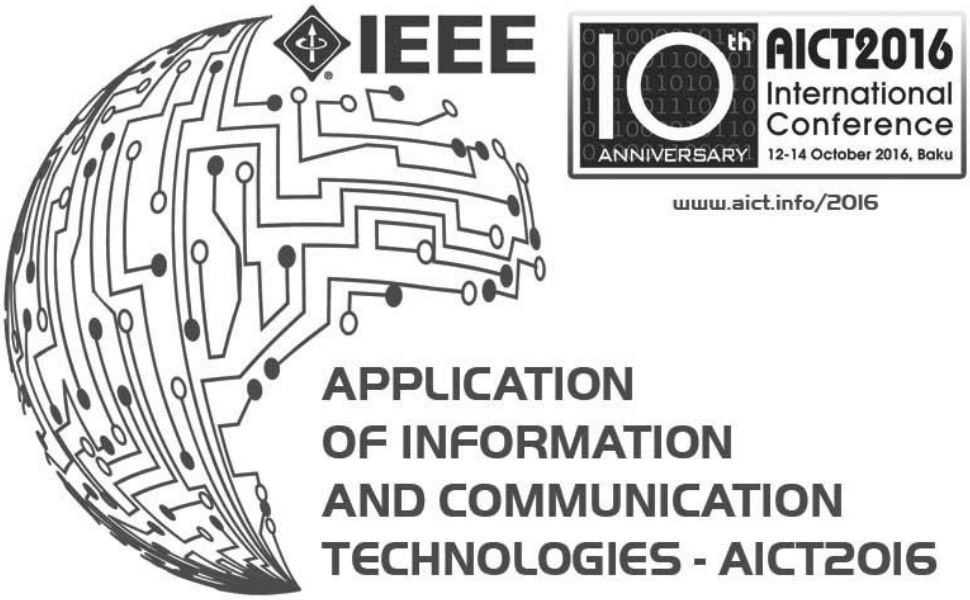
[6] O.V.Budaeva. 2010. Assessment of development of information economy in different countries//Vestnik ESSTU: 4 (31), 109-115

[7] A. A.Techatassanasoontorn, H.Huang, E.M.Trauth, S.Juntiwasarakij. Analyzing ICT and Development: Thailand’s Path to the Information Economy. Journal of Global Information Management: 2011, 19(1), 1-29.

[8] M.Li, M.Gao. Strategies for developing China’s software industries. Information Technolo-gies and International Development, 2003, 1(1), pp.61–73.

[9] E.A.Petrova. Foreign experience of informatization and features of its implementation in Russia // Scientific Journal "Basic research". 2007, 11(2), 65-73.

[10] M.A.Nikitenkova. Innovative activities of U.S. ICT corporations in the context of globalization // "Russia and America in the XXI century", 2012. № 2.



CONFERENCE PROCEEDINGS

SESSION 7.

ICT IN EDUCATION, HUMANITIES, SOCIAL SCIENCES AND RESEARCH

12-14 October 2016, Baku, Azerbaijan

ICT in Humanities: phenomenon of Corpus Linguistics

Russian national experience

Galina Kedrova
Philological Faculty
Moscow State Lomonosov University
Moscow, Russian Federation
kedr@philol.msu.ru

Maria Volkova
Center of New Technologies in Humanities
Moscow State Lomonosov University
Moscow, Russian Federation
contact@philol.msu.ru

***Abstract*— The paper presents various Russian language corpora to discuss professional advantages and cultural benefits of linguistic corpora technology in comparison with the pre-computational and pre-corpora state-of-the-art in language research and Arts and Humanities. As the most faithful 'mirror' of political, intellectual and spiritual life of a nation during current state and in historical perspective, linguistic corpora could be also applicable as innovative research methodology in wide variety of research projects in Social Sciences and Humanities. Some valuable examples are presented.**

***Index Terms*—Language corpora, Linguistics, Russian National Corpus, research methodology in Corpus Linguistics**

I. Introduction

Language sciences were the first field in Humanities that benefited most from development of ICT. Advances in modern computer technologies have recently moved to the forefront of research activity a new linguistic sub-discipline – Corpus Linguistics. The exceptional role of modern language corpora in language sciences and in other Humanities is that they serve as unique reliable source of objective information on the actual functioning of language. And which is even more importantly, the language taken not only as a system (as in structural linguistics), but also as linguistic performance, or "Parole" in terms of Ferdinand de Saussure [1], i.e. a macrocosm of all written and oral texts already produced or to be produced in prospect.

Corpus linguistics also provides new theoretical approach for research and education, while it "re-unites the activities of language gathering and theorizing and argues that this is leading to a qualitative change in our understanding of language" [2]. It is already a kind of common statement that modern corpus linguistics turned to be an exceptional tool to bridge the gap between formal and intuitive approach in Social Sciences, in Arts and Humanities. One of the received definitions for 'language corpus' is as follows: Language corpora are "large quantities of natural, 'real' language, taken from a variety of sources and lumped together in a computerized system, so that people, dictionary people in particular, can study the meanings and patterns that emerge" [3]. The first and still very intensively and successfully used application of language corpora resources has been in second language teaching and learning (for overview of contribution of the corpus linguistics into ESL domain see [4]).

Not every compendium of texts however big it may be should be considered a corpus. On the first hand a text collection should be classified and categorized to form a database. Each text item (word, collocation, sentence, chapter, etc.) must be processed according to regulations known as annotation, or tagging. Through tagging activity the peculiar linguistic information about each word's relevant characteristics (grammatical, stylistic, phonetic, syntactic, paralinguistic categories such as gender, grammatical meaning, word class mark, etc.) is added to the corpus item in the form of tags. The tagging procedures and practices, however very time consuming and labor-intensive, should be considered as universal and indispensable approach in building language corpora of every type. Recently more tagging operations became automated, as multiple sophisticated computer algorithms turned to be more effective in the domain of computer text analyses. Linguistic corpora become thus a unique source of statistically proven, verifiable and reliable information on language operation. Thus, language corpora resources and linguistic corpus facilities find ever-growing use in multidisciplinary research in Social Sciences and other Humanities as well [5].

The main purpose of the corpus is to facilitate research on the lexicon and grammar of a language, as well as processes of language change within a certain, ad hoc defined period of time, normally within one to two centuries. The other purpose of the corpus is to serve as a reference point for lexical, grammatical, and accentological questions, and as source of changes in the history of language [6]. However, lately it has become obvious that besides Linguistics per se, a wide variety of other Social Sciences could and should benefit from language corpora data.

II. Corpus Linguistics in Russia

We must approve that Russian linguists achieved outstanding results in modern corpora building. One of the main reasons is that since the whole of our linguistic tradition from the beginning of the 20th century had been speech-

oriented, focusing on how languages work and how words are used to shape concepts and reflect our perception of the world [7]. There always have been many outstanding Russian scholars recording large quantities of the language data in literature, folklore, and poetry. Many of these data have been used to form a cornerstone of the Russian lexicographic tradition, were worked up into dictionaries, encyclopedia, and concordances based on the real data which have been collected for decades and even centuries by several generations of Russian and foreign linguists. Therefore Corpus linguistics in Russia is beyond all doubt in the vanguard and continues to forward very successfully trailing the path in various sectors of general and specialized linguistic corpora construction activity [8]. Therefore we have now a whole variety of specialized Russian language corpora, among them only few the most important to mention:

Corpus of Standard Written Russian (http://www.narusco.ru/) – a well-balanced (normalized according to real frequencies for every type of text) collection of literary and scientific Russian texts produced in time period of 1950 – 2012, with accent mark on every word item.

Russian Emotional Corpus (REC) (http://harpia.ru/rec) – a multi-modal corpus with functional annotation covering facial expressions, hand, body and head movement which allows analysis of communicative functions of these actions, for example, the expression of inducement or request of feedback, or ways to express joy and happiness.

Russian part of SPAAC (Speech Act Annotated Corpus) – Speech Corpus of Everyday Communication (http://narusco.ru/resourses.htm) that comprises annotated Russian speech records taken through continuous recording of 24 hours speech activity of an experimental subject. This corpus is compiled with the Russian oral speech corpus.

Online Corpus of Russian Sign Language (RSL) (http://rsl.nstu.ru/site/index/language/en) has been created on the basis of the annotated RSL corpus developed during the project "Corpus-Based Research on Morphosyntax and Vocabulary of the Russian Sign Language" supported by the RFBR [9]. The texts of the corpus represent two different local variations of RSL – the "Siberian" and the "Moscow" one.

Russian Learner Translator Corpus (RusLTC) (http://dev.rus-ltc.org/search) is mostly used in translation quality assessment research and in didactics of translation. Research projects conducted within the RusLTC corpus help to investigate the linguistic features of student translations from English into Russian against comparable non-translated Russian texts [10].

Corpus of Russian student texts (http://web-corpora.net/CoRST/search) is a collection of Russian texts written by students of different universities. The texts are annotated in several ways (meta-textual annotation, morphological annotation and error markup), which enables users to perform many types of search in the corpus. The corpus includes texts written by students of the following academic majors: economics, sociology, political science, law, psychology, journalism, linguistics, history, philology, logistics, mathematics, and philosophy. Generally, the corpus provides information about the gender and age of the author, and also about the academic year (1st year bachelor, 2nd year master, etc.). Main types of texts are as follows: course papers, term papers, bachelor's and master's theses, essays, abstracts, reports, summaries, autobiographies, etc. [11].

The Russian Language component of the CHILDES (Child Language Data Exchange System) Project – http://childes.psy.cmu.edu/ is aimed to fix and analyze children's speech from birth until the end of teenage period. One of the recent inputs to the Corpus is a sub-corpus of recordings produced by primary school bi-lingual Kazakh children in Almaty city [12]. This resource is a valuable tool for investigation of formation of various kinds of bilingualism.

Alongside with intensive work on building various types of specialized Russian language corpora, the main authoritative source on the Russian language taken as multifaceted, multi-modal social phenomenon is still the Russian National Corpus (RNC) (http://www.ruscorpora.ru/en/index.html). Currently the RNC comprises 600 million words and encompasses 10 domain-specific sub-corpora. One of the most informative sub-corpuses of the RNC is the one with modern and old (from XVIII-XIX centuries) written texts. It is also the core collection database, and the largest among all the RNC's sub-corpora.

The sub-corpus includes various types of texts representing standard Russian language performance:

- Modern fiction of various genres;
- Modern drama;
- Memoirs and biographies;
- Journalism and literary criticism;
- Scientific, popular scientific and teaching texts;
- Religious and philosophical texts;
- Technical texts;
- Business and jurisprudence texts;
- Day-to-day life texts, including texts not intended for publication (letters, diaries, etc.).

All types of texts in the Corpus are represented in proportion to their share in real-life usage. For example, the share of fiction (including drama and memoirs) does not exceed 40% [13].

The most often used in research projects collection of the RNC resources is a sub-corpus "Main corpus" (http://www.ruscorpora.ru/search-main.html). At the moment the sub-corpus comprises 265 millions of word tokens. It includes texts representing standard Russian and is subdivided into 3 parts, each of which has its distinguishing features: modern written texts (from the 1950s to the present day), a sub-corpus of real-life Russian speech (recordings of oral speech from the same period), and old texts (from the middle of the 18th to the middle of the 19th centuries) [13].

There are also other types of sub-corpora within the Russian National Corpus, for example: sub-corpus of Dialectal speech, sub-corpus of Russian Rhythmical structures, Syntactic sub-corpus, Russian mass-media collection, sub-corpus of Poetry, Parallel text sub-corpus, Multi-media sub-corpus, etc. All these sub-corpora were designed to answer special questions of the modern language theory and practice. On the

first hand, questions dealing with inter-disciplinary knowledge in social sciences and humanities.

III. Linguistic and Cultural Research Projects Using Modern Linguistic Corpora

As has been already mentioned earlier, Corpus linguistics is challenging to answer questions relevant to interdisciplinary endeavors in the field of translations and literary studies, language teaching and learning, as well as a whole host of domains within communication studies. However, hot debates about different value and values attached to corpora are still going on within the professional society [14]. In the current paper we will present some examples of interdisciplinary students' research projects using Russian National Corpus. All the work has been executed according to master programs' curriculum at the Faculty of Philology of Moscow State Lomonosov University.

One of the ongoing students' research projects deals with study of trends in perception and evaluation of foreign citizens, or foreigners, in the Russian society changing through time. It is well known that cultural stereotypes in assessing a foreigner (a stranger) are peculiar to every national society, or ethnic group within a society [15]. Investigation of various attitudes towards a self and towards a foreigner (a stranger) is thus of special interest. Nowadays they become even more critically important because of intensive migration processes going on all over the world.

To describe evaluation of social attitudes towards this or that social event has been always of special interest for local government as well as for other political forces. Sociological survey, or opinion poll has been traditionally considered as most reliable source to answer the question. However, great deal of information could be also extracted through analyzing everyday language practices. Before the 'Corpus era' linguists were to make search for appropriate contexts through random sampling of available texts and other documentary resources with no guarantee of completeness, sustainability, relevance and reliability of information found. They were wasting time browsing texts, writing down conversations (samples of conversational speech), lurking underhand communicating people's behaviors, and taking notes afterwards. With the new Corpus linguistics methodology the situation has changed radically.

Russian National Corpus enables to perform all kind of simple and advanced search of textual context for key words or for any undefined lexical item within fixed grammar or stylistic category in the chosen sub-corpus data. For example, within investigation of perception peculiarities in the Russian national collective conscience and even collective unconsciousness of a subject with English ethnical background a student made a contextual vocabulary search through several RNC databases. In this research project one of the experimental targets has been investigation of selected set of special Russian collocations, that is: any qualitative adjective that appears in a text in the vicinity (close proximity) of the test word – a noun (in the current project the keyword has been the word 'англичанин', or Englishman). Screenshot of the search template of RNC's main sub-corpus data with appropriately filled fields is presented in Fig. 1.

RNC's search results provided us with valid collection of collocations of the type "adjective + noun" (where noun has been 'Englishman', or – as in other project's areas – 'German', 'French', etc.), as well as with all relevant adjacent contexts, taken from data in the sub-corpus of Russian texts of various genres and types covering time period of 1800 – 2014. The total amount of phrases with the searched phrase was up to 983 items in 653 documents (screenshot of the RNC's web-page see in Fig. 2).

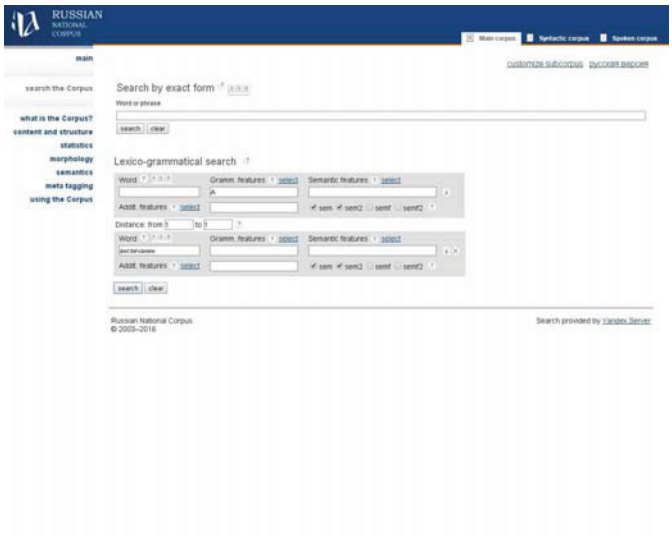


Fig. 1. Example of RNC's search template for collocation 'any adjective + Englishman'.

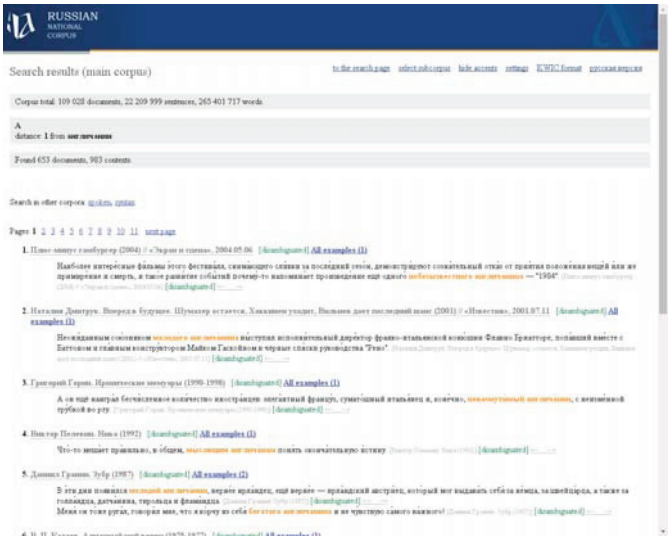


Fig. 2. Search results for contexts of collocation 'any adjective + Englishman'.

All collocations provided by the corpus containing an adjective placed before the word (noun) 'Englishman' / 'Englishmen' enabled primary analysis of qualitative characteristics of a person of English nationality in the perception of Russian people changing throughout more that

two hundred years period. However, our rough analysis evidences that Russian people's evaluation of a typical Englishman turned to be rather stable. Main experimental data on Englishman's representative qualities assessed by Russians, as provided by the RNC's texts collections, can be divided into three broad qualitative categories: Appearance, Character and Manners. The typical adjectives connected with the Englishman in Russian texts are as follows.

Category I. Appearance: blond, meagre, blue-eyed, heavily-built, stiff, a long Englishman with a horse profile, yellow-teeth Englishman, sinewy, fat, bony, red-cheeked, inconsistent, broad-shouldered, robust, tall, red-haired, sporty, dry, with gaunt feet, gaunt, slim, thin.

Category II. Character: gamblesome British, passionless, haughty, proud; asocial Englishman, arrogant, ironic, crafty, silent, impassive, optimistic, careful, brave, pedantic, self-satisfied, ambitious, pragmatic, reserved, sentimental, gloomy, severe.

Category III. Manners: obtrusive, mannerly, polite, conservative, amiable, exasperating, prudent, strait, intelligent, astute, sagacious, meticulous, businesslike, decorous, offish, crazy, oddish, stylish.

Similar procedures have been performed to the set of collocations extracted from the RNC with the same key words' grammatical properties and syntactic structures, nominating people of various nations (German, French, Tatar, Chinese, etc.). Results of the project could be used in translation practices, cultural and political studies, social history, as well as in many fields of modern and historical linguistics.

IV. Interdisciplinary Value of Corpus Linguistics

As Guy Aston, Professor of English Language and Translation at the University of Bologna writes: "One of the interesting things about the wide range of areas in which corpus linguistic methods are employed is that if you are interested in one application area, you can often learn things from other areas: there is a lot of scope for interdisciplinary in applied Corpus Linguistics" [14].

One of the most valuable capacities of the Russian National Corpus is that it has embedded algorithms for statistical analysis applicable to any type of data. Therefore our current project's experimental samples provided us not only with reliable and objective data on word meaning and types of texts (genres) with experimental key words collocations' lexical and grammatical information, but also with valuable information on the frequency of occurrence of the experimental token ('Englishman') in all the Russian texts published during the years 1800 – 2014.

In the main sub-corpus of RNC search key word 'Englishman' has been found in 3525 documents, 12115 times in total (see Fig. 3). Results of statistical analysis of frequency of occurrences of the key word under investigation (Russian equivalent of the word 'Englishman') are provided as frequency values on a screenshot of the RNC's web-page in Fig. 4. (English version of the web-page is currently not available).

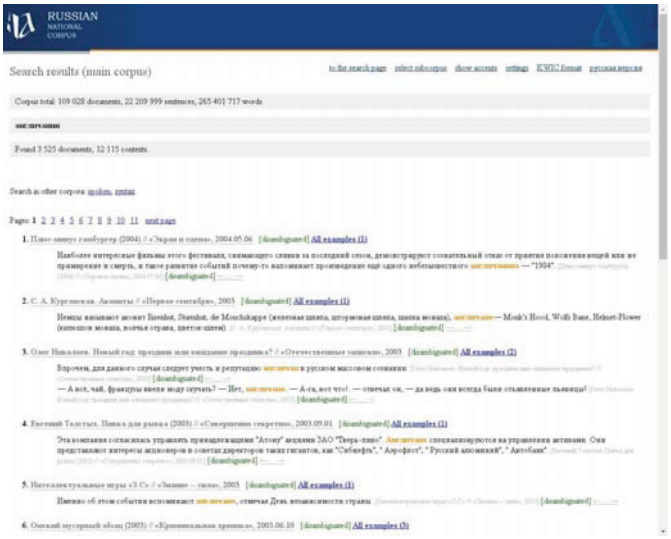


Fig. 3. Search results for contexts of the word 'Englishman'.

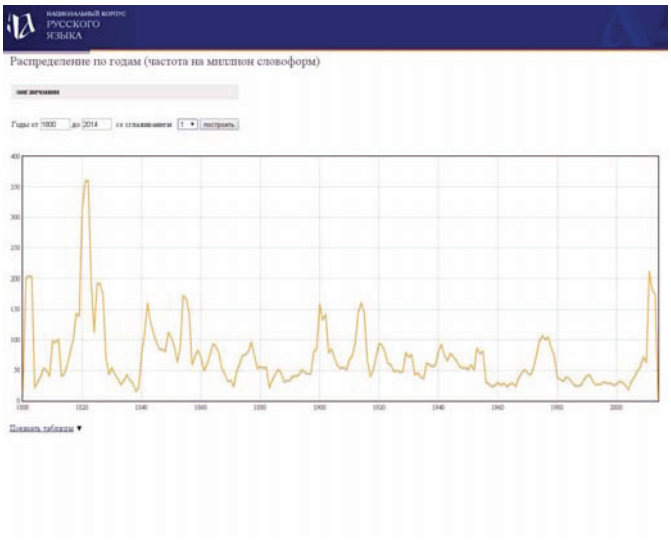


Fig. 4. Frequency of occurrence of the word 'Englishman'.

Statistical analysis of frequencies of occurrence of the word 'Englishman' in Russian text reveals strong correlation of picks frequency values and certain events in bi-lateral relationship of the Russian Empire (Russian Federation) and Great Britain in 19th century. The peak values of frequencies coincide with most significant events in these relationships such as the Napoleonic Wars (1803–1815) where Russia and Great Britain were allies in the coalition against France, and Greek War of Independence (1821-1829), also known as the Greek Revolution. In the revolt against Ottoman Empire Greeks were assisted by the Russian Empire, Great Britain, Kingdom of France, and several other European powers. Another peak value on the graph around the year 1854 corresponds to the Crimean War when Russian Empire and Great Britain were enemies. This continuously balancing interrelations of Russia and Britain in the 19th century is often defined in international politics as the "great game" between Russia and Britain [16],

and was clearly indicated in the full assembly of texts of the time.

Other interesting examples of interdisciplinary value of the Corpus capacities could be illustrated with the results from similar student's research work. The target Russian word in the student's project was a noun 'кебаб' ('kebab' in English). In the Russian language the key word is a loan word from Turk languages that has been recorded sufficiently frequently in the Russian texts at the end of 18th – beginning of 19th centuries. We present here results of general statistical analysis of co-occurrences of the key word 'kebab' (i.e. its frequency values changing in time) on a screenshot of the RNC's web-page (Fig. 5). (English version of the web-page is currently not available).

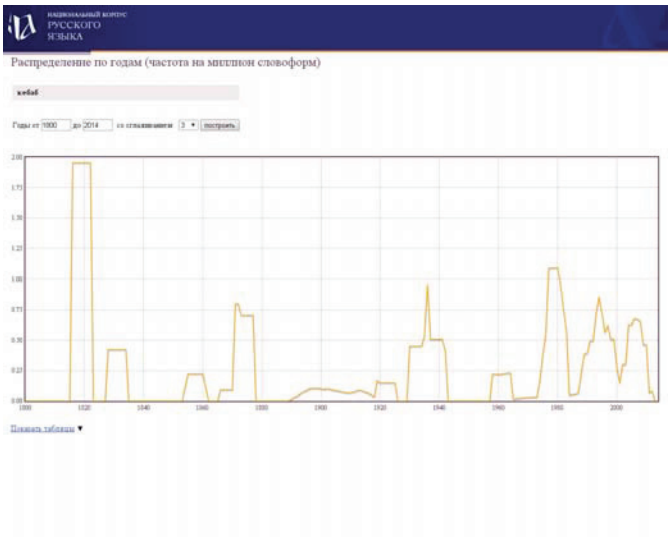


Fig. 5. Frequency of occurrence of the word 'kebab'.

One can easily note that the frequency of the key word's occurrence is extremely uneven in the texts within two centuries' period. Absolute peak of occurrences in texts falls upon the period from 1816 until 1822 years. It corresponds with the main trends of the Russian foreign policy in Caucasus. Namely, it's the Gulistan Treaty between Russia and Persia in 1813 that has launched a long process of reunification with Russia of the transcaucasian states (Georgia, Azerbaijan), and beginning of the initial phase of series of the Caucasian wars.

Thus, one may state that the language corpus data strongly supports the basic idea of a language performance as universal 'mirror' of intellectual, spiritual, political and social life of a nation. The corpora data therefore should be considered as one of the most objective and reliable source of information not only in Linguistics (primarily in investigation of human language and human cognition) but also in multiple areas of interdisciplinary research in Humanities and Social Sciences.

V. CONCLUSION

Our initial overview of the Russian corpus linguistics and some examples of the problems that could be resolved with language corpora data supports exceptional role of the language corpora as unique reliable source of objective information on the actual state of language practices. As the most faithful 'mirror' of political, intellectual and spiritual life of a nation during current state and in historical perspective, linguistic corpora could be also applicable as innovative research methodology in wide variety of research projects in Social Sciences and Humanities. Corpus linguistics provides new theoretical approach for the whole field of humanitarian research and education. We consider thus digital literacy in using language corpora facilities for research in Social Sciences, and particularly in language learning, as one of the primary targets in professional formation within wide variety of disciplines.

REFERENCES

[1] F. de Saussure, Course in general linguistics, 3rd ed., R. Harris, Trans. Chicago: Open Court Publishing Company, 1986.

[2] M. A. K. Halliday, "Quantitative studies and Probabilities in Grammar," in Data, Description, Discourse. Papers on the English Language. In Honor of John McH. Sinclair, M. Hoey. L., Ed. London: HarperCollins Publisher, 1993, pp. 1-25.

[3] D. Summers, "English in the Raw," Modern English Teacher, Vol. 1, No. 4, 1992, p. 14.

[4] J. Mukherjee, "Corpus linguistics and language pedagogy: the state of the art - and beyond," in Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods, S. Braun, K. Kohn & J. Mukherjee, Eds. Frankfurt am Main: Peter Lang, 2006, pp. 5-24.

[5] T. McEnery, R. Xiao, and Y. Tono, Corpus-based language studies: An advanced resource book. Taylor & Francis, 2006.

[6] What is the Corpus? URL: http://www.ruscorpora.ru/en/corpora-intro.html . (*On-line resource*)

[7] N. Gvishiani, English on Computer: A tutorial in Corpus Linguistics. Moscow: Vyshaja Shkola, 2008, p. 178.

[8] V. Plungian, "Linguistic corpus as a tool and as ideology: several lessons from modern corpus linguistics," Russky yazyk v nauchnom osveshchenii, No. 16(2), 2008, pp. 7-20.

[9] Online corpus of Russian Sign Language, URL: http://rsl.nstu.ru/site/index/language/en. (*On-line resource*)

[10] M. Kunilovskaya and T. Ilyushchenya, "Russian Learner Translator Corpus in translator training," 4th Conference Using Corpora in Contrastive and Translation Studies, United Kingdom, pp. 32-33, 2014.

[11] Corpus of Russian Student Texts, URL: http://web-corpora.net/CoRST/search/?interface_language=en (*On-line resource*)

[12] A. Shayakhmet, "The use of CHILDES in children's Kazakh-Rissuain bilingualism study," International Conference Corpus Linguistics-2015, S-Petersburg, 2015, pp. 443-450.

[13] Content and structure of the Corpus, URL: http://www.ruscorpora.ru/en/corpora-structure.html (*On-line resource*)

[14] V. Viana, S. Zyngier, and G. Barnbrook, Eds. Perspectives on corpus linguistics, Vol. 48. John Benjamins Publishing, 2011.

[15] I. Kobozeva, "German, Englishman and French: investigation of stereotypes of national characters with analysis of ethnonyms' connotations," Moscow State University Bulletin, Series 9, 1995, pp. 102-117.

[16] W. Thompson, "Principal Rivalries," The Journal of Conflict Resolution, vol. 39, no. 2, 1995, pp. 195-223.

Interactive Whiteboard in Motivation of Pupils of Primary School

Gabriel Bánesz, Danka Lukáčová, Lukáš Vaněk, Peter Hrubý
Department of Technology and Information Technologies
University of Constantine the Philosopher
Nitra, Slovakia
gbanesz@ukf.sk; dlukacova@ukf.sk; lukas.vanek@ukf.sk; peter.hruby@gmail.com

***Abstract*— In our report we want to present the results of a survey focusing on the use of interactive whiteboard technology in the subject called technology at primary schools to increase students' motivation in teaching technical subjects. Technology is a very specific subject in primary schools, because it combines the teaching of theoretical information and practical skill practicing. The research was to find out or check whether pupils of experimental groups who were using an interactive whiteboard during teaching, will be more motivated to learn than students of the controlling groups. We used Kozéki's questionnaire to detect pupils' motivation to motivate students to learn. We found out that in the experimental group the motivation in the affective field (extrinsic motivation) is positive and predominant. This is demonstrably more powerful than the motivation of the control group, which belongs to the field of the cognitive effects, also with positive character. The research results are useful in the planning of teaching in terms of allocating the interactive whiteboard into the teaching process.**

***Index Terms*— education, interactive whiteboard, multimedia technologies, teaching practice, questionnaire.**

I. Introduction

Interactive whiteboards are included among the relatively new technical means in teaching. In terms of operation it is a large-scale touch screen (from a technical point of view there is a big difference between the interactive whiteboards and touch screens) with the dimensions of a school board, which together with related software allows you to create interactive lessons, where a multimedia application is running on the desktop board that can be interactively altered. Instead of the interpretation of the teacher this increases clearness, and also the students are involved into the activities, thus increasing their active participation in the teaching process. These funds are commonly used for teaching general - education subject now.

Subject called technology taught in primary school is a very specific subject, as it combines theoretical knowledge with the practical skills of students [1]. By 2015, the scope of teaching technology in primary schools in Slovakia was very small - one lesson per week in any class on the secondary school, the motivation of students was on a very low level against the subject. We watched mainly the increase of the level of teaching the subject technology and motivation of students to the subject of technology by the introduction and the use of interactive teaching aids in the teaching of a technology. The application of this method into the teaching requires a preparation of a teaching model based on the use of interactive teaching aids, as well as the preparation of interactive teaching resources for its implementation to the teaching process. Hence the aim of the research is: to determine whether students of the experimental groups will be motivated to learn better than students of the control groups. We have formulated a hypothesis to reach the aim:

H: Students of the experimental group taught with interactive teaching aids will reach a generally higher level of motivation in the questionnaire to measure the level of motivation than control group students taught in the traditional way.

We had to take into account several criteria in the choice of schools, where we conducted the research:

- The teaching of the subject of technology in the seventh, eighth or ninth grade.
- Equipped interactive boards.
- Availability (possibility of personal participation in the experimental teaching)
- Willingness to cooperate (and to provide an interactive board).

Planning and preparation of research consisted of several parts. One of them was reduction the choice of primary schools according to the specific criteria. We have created a list of primary schools in Nitra and in the nearby region from the official lists of schools from the availability point of view. Based on the statistical information we assumed primary schools that use interactive whiteboards, as well as teaching technology according to teaching plans and curricular. The selection was further carried out in the agreement with the school management and teachers. For better friendliness we chose the tactics of personal visits to schools and personal contacts with the Director or representative and the teachers, thus we have met with a great willingness to cooperate. We moved further in the selection of those schools that meet other criteria such as a sufficient number of students to form two groups for the whole school (preferably two parallel classes), teaching technology and a willingness to let us implement teaching on an interactive whiteboard. The research

participated 144 pupils in overall, who were divided based on the results of a pre-test into an experimental (73 students) and a control (71 students) group.

II. Research methods

Based on the curriculum we have chosen a curriculum in the field of electrical engineering, we focused on a simple electrical circuit and its elements, switching elements, and serial and parallel connection of elements in particular. The curriculum was chosen for its theoretical nature, to be able to compare conventional classroom teaching with a similar classroom teaching updated with the support of interactive teaching aids. At the same time, it was a challenging curriculum for pupils in which they have a long low success rate. On the base of curriculum, the preparation for the lesson was processed in detail. Eight preparations were created for eight lessons totally.

The same teacher taught both the experimental and control groups, with the same content and scope of the curriculum.

Identification of the variables:

1) Independent variable - the method of implementation of learning - learning through interactive teaching aids
2) The dependent variable - attitudes of students

Intervening variables:

- The same teacher
- The same content and scope of curriculum.

As a tool for identifying the pupils' motivation, we used Kozéki's questionnaire to motivate students to learn. This questionnaire was created by the Hungarian artist B. Kozéki focuses the teaching activity of students in the school. It is barely used in researches but it was used by authors, for example J. Lokša, I Lokšová or M. Zelina [2], [3]. This questionnaire is divided into three areas representing the affective motivation area (A) - extrinsic motivation, the area of cognitive motivation (K) - intrinsic motivation and motivation area of effect motivation (M) - internalized motivation. Each band can be characterized as follows:

Affective area (A) - themes arising from the student's emotional relationship to his environment and its members. Here are three sub-areas, and the emotional relationship with parents and the identification with their requirements and examples (A1), emotional relationship with the teacher, their example and maintain their trust (A2) and emotional relationship with the classmates and collective, social inclusion and collective recognition and popularity with others (A3).

Cognitive area (K) - themes arising from the possibility of developing knowledge and skills. Three sub-areas can be distinguish here as well. Here belongs the own desire for cognition and faith in own resources (K1), a desire for knowledge and improvement of skills, an interest in knowledge, intellectual motivation (K2) and motivational impact of the activity and curiosity, the joy from discovery (K3).

Effects area (E) - themes arising from the social standards and expectations of the environment. Here are also three sub-areas, namely self-esteem and motivation to learn and the effort of the pupil in order to achieve specific performance and endeavor to meet expectations (E1), the effort to serve the collective norms, incentives based on the efforts and values of the collective (E2), upheld in social norms values and ideals (E3).

The questionnaire consists of 81 items - verdicts. Each statement relates to one of the stated sub-areas – all of them include 9 statements (to each area belongs 27 statements). All statements are indicated by the number of points, and -1 point if it is a negative motivation, 0 points if it is a neutral motivation, and +1 point if it is a positive motivation. The statements from these three point areas are evenly allocated to each motivational area and sub-area, therefore in each region there are 9 negative, 9 neutral and 9 positive statements, in each area and sub-area are 3 negative, 3 neutral and 3 positive statements. The tested person gets the list of these statements and should mark only those statements with which he agrees. During the evaluation the points are counted in individual areas and sub-areas (in each area from -9 to +9 points in each sub-area -3 to +3 points can be obtained). This will yield the scale evaluation of motivation in different areas and sub-areas. Based on these results, it is possible to determine the predominant motivation area of learning for each student (or two if they are equivalent), depending on which area the most points were obtained and also whether it is a positive (joy, success ...) or negative (fear...) motivation.

By Kozéki's survey we monitored the type and level of motivation of students of experimental and control groups for learning. The questionnaire result from each pupil was triplet figures: the number of points for the affective area, the number of points for the cognitive area and points for effective area. In each area from minus 9 points, through zero by plus 9 points could be reached. More Points (whether positive or negative) represent a stronger motivation, while a negative number represented a negative and a positive number positive motivation. Zero means zero - no motivation.

During the evaluation of the motivation of each student we seek for maximum - the largest absolute value of between three scores of different motivational areas. So we identified the dominant motivation area of each pupil. If the student is expected in two or all three equally large absolute values, we also identified several areas as dominant. From the number of motivational areas, we identified the dominant motivating area for the experimental and the control group. Then we calculated the percentage of the maximum levels of motivation for the experimental and the control group, which we determine the strength of the motivation in the group.

In the experimental group we have identified the affective area as the most dominant (a total of 32 cases). Students further identified the cognitive area the most common times (30 cases), the fewest cases were in the effective area (26 cases).

The control group had identified the cognitive and effective area in the most cases (both with 30 cases), the affective area was signed fewer cases (23 cases). The comparison can is seeable from the graph.

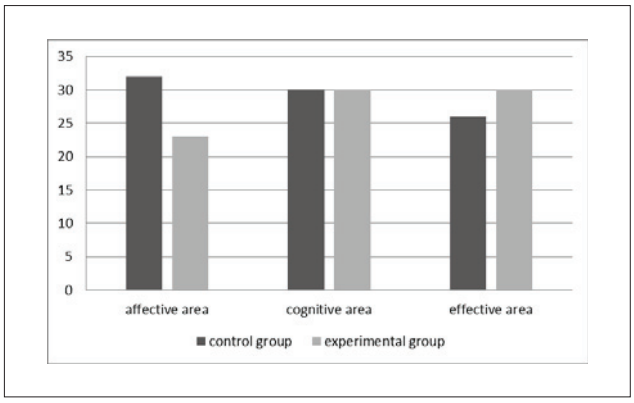


Fig. 1. Comparison of results in the controlling and experimental group.

From the graph we can see that in the experimental group affective motivation prevails and in the control group the cognitive motivation and effective motivation.

From the absolute values of the most dominant areas of students, we calculated the average level of motivation of each group. We found that students of the experimental group received a total of 3.1 points out of nine (the higher the number, the stronger motivation), the pupils of the control group gained 2.37 points. This shows that the experimental group students were more strongly motivated.

Because there is a positive motivation (based on joy, expectation of good) and negative motivation (possibly based on fear, worry), we calculated the values of the dominant areas of students of each group and the total value of motivation (we took into account the polarity of motivation), to find out whether the motivation for the individual groups is positive or negative. The experimental group gained 1.09 points (from -9 after +9), the control group 0.71 points. This shows that the motivation was positive for both groups (number is greater than zero).

By the use of Kozéki's questionnaire we managed to prove that the predominant motivation of the experimental group in the field of affective motivation (extrinsic motivation) is positive. It is demonstrably more powerful than the motivation of the control group, which in turn is within the cognitive an effective motivation, also positive. We found that students of the experimental group received a total of 3.1 points out of nine, while the pupils of the control group gained just 2.37 point. This shows that experimental group students were more strongly motivated. From these results, we confirmed the hypothesis.

III. Discussion

The positive effect of interactive teaching aids for teaching is describes by many other authors in their articles and works.

For example, Dolný in his article describes the benefits of using interactive teaching aids as increasing motivation, deepening the acquired knowledge of the subject matter and positive use of ICT in the educational process. These results are compared with the information obtained by us and they confirm each other, both in terms of results obtained through Kozéki's questionnaire and the work itself with interactive whiteboard and with computers [4].

Martinková [5] addressed in her work the use of video recordings in teaching technology in the context of the interactive whiteboard. On the basis of the evaluation of teaching in which the video was used as a teaching aid claims that the pupils reflected greater effort, active and creative approach to the acquisition of knowledge, besides the strong motivation of students particularly played a role. These discoveries agree with the results of our research.

Dömischová [6] dealt with the views of teachers on the use of interactive boards in teaching languages, where primary school teachers defined its positives. They state the biggest positive in the strong motivation of students. In this way, although in other subjects, it indirectly points to the audited incentive effect of the use of interactive teaching aids. It also points out that the pupils lose the attraction by regularly used interactive whiteboards.7

Tóblová and Soviarová [7] deal with the benefits of using interactive whiteboards in education in their work. They present the results of a survey in which the noted that work on the interactive whiteboard motivates students to try to achieve better results and to further learning, and it also increases the level of their knowledge in the field of information - communication technologies.

In his case study, Wood and Ashfield [8] write about the creative teaching of literacy and math through interactive whiteboards, where they also describe its motivation influence to the activities of students in autonomous learning. This shows that, internationally, we can find our discoveries in the field as confirmed.

IV. Conclusion

In our research, we have compared the experimental and control group motivation through Kozéki's motivation questionnaire on a sample of 144 primary school pupils. The experimental group was taught by using an interactive whiteboard, while the control group did not use an interactive board in teaching. We found that the affective motivation proved to be the dominant motivating area of the experimental group, while with the control group the cognitive motivation and effective motivation were equally dominant. The experimental group proved to be strongly motivated in determining the intensity of motivation. The polarity of the resulting motivation for both groups was positive. The results of this research can be used in lesson planning for technology on primary schools. The work with an interactive whiteboard is appropriate to include as motivation for new information, or for new activities to be subsequently carried out by students.

Acknowledgment

Publication of the paper was funded from the project KEGA 021UKF-4/2014 Remote Real Experiments in School Practice.

REFERENCES

[1] J. Dostál, V. Prachagool, "Technology Education at a Crossroads – History, Present and Perspectives", In Journal of Technology and Information Education. Olomouc: UP, 2015, Vol. 8, Issue 2, DOI 10.5507/jtie.2016.006

[2] J. Lokša, I. Lokšová, "Analysis of pupils' motivation for learning", In Pedagogical revue, XLIV, 1992, č. 6, p. 438-454.

[3] M. Zelina, "Students' motivation to learn", In Journal of School Unified, XXXV, 1983, č. 4, p. 359-371.

[4] B. Dolný, "Multimedia teaching aid for thematic unit Technical electronics in the 8th year of primary school", In The mutual awareness - an effective way to development of scientific and educational activities. Nitra: UKF, 2008, p. 22 - 26. ISBN 978-80-8094-300-4.

[5] A. Martinková, "Use of interactive whiteboards in technical education and computer science", In The mutual awareness - an effective way to development of scientific and educational activities. Nitra: UKF, 2008, p. 55 - 59. ISBN: 978-80-8094-300-4.

[6] I. Dömischová, et al, "Interactive whiteboard in teaching at elementary school", In Modern education: Technology and Information Technology. Olomouc: UP, 2011, p. 36 - 37. ISBN 978-80-244-2912-0.

[7] E. Tóblová, A. Soviarová, "Interactive whiteboard and its contribution in education", Available: http://www.movz.upol.cz/files/articles/eva-toblova/toblova_e.pdf.

[8] R. Wood, J. Ashfield, "The use of the interactive whiteboard for creative teaching and learning in literacy and mathematics: a case study". Available: http://www.pgce.soton.ac.uk/ict/NewPGCE/PDFs/Tyhe%20use%20of%20IWB%20for%20creative%20teaching%20and%20learning%20in%20literacy%20and%20marthematics%20case%20study.pdf.

Expert Approach to Statistical Assessment of Education Quality: The Case of Azerbaijan

Sardar Shabanov
Scientific Research Institute of Economic Studies
Azerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
sardar.shabanov@unec.edu.az

Fariz Quliyev
MBA department
Azerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
f.guliyev@unec.edu.az

***Abstract*—We consider the quality of education from two points of view in Azerbaijan. The first point of view is based on the international program for student assessment - PISA, and the second one on the scientific and statistical materials of the State Student Admission Commission of the Republic of Azerbaijan. The distribution law of the applicants to universities at the entrance exams was investigated. Final recommendations were made on improving the quality of education in Azerbaijan.**

***Index Terms*—Human capital, quality of education, PISA, normal and chi-square distributions, Bloom's taxonomy.**

I. Introduction

Education holds ones of the important parts in the theory of human capital [1]. In general, it is accepted that an educated person has more opportunities for success compared to an uneducated one. From this perspective matters of quality is brought forwad in the article. Literacy has been remarkably increasing in the world [2]. Countries with low levels of human development index has an average of 4.2 years of study time in high schools whereas in countries with high levels of human development index the indicator is 11.7 years. In Azerbaijan, this indicator is 10.4 years.

It is obvious that direct investments play a crucial role in the development of any sector. In this aspect, the report on the execution of the 2014 state budget for education expenses [3] by specializations play an important role. Current expenses on education was AZN 1 553.9 million in local currency, which is more by AZN 116.1 million or 8.1% compared with that of 2013. Personal weight of the education expenses in state expenses was 8.3% which is more by 0.8% compared to 2013 figures. Combining the funds set aside by state budget education expenses made 10.4% of the expenses of the state budget or AZN 1 943.6 million in nominal terms. The distribution [3] of expenses by directions of the ministry of Education of Azerbaijan Republic consists of 5 blocks. The major part in education expenses contains incomplete and secondary education block. Costs in this block contains 52,9% of all the costs in this category. The second largest block contains 19,8% and directed to applied researches and other services in education. The third largest block by volume is in post-secondary category with 10.3% of the total cost. The preschool and primary education holds 10.2% of the total consideration. The block in technical-vocational schools and liceums has 4.3% of the total share. Finally, the last block in the list is for special purpose and boarding schools which has 2.5% of the total costs. Note that, all the funds have been expensed as assigned.

II. State Of The Problem

In our understanding qualitative aspects of education must be paid attention alongside with its quantitative indicators. In this sense, a remarkably large part of the expenses must be committed to teachers who are the driving force in all spheres of education. Special attention must be paid for the preparation of quality teaching staff. A comprehensive and effective action plan must be prepared to improve the vocational training of the teaching staff. The ultimate opportunity here should be a matter of robust logic. For that purpose, it is suggested that trainings should be organized to solve logical problems. ***Foreign language*** (specifically english and russian languages) trainings must be organized, ***modern methods in teaching and trainings*** must be taught and use of ***computer and multimedia*** resources must be arranged. In order to enable a convenient use by teaching staff ***exemplary video-classes*** on various topics must be placed in the web site of the Ministry of Education of Azerbaijan Republic.

The share of education in GDP is 2,46% in Azerbaijan. A global comparison of countries demonstrates that this indicator should be increased.

TABLE 1. Share of Education and GERD on GDP, %

Contry	Education	Gerd
Lesotho	13.98	0.01
Cuba	12.84	0.47
Norway	7.37	1.66
Finland	7.19	3.31
USA	5.22	2.73
EU(middle)	-	1.91
Russia	4.15	1.11
I.R. of Iran	3.06	0.75
Turkey	2.86	0.95
Kazakhstan	2.89	0.17
Azerbaijan	2.46	0.21
Armenia	2.40	0.24
Georgia	1.98	0.13

The level of literacy is substantially high in Azerbaijan. Such that, based on statistical information [4, p.12] the level of literacy in Azerbaijan is 99.8%. Quantitative indicators of the level of literacy is provided in the table below.

TABLE 2. LITERACY LEVEL IN AZERBAIJAN

Level of education	2012	2013	2014
Population with secondary and post-secondary education older than 15 (per 1000 people)	969	969	971
including Post-secondary	125	125	126
Vocational	85	85	85
Secondary	629	629	630
Secondary general education	130	130	130

The structure of the literacy level per 1000 people in Azerbaijan for 2014 makes up 126 people with post-secondary level of education, 85 people with vocational level of education. These indicators are roughly twice below the levels of leading countries in the world. In Scandinavian countries, the number of people with post-secondary education is very high, ex: In Norway, almost all people with 25 years and over are educated. A comparative analysis shows that a huge platform exists to increase all the quantitative indicators at all levels of education only a special program must be prepared and developed for this purpose. An existing practice of arranging two year institutions are used in order to direct students with relatively law potential to universities and teach them concrete specializations. As a result, the number of citizens who have mastered specific specializations, gained a two year education and who earned a post-secondary education of some degree can be increased.

III. DECISION

The quality of secondary level education in Azerbaijan has been viewed from two aspects. One of this aspects is based on international comparison while the other one is based on uniform examination results organized by State Student Admission Commission (SSAC) of the Republic of Azerbaijan.

At first, let us look at the PISA competition which has taken place in the year of 2006 and 2009. Fifteen year old teenagers from Azerbaijan has participated in PISA 2006 and 2009 competitions joined by 65 countries [7-9]. More than 5000 students from Azerbaijan have joined and participated twice in both competitions. Grading was carried out in three directions: reading comprehension, mathematics, and other sciences. High school students from Azerbaijan have scored 353 points in reading comprehension from possible 600 points and ranked 54th in 2006. Students ranked 32nd having scored 476 points and 55th having scored 382 points. On average it was ranked 54th place among 64 countries.

In 2009, PISA competition they scored 362 points in reading comprehension by advancing 9 points, 431 points in mathematics by lagging -45 points, and 373 points in other sicences by lagging -9 points, thus on average ranked 64th place in overal ranking. The outcomes of global grading demonstrates that the secondary education has lagged behind 2006 through 2009. Unfortunately, it should be noted that Azerbaijan has not participated in 2012 and 2015 PISA competitions.

The second matter of qualitative approach in education is based on the method employed by SSAC. Based on the analysis of statistics provided by SSAC it was disclosed that 64.7% applicants has scored less than 200 points which means a failure (a grade of "2").

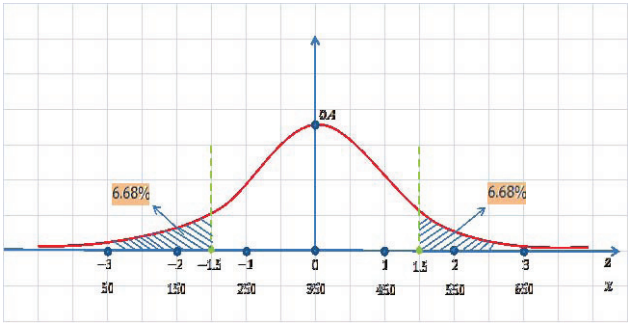


Fig.1. Personal weight of applicants with a score of less than 200 points, %

Let's suppose that points gathered by applicants in final exams fits the law of normal distribution [6].

$$F(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt \qquad (1)$$

$$F(z\le -1.5)=F(z\ge 1.5)=0.5-F(z\le 1.5)= \\ =0.5-0.4332=0.0668=6.68\% \qquad (2)$$

In this case, the value of this indicator is calculated to be 6,68%. If our supposition holds true then the real value (64.7%) would be roughly 10 times bigger than the theoretical value (6.68%). Yet a deeper analysis demonstrates that the distribution of this random variable is not normal.

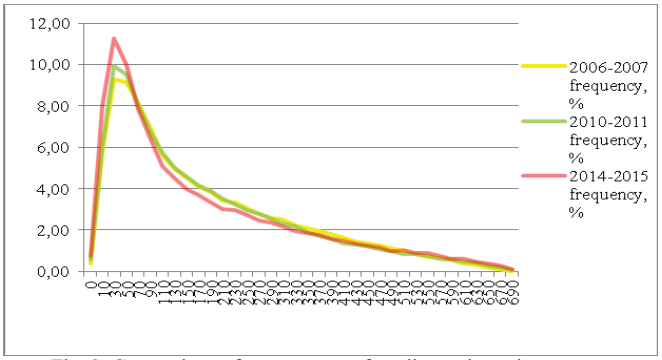


Fig. 2. Comparison of acceptances of applicants in various years (calculations by the authors)

Speaking of the Fig.2 one could suppose that this random variable has a characteristics of chi-square distribution.

TABLE 3. THE STATISTICS OF THE DISTRIBUTION OF SCORES IN FINAL EXAMS (CALCULATIONS BY THE AUTHORS)

Statistical Characteristics	Entrance Year		
	2006/2007	*2010/2011*	*2014/2015*
Estimation	171.53	171.53	171.53
Standart Error	223.53	223.53	223.53
Mode	30.00	30.00	30.00
Median	130.00	130.00	120.00
Count of Applicants	103 196	103 185	92 643

For every examination year values of estimations, median and modes that are remote from each other once more demonstrates that the process under research does not fit the normal distribution law. In addition, ([5, p.41]) applicants with weaker than average scores dominates the list.

In 2014/2015, 43,89% of the applicants who participated in final entrance examinations scored 0-100 points and 20,76% scored 100-200 points. Applicants who scored between 200-300, 300-500 and 500-700 points points made up 13,61%, 15,73% and 6,01%, respectively. Applicants in Azerbaijan section was formed of 84739 people and in Russian section 7904 people. The median in Azerbaiajn sector was 116,92 points, in Russian section 186,43 points.

According with the learning issues of Bloom's taxonomy [11] (look at the fig.3) majority of our high school students can not abstain from memorizing and face difficulties in reading comprehension. So, the students again act indecisively in the third stage of learning, which is application and subsequent analysis. Most of them do not even reach the next step of synthesis and assessment. For that reason, high school teaching should not consist of memorizing or learning the facts by heart. It is important to move from memorization based surface learning to comprehension based learning.

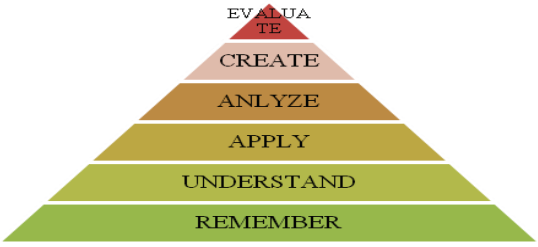


Fig. 3. Bloom's Taxonomy of Education

IV. CONCLUSION

Almost each students in higher education are high school graduates. Considering the Bologna scoring which includes 0 to 50 points as an "F" mark (failure) then the entrance score to higher education institutions should not be less than 351 points in general (700/2+1). So, 351 points should be determined as the floor level in entrance exams excluding specialisations requiring special talent (performing, painting, sports, etc.). If specialisations requiring special talent are excluded there are only 15 operating universities out of 39 in Azerbaijan which have an average entrance score of 351 points as a minimum level ([7]).

Azerbaijani youth study in schools with good facilities and condition which means that a lot of work has been done for the material-technical and logistical base in Azerbaijan. We believe that the real salary (AZN 170-190) of teachers in high school should be raised substantially and considered that the average salary in Azerbaijan is AZN 455.

Part of the applicants who are prepared less and accepted to higher education institutions carry these drawbacks to the same higher education institutions. Yet the quality of graduates from higher education institutions is firstly measured by the requirements of labor market which excites a serious deal of reasearch in order to advance certain ideas in this sphere.

In our view, regardless of which fields of economy takes priority, the education services sector should have a high consideration. Professional, vocational, and secondary special education should be adaptively regulated with respect to current and future demands of the market.

ACKNOWLEDGMENT

Authors have numerously discussed the conclusions in the scientific seminar of the Scientific Research Institute of Economic Studies. We would like to express the deepest appreciation to each and every member of the seminar for the valuable notes provided to us, especially to Prof. Y.H.Hasanli and Prof. R.R.Guliyev who are the scientific supervisors of the seminar.

REFERENCES

[1] G. Becker, "Investment in Human Capital: A Theoretical Analysis", Political Economy, vol.70, Issue 5, Part 2: Investment in Human Beeing, pp.9-49, Oct., 1962.

[2] Social Development Indicators/ Center for Sustainable Systems. University of Michigan. 2014. "Social Development İndicators Factsheet". Pub. No. CSS08-15. October 2014. http://css.snre.umich.edu.

[3] The Report on the Implementation of the State Budget of the Azerbaijan Republic for 2014 Year. Baku, 2015.

[4] Education, Science and Culture in Azerbaijan. 2014. "State Statistical Committee of Republic of Azerbaijan", Baku, 2014.

[5] Statistical Yearbook of Azerbaijan 2014. "State Statistical Committee of Republic of Azerbaijan", Baku, 2014.

[6] D. Lind, R. Mason, W. Marchal, Basic Statistics for Buisiness and Economics. McGraw-Hill., 3rd ed., 2000.

[7] Students Admission 2014: Results. Abiturient journal, № 12, SSAC, Baku, 2014.

[8] Quality of learning outcomes gives an indication of the level and distribution of countries' student performance in the subjects Mathematics, Reading, Science and Problem Solving. PİSA 2006, http://pisacountry.acer.edu.au/

[9] PISA (The Programme for International Student Assessment) 2006: Science Competencies for Tomorrow's World, Survey, OECD 2007, http://www.oecd.org/pisa/pisaproducts/39725224.pdf

[10] PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I), OECD 2010, http://www.oecd-library.org/

[11] B.S. Bloom, (Ed.). M.D. Engelhart, E.J.Furst, W.H.Hill, D.R. Krathwohl, Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain, N.Y., Longmans, Green and Co., 1956.

Information Systems as a Component Part of Tertiary Education Quality Assurance

Alena Hašková
Faculty of Education
Constantine the Philosopher University in Nitra
Nitra, Slovak Republic
ahaskova@ukf.sk

***Abstract*— Quality of higher education has been considered as the basic priority of higher education development during last several decades in many countries of the world. Main principles of quality assurance were formulated in an official document known as European Standards and Guidelines (ESG). According to them information and information systems are two of the key components creating and supporting quality of each higher education institution. The paper explains importance of information with regard to different groups of stakeholders and at the same time it points out the issue of information systems with regard to the way in which people use the information. The key attention is paid to the presentation of a research focused on the use of information and information systems in higher education institutions in Slovakia and in a broader scope of European institutions, to support and enhance quality assurance processes. The paper presents also other activities which should implement the European quality assurance strategies also in the Central Asia.**

***Index Terms* — Higher education institutions, internal quality assurance systems, information, information systems, EGS – European Standards and Guidelines.**

I. Introduction

Higher education is the backbone of any society. Its quality decides the quality of human resources in the country. Today higher education stands at a cross roads. On one side there is a high demand for access to higher education, and on the other the quality is questioned. Which way to go in this situation? Which road to follow? Whichever a higher education institution decides to follow, it will find out in the end that the roads are only two sides of the same coin. To survive in the competitive world of globalization each higher education institution has to pay attention to the quality of the education it offers.

The core mission of higher education institutions (HEIs) still remains providing learners with the opportunity to learn and ensuring equipment of graduates with the knowledge and skills they need for their personal development, and also enabling the graduates to contribute to the development of the society they live in. What is rapidly changing is the scope of the stakeholders in university context and the requirements and needs the knowledge societies of the 21st century have on the university graduates coming to the labor market.

To promote the quality of higher education the European Network for Quality Assurance in Higher Education (ENQA) was invited by the ministers of the Bologna Process signatory states to develop standards, procedures and guidelines on quality assurance in the area of higher education ([8], ([9], [10], [17]). *The European Standards and Guidelines for Quality Assurance in the European Higher Education Area* (ESG, [11]) were adopted in 2005 and based on them, the higher education institutions (HEIs) were advised to introduce or develop their own internal quality assurance systems following these guidelines) .

To support their implementation into the practice of higher education institutions at both national and international levels, and to identify the main achievements and limitations of their implementation mainly into HEIs` internal quality assurance systems various activities, monitoring, researches have been carried out. Thereinafter the focus is put on some of the activities related to this area, in which Slovakia represented by Constantine the Philosopher University in Nitra has participated.

II. Information Systems in Quality Assurance of Higher Education Institutions

Standards and guidelines included in the document *The European Standards and Guidelines for Quality Assurance in the European Higher Education Area* (ESG, [11]) have been designed to be applicable at all higher education institutions and quality assurance agencies in Europe. They aim to promote mutual trust while respecting diverse national and institutional contexts and subject areas ([11], [6]).

Information systems and public information have been included as two of the basic seven components [11] into the Part 1 of the ESG. This confirms information and information systems to be important and significant factors on which HEIs quality assurance is based on. But different groups of HEIs stakeholders require different types of information. To develop information systems meeting the needs of all of the stakeholder groups (students, students´ parents, administrative staff, technical staff, HEI teachers, researchers, self-government bodies, funding bodies, policy makers and decision making

management, self-evaluation bodies, external evaluation agencies etc.) is a challenge for each HEI ([1], [5] .

As to the information systems, it is important to decide which information is needed for which stakeholder group and in which way it should be made available to it. Moreover less attention should be paid to the used technologies and more to the way in which people interact with information ([2], [3], [4]).

III. Research on Implementation of Information Systems in HEIs Internal Quality Assurance Systems

A. Background and Methodology of the Research

To identify main achievements and limitations of the implementation of information systems in higher education institutions´ quality systems was a part of the international research project IBAR - *Identifying barriers in promoting The European Standards and Guidelines for Quality Assurance at institutional level and making recommendations as to how these might be addressed* (IBAR) which the Slovak Republic participated in ([7], [15]). The other project partners were HEIs representing six other European countries, in particular the Czech Republic, Latvia, Poland, Portugal, the Netherlands and the United Kingdom.

The main task of this part of the IBAR project research was to analyze information systems (collection, analysis, use of the information, for what purposes, public/not public availability of the information) used within the HEIs in the project partner countries that are related to institutional quality assurance processes (i.e. both forms of information described in the ESG 1.6 and ESG 1.7 were taken into the consideration: ESG 1.6 - information related to comparison of the context of collection, analysis and use of specific information for HEI's internal purposes to provide self-knowledge on institution and effective quality assurance; ESG 1.7 - information related to public access to objective data on HEIs; see at IBAR http://www.ibar-llp.eu/results.html).

Hereafter, main findings of the survey from the case studies carried out at the 4 Slovak HEIs, which consequently served as an input for processing the international comparative study, are presented ([13], [14], [15]).

The research data necessary to elaborate the case studies of the four chosen public HEIs were collected by the means of content analysis, structured and semi-structured interviews with representatives of the top, middle and lower management, employees and students and questionnaire surveys of the same target groups.

The respondents of the interviews and questionnaire surveys from the each involved HEI expressed themselves to such questions as:

- Is there an overall information policy at the institutional level of the HEI? How is the national information policy/regulation referring to the provision of information about the HEI and by the HEI (to the society, various institutions, HEIs stakeholders, students reflected in it?
- What is the policy and practice in the HEI regarding collecting and use of ESG related information (ESG points 1 – 8) and for what purpose is it used (e.g. decision-making, corrective action, improvement, innovation)? Which kind of information in relation to which target groups (society, various institutions, HEIs stakeholders, students, employees, etc.) is the object of the information policy?
- Which other quality related information is contained in the policy and practice of the HEI?
- What are the instruments motivating the HEI to create and develop their institutional information policy, to develop their information systems, to make public some kind of information?
- What is the policy and practice in the HEI regarding publishing the information (ESG points 1 - 8). What is kept internal and what is published externally?
- To what extent is the provision of the information formalised and checked in the internal quality assurance processes? How effective is the institutional policy and practice for gathering and using information for improvements of
- What significant changes or improvements were implemented in the HEI information policy and practice during the recent five years?
- What is the institution (HEI/faculties/departments) satisfied with and what is it not satisfied with in relation to provision of information on quality of education? Is there any system in place to identify needs for changes and improvements in the HEI information practice? Are there planned any changes?

B. Research Results

The aim of the case studies carried out at the selected Slovak HEIs was to identify key barriers and good practices in relation to the implementation of the ESG 1.6 and ESG 1.7 [11] in HEIs' internal quality assurance processes.

The main findings resulted from the case studies, together with the relevant recommendations can be summarized as follows:

National policy on disclosure of information concerning HEIs

- Obligation of HEIs to disclosure information is defined in a large scope of legislation.
- Information received from HEIs is analysed and processed by the Ministry of Education, Science and Sport of the Slovak Republic (further: SR) respectively by other competent public institutions, some of their aspects (based on the decision of public bodies) are made accessible for general public.
- Based on the national legislation there is a trend to motivate HEIs to disclose information related to indicators of quality in education.

Examples of good practice:

- Creation of national website where information on the range of unemployment rate of graduates is published

(administered by the Ministry of Education, Science and Sport of the SR),
- List of recommendations which information on HEIs should be disclosed to public or published on the web site of the Ministry of Education, Science and Sport of the SR.

Barriers:
- Incompatibility of information systems, which are administered at the national level by various public bodies,
- One-sided communication – from HEIs provision of correct and on time given information is required by the Ministry, but there is a lack of analytical information delivered by the Ministry to HEIs (e.g. on employability of graduates of particular study branches),
- High level of bureaucracy, duplicity of information required to be disclosed to the public bodies by HEIs and a lack of criteria to evaluate the usefulness/value of the delivered information.

Recommendations:
- To simplify the process of obtaining information from HEIs,
- To interconnect existing information systems to prevent and eliminate duplicity of information,
- To interconnect all public information systems administered by different public bodies concerning employability of graduates,
- To provide relevant information flow on the results of analysis from the information systems to individual HEIs.

Institutional policy on collection, analysis and use of information on HEIs
- HEIs have established relatively powerful IT background (IT portals, websites, IT information systems), which they use to collect information on the key areas (student population profile, student assessment, research results),
- HEIs regularly and systematically collect information on the assessment of student progress and success rates.

Barriers:
- High administration and financial requirements on the monitoring of graduates' employability,
- Lack of HEI systems enabling provision of information on graduates and their employment,
- Low motivation and engagement of students in the processes of study program quality evaluation,
- Effectiveness of teachers and their quality is assessed mainly on the basis of quantitative indicators (number of lessons, research/art outputs, and usually depends on the criteria of public financing and HEIs' accreditation.

Examples of good practice:
- Establishment of Alumni Clubs maintaining contacts with the graduates.

Recommendations for HEIs:
- To establish systematic procedures on provision of information on graduates and their employability,
- To provide special information on jobs and careers HEI's graduates by the relevant institutions,
- To improve internal communication among different levels of HEIs about the measurers improving quality of education which were identified based on the results of the students' questionnaires,
- To optimize the output formats of analyses to ensure their better applicability for the decision making processes.

Information disclosure
- Internal policy of HEIs is currently shifted from the disclosure of basic mandatory information to a disclosure of a broader scale of different information, caused by the growing competition of the HEIs.
- Disclosure of information on the key activities of the HEIs has become a common practice (e.g. information on curricula, study plans, expected results and graduates` profiles, obtained qualifications),
- Disclosure of information on graduates in terms of their employability is absent. Some information is disclosed by the Ministry of Education, Science and Sport but it is not on sufficient level.

Barriers:
- High administrative requirements related to the processes of the information disclosure,
- Lack of internal systems and methodologies for disclosure with links to direct responsibilities.

Recommendations:
- To establish a system of disclosure of exceptional student`s achievements for potential employers,
- To establish systems of information disclosure about graduates employability with the accent on the fields and areas in which the graduates applied for work at the labor market, including information on their income levels. This system should be set up without any increase of the other administrative load,
- To develop and implement new information systems which will be able to analyze the recorded data and process them according to the requirements of the final (target) recipients (university management, students, employers, etc.).

Consequently the above mentioned findings were used as the input data for the international comparative study [14] prepared as one of the main final results of the IBAR project.

IV. Dissemination of the Research Results

In European context the Bologna process and its agenda has included improvement of the higher education quality from the start. Consequently, the importance of higher education quality together with the interest in the Bologna process priorities has been accepted also in the Central Asia region. All countries of this region declared quality assurance processes and mechanisms as the national priority for Erasmus+, and Bologna

process type reforms were set as the priority at the regional level. These facts have been reflected in the project Enhancing capacities in implementation of institutional quality assurance systems and typology using Bologna process principles (IQAT, www.project-iqat.eu/).

The newly started IQAT project is co-funded by the Erasmus+ Programme of the European Commission and it can be perceived as a continuation of the IBAR project having as one of its aims dissemination of the achieved results [16].

The project is co-ordinated by the Czech University of Life Sciences in Prague (Czech Republic) in co-operation with the Centre for Higher Education Studies (Prague, Czech Republic). The other program countries project partners are University of Alicante (Spain), University of Latvia (Riga, Latvia) and Constantine the Philosopher University in Nitra (Slovakia).

Partner countries involved in the project are Kazakhstan and Uzbekistan represented by the following higher education institutions::

- S. Seifullin Kazakh Agro Technical University (Astana, Kazakhstan),
- International Educational Corporation (Almaty, Kazakhstan),
- Karaganda State Technical University (Karaganda, Kazakhstan),
- Tashkent Chemical-Technological Institute (Tashkent, Uzbekistan),
- Samarkand Agricultural Institute (Samarkand, Uzbekistan),
- Andijan State University (Andijan, Uzbekistan).

General goal of the project is enhancing capacities in implementation of Bologna Process reforms at institutional level in higher education institutions in the Central Asia region (particularly in Kazakhstan and Uzbekistan). This general goal of the project is detailed in several specific project aims, which are:

- to map the quality assurance issues at HEIs in the partner countries involved,
- to offer interactive trainings on quality assurance issues at HEIs for the representatives of the involved HEIs from the partner countries,
- to develop drafts of internal quality assurance systems at HEIs in the partner countries (exploiting U-map methodology and typology of HEIs),
- to carry out a pilot implementation of the new or innovated or modified drafts of the internal quality assurance systems at the HEIs in the partner countries,
- to development final versions of the internal quality assurance systems.

A very important innovative feature of the project is the suggested exploitation of the methodology of mapping HEIs (U-Map). This methodology should enable to develop the institutional profiles taking into account grouped institutional activities and show the diversification of higher education systems in practice. This mapping is closely connected with quality in higher education which should not be assessed equally in all HE institutions but HEIs´ diversification, expressed in their missions, should be seriously taken into consideration. The mapping has been in its pilot stage in Europe and it will certainly be very interesting to see if, and to what extent, the mapping can be successfully exploited outside Europe.

V. Conclusion

Under current conditions in the market of education services the investment into education has, similarly as in case of other investments, its economic side. Therefore to assure the quality of education processes and their management becomes a priority task for each higher education institution. That is why the HEIs have to reflect the on-going changes in their stakeholders´ needs and external environment.

There is no doubt that currently information systems form an integral part of the management and quality assurance processes at HEIs in each country. Moreover there is also a relatively good knowledge about the quality assurance systems at the national levels of the particular countries but there is still missing enough information and experience related to the internal quality assurance systems at the institutional level. That is why projects like the IBAR or IQAR project are so important in both national and international contexts ([18], [19], [20]).

References

[1] A. Burrows, and L. Harvey, “Defining quality in higher education: the stakeholder approach”. AETT Conference on Quality in Education, University of York, 1992.

[2] T. H. Davenport, “Information ecology”. Oxford: Oxford University Press, 1997.

[3] R.D. Galliers, “Strategic information systems planning: myths reality and guidelines for successful implementation”. European Journal of Information Systems, 1/1991, pp. 55-64.

[4] R. D. Galliers, “Towards a flexible information architecture: integrating business strategies, information systems strategies and business process redesign”. Journal of Information Systems, 3/1993, pp. 193-213.

[5] J. Guang, W. Nunez, and F. H. Welsh, “Institutional strategy and information support: the role of data warehousing in higher education”. Campus-Wide Information Systems, 19/2002, 2002, pp. 168-174.

[6] D. Crosier, and P. Ruffio, “Higher education in Europe 2009: Development in the Bologna process”. Brussels: EACEAC, Eurydice, 2009.

[7] H. Eggins, H. “Drivers and barriers to achieving quality in higher education”. Rotterdam: Sense Publishers, 2014.

[8] EHEA, “The Bologna declaration of 19 June 1999: Joint declaration of the European ministers of education”. Available at: www.ehea.info/Uploads/Declarations/BOLOGNA_DECLARATION1.pdf

[9] EHEA, “Realising the European higher education area: Communiqué of the conference of ministers responsible for higher education in Berlin on 19 September 2003. Available at: http://www.ehea.info/Uploads/Declarations/Berlin_Communique1.pdf

[10] EHEA, The European higher education area – Achieving the goals: Communiqué of the conference of European ministers responsible for higher education”, Bergen, 19 - 20 May 2005.

Available at: http://www.ehea.info/Uploads/Declarations/Bergen_Communique1.pdf.

[11] ENQA, "Standards and Guidelines for Quality Assurance in the European Higher Education Area". Helsinki: ENQA, 2006.

[12] Z. Gadušová, "European project management". Topical issues of modenization: the economy, education, management, politics and business, Aktual'nyje voprosy modernizaciji: ekonomika, obrazovanije, upravlenije, politika i biznes. Semey: Kazachskaja finansovo-ekonomitceskaja akademija, 2012.

[13] A. Hašková, et al., "Quality and information". National study WP 11 Slovakia, international project IBAR: IBAR results WP 11: SK – Slovakia. Praha: Centre for higher education studies, 2012. Available at: http://www.ibar-llp.eu/assets/ files /wp11/IBAR_WP11_SK-National_report.pdf.

[14] A. Hašková, Ľ. Lachká, and Ľ. Pilárik, "Information and Internal Quality Assurance in European Higher Education Institutions". Journal of the European Higher Education Area, 4/2013, pp. 2-20.

[15] IBAR. Available at: http://www.ibar-llp.eu/results.html.

[16] IQAT. Available at: www.project-iqat.eu/#!about-the-project/ee7jd.

[17] T. Loukkola, and T. Zhang, "Examining quality culture: Part 1 – Quality assurance processes in higher education institutions". Brussesls: EUA, 2010.

[18] Ch. Orsingher, "Assessing quality in European higher education institutions: Dissemination, methods and procedures". Heidelberg: Physica-Verlag, A Springer Company, 2006.

[19] M. Pisoňová, and A. Nagyová, "The auto-evaluation in the process of improving the quality of educational and non-educational institutions. Procedia - Social and Behavioral Sciences, 149/2014, pp. 724–732.

[20] P. Wiliams, "Implementation of the quidelines adopted by the European ministers responsible for higher education". Council of Europe higher education series, No. 9: The legitimacy of quality assurance in higher education, Strasbourg: Council of Europe, 2010, pp. 75-80.

Audiovisual Technologies for Foreign Laguages Teaching

Veronika Pisarenko
Foreign Languages Department
Southern Federal University
Taganrog, Russia
jetaime2014@yandex.ru

Shavadi Arsaliev
the Institute of Chechen and General Philology
The Chechen state university
Grozny, the Chechen Republic, Russia
shavadiarsaliev@gmail.com

***Abstract*—Audiovisual technologies are considered to be very effective in foreign languages teaching because of their great possibilities to use the whole system of the human perception. Recently, research activities in this area emphasize on computer based foreign languages teaching. In this paper, the basic concept of audiovisual technology for foreign languages teaching is introduced. According to the types of video information used in education process, the technologies are classified into three classes: using films; using video courses; using all other video information. The effectiveness and the quality of education are evaluated by means of two parameters: change of level of competence and students' satisfaction of training methods. Finally, a discussion pointing out the current and future developments in this area is presented.**

***Index Terms*—Audiovisual technologies, foreign languages teaching, higher education, methodology, pedagogics, pedagogical algorithms, techniques of teaching, technologies of education, video resources.**

I. Introduction

Audiovisual Technology represents a new approach to teaching methods and techniques and this is the current reality in education. Audiovisual Technology dramatically improves efficiency and availability of teaching resources [1].

There are many definitions and concepts of *technology* in literature. *Audiovisual Technology* is a term for the abstraction of the process of constructing a set of operations, methods and techniques based on video resources and resulting in competence formation [1]. This technology is based on use of human audio and visual channels of information perception. It permits to organize the perception in maximal value. This is very important for teaching foreign languages. Video resources allow to demonstrate the object of studies – a foreign language – in real conditions. Video permits to demonstrate not only a language but the whole environment where it exists including cultural, social, ethnic, and many other aspects [2, 3, 4].

II. Materials and Methods

To consider the problem of audiovisual technologies in training it is necessary to recognize that all video materials for foreign languages teaching are classified by us as: 1) video courses; 2) video films, and 3) other video information or information. At work with video courses which represent a single whole including the tasks and educational problems, created by authors of video courses, the teacher follows strictly the structure of educational process offered by authors of video courses. The process of work with video films also has the specificity.

2.1. Video Films in Foreign Languages Teaching

Use of video films promotes realization of the requirement of a communicative technique - to present the process of language mastering by comprehension of live culture speaking another language; an individualization of training and development of motivation of speech activity of students [5].

Another advantage of video films is their emotional influence on students. Therefore the special attention should be given to the process of formation of the personal relation to the materials used in training process. It is the main objective of the personality focused training paradigm. Successful achievement of this purpose is possible firstly at regular display of video films, and secondly, at their methodically organized demonstration. The use of video films helps to develop various kinds of student's activity, and also their attention and memory. At video material viewing there is an atmosphere of joint informative activity in audience that promotes the increasing in attention concentration. Use of various channels of the information perception influences positively durability of storing of a regional geographic and language material. Psychological features of video films influence on students promote an intensification of educational process and create favorable conditions for the communicative competence formation. The teacher defines, whether it is necessary to include a video film consistently in educational process, or to use separate episodes taking into account conformity of subjects of a video film to a basic word stock and the communicative situations included in the foreign language program for a concrete grade level. The information quantity and character should correspond to quantity and quality of the information which the student studying a foreign language is capable to acquire at a given time.

Efficiency of use of a video film in educational process depends on exact definition of its position in training system, on coordination of educational possibilities of a video film with training problems, on rationality of organization of work, and on functional features of a video film.

There are the following *functional features* of a video film used in educational process:

Informative and training function – the student is involved in the film's subject line and in the process of information transfer simultaneously; this information will be used in the course of training;

Illustrative and evident function – video film shows a subject line in the art form;

Organizational and operational – it is realized in concentration and the subsequent management of students' attention by means of a subject line and art features of its embodiment;

Educational – it is embodied in realization of certain ideas in the art form and statement of problems which the student will discuss further;

Integrating – in the course of viewing there is a process of integration of various kinds of perceptions, and also various aspects of language - phonetic, lexical etc.[6]

There are the following 4 stages in the audiovisual technology structure with use of video films [7]: *preliminary work* – preliminary removal of language and cultural difficulties, statement of an educational task; *perception* - development of skills of the information perception; control of understanding of the basic maintenance; development of language skills and abilities of oral speech.

Preliminary work. The students are informed about the name of a film and they are offered to come out with assumptions of film's maintenance, time and a scene of action, possible characters. Phrases for assumption expression are introduced and trained necessarily. Then the new lexical material which is necessary for understanding of the maintenance of a film is introduced and intended for active possession. Introduction of new lexicon precedes demonstration of each part of a film. The special attention is given to phraseological units. The regional geographic comment is given necessarily.

Perception of a video film (in parts). Before viewing of each part students can be offered questions which are necessary to be answered after viewing.

Control of understanding of the basic maintenance. In the beginning students answer the questions offered by the teacher before viewing. Then the following educational tasks are proposed to students: *Choose the right answer from the offered ones. Arrange the phrases according to the film plot. Break the film into logic parts and choose a heading for each of them of the offered ones. Associate the following statements (remarks) with the film heroes. Associate the following events with a place of action. Name participants of the following dialogues. Choose one correct variant of the phrase continuation etc.*

Development of oral speech skills and abilities. It is possible to stimulate communicative activity by means of various tasks. It is necessary to begin with usual discussion of dynamics of a plot, features of heroes' appearances, characters and actions. The following types of tasks can be used: *Describe the place and action time. Remember the names of the main characters of the film. Describe the appearance of the main characters of the film. Express your opinion on characters of heroes of the video film. Restore the chronology of the film events. Share your general impressions about the film. State the problems which were brought up by the authors of the film.*

Further we pass to the problem discussion. The teacher has to stimulate the students' oral statements not only about the film maintenance, but also about the problems and the ideas given by the film. The following types of questions and tasks can provide the discussion: *How will you characterize the action time? What do you know about the scene of action? What associations are caused by the given scene of action? What emotions are caused by heroes of the film? Why? Whose character from heroes of the film is closest to yours? If you were on the place of the film heroes what would you do in this situation? Predict the further succession of events. Which problem do you consider as the most important one? If you were the film director how would you illustrate a similar problem in your film?*

For the decision of each tasks students should know not only the general maintenance of a video film, but also remember the details, and also be able to estimate events, to give the characteristic to characters, using words and expressions from speech support of a video film. Practical experience shows that video films in a foreign language often represent difficulty for students of technical college because of great volume of a language and speech material. As the practice shows it is much more effective to use separate video information [8]. Technologies of their use will differ from ones considered above.

2.2. Pedagogical Algorithms of Audiovisual Technologies Realization

For realization of audiovisual technology of foreign languages training with use of video information it was suggested to use a *pedagogical algorithm*. Algorithmization represents the process of working out and realization of algorithms for the trainee. An algorithm in pedagogics represents a sequence of educational operations leading to the object in view. Algorithmization of educational process on a foreign language is of great importance, in particular, in case of tutorials use. It is necessary to notice that application of tutorials for foreign languages training represents a well developed problem in a teaching technique. Nevertheless, occurrence of new tutorials (video, computers) demands working out of a special technique of work with them, allowing to use them most effectively and to receive a good level of results [9, 10].

We consider any material which is demonstrated in educational process with use of video equipment as a *video material* (a video fragment, a video course, a video program etc.).

What matters here is the way in which video materials are used in the training process. Their demonstration has to be subordinated to the certain educational purposes and, hence, it has to be organized in a special way. The expediency of use of video in educational process is provided by the following facts: 1) availability of video information which can be

recorded from various sources; 2) presence of a certain experience of using video equipment and a video production; 3) possibility of more active creative activity of the teacher.

In realization of audiovisual technology based on video information it is necessary to distinguish the following stages [1-3, 6].

- **Previewing** - statement of the purpose of viewing, the formulation of tasks, removal of lexical and grammatical difficulties.
- **Video information presentation** - primary viewing; viewing repetition if necessary.
- **After-viewing** including:

  1) the control of understanding having the various purposes and the forms;

  2) formation, development and perfection of the foreign language communicative competence on the basis of lexical, grammatical and cultural video information maintenances;

  3) total control.
- **Actualization** of the foreign language communicative competence as a result of work with video information.

The **previewing stage** has some variants of realization:

- The formulation of the global purpose of work and a specific goal of viewing and statement of an educational problem – what exactly should be made by students during presentation of a video information and after it?
- The formulation of the global purpose of work and a specific goal of viewing and statement of an educational problem, removal of lexical and grammatical difficulties, linguistic and cultural comments to a video information, having for an object the removal of difficulties of social and cultural character.

Removal of difficulties assumes introduction of a new lexical and grammatical material, its training on the basis of exercises, and also the linguistic and cultural comment of the social and cultural phenomena which will appear in the course of viewing.

The video information presentation can be carried out one or many times. It depends on the viewing purpose. The variety of video information viewing purposes allows to offer the following classification:

- *Illustrative viewing* – for the purpose of reception of evident illustrations for any language material, for example, the grammatical phenomena, the pronunciation skills, social and cultural phenomena;
- *Fact-finding viewing* – for the purpose of general idea reception about a video information and its subjects; viewing with the general coverage of the maintenance and concentration on understanding of the most essential; the minor, insignificant facts and unimportant details can be left; synthesis prevails over the analysis; this assumes the ability to define a theme, to allocate the basic idea, to find the main facts from the text; as a result students get skills of forecasting, allocation of semantic marks and support in the text, learn to guess value of keywords and to do without the words which value is not known to them; fact-finding viewing is an indispensable component of studying and critical viewing and can precede them. Generation of oral and written statements based on video information can be result of viewing, for example, fact-finding viewing is carried out with concentration on the subsequent transfer of the video information basic maintenance in the form of a short oral or written statement - the resume, summaries etc;
- *Studying viewing* – with extraction of the full information, with concentration on full and exact understanding of all basic and minor facts, their mental processing and storing; at studying viewing the student assumes to reproduce or to use the received information subsequently in this or that form: to retell orally or in written form, to estimate, to make comments, to explain the information, in common to discuss the video information maintenance; this kind of viewing is considered to be the most difficult, and requires considerable efforts and time;
- *Critical viewing* – for the purpose of the critical analysis, assuming an estimation of the video information maintenance and its correlation with the personal point of view, knowledge, with own life experience; the estimation is impossible without full and exact understanding of a video information, thus, critical viewing is based on studying viewing;
- *Search viewing* – for the purpose of a fast finding certain information in a video material (facts, characteristics, digital indicators, instructions, dates, proper names etc.); students have to be informed that the video material contains the information needed; extraction of the semantic information occurs automatically; search viewing assumes ability presence to be guided in logic and semantic structure of a video information, to choose from it the necessary information on a certain problem, to choose and unite the information of several video information on individual questions; search viewing is used for work with professionally oriented video information.

Besides, the presentation of video material can be realized in the form of the *developed* or *fragmentary exposition*. The given terms have been offered by I.D. Salistra and, in our opinion, till now are the most successful. I.D. Salistra suggested to consider exhibiting in corresponding reality situations of a certain speech material as the primary goal of an exposition [13]. *Developed exposition* represents the complete video information having a subject basis; *fragmentary exposition* contains separate fragments, situations.

The *after-viewing stage* is the process of studying following a video information perception. First of all, this is the control having various purposes and forms. Control can be carried out in various forms: oral, written, face-to-face, individual, question-answer; in different volume - control of the plot understanding, control of perception and discussion of separate details of a plot. Control can be intermediate and total. The *after-viewing stage* is aimed at formation, development and perfection of a foreign language communicative competence on the basis of lexical, grammatical and cultural video information maintenances. This is a question of introduction, training and fastening

language, speech and cultural material on the basis of the decision of various educational problems.

Actualization of a foreign language communicative competence represents a result of work with video information. It represents actually a stage of realization of the global purposes of work with video information. Actualization stage continues during all further use of a foreign language by the student. Frameworks of this process seem hard to be defined exactly. Nevertheless, we consider this stage should be necessarily allocated.

If we consider all stages of work with video information mentioned above, a variety of the training purposes, viewing types, specificity of the subject "Foreign language", it is possible to offer a number of operations sequences. They can be presented as algorithms of work with video materials which we suggest to use in foreign languages training.

Illustrative viewing. The main objective is demonstration of language or cultural phenomena. For example, we can use teaching material containing regional geographic information, the basic sights of any city of the studied language country. The grammatical phenomenon studied by students is possible to be illustrated by a corresponding fragment from speech of one of heroes of video information in which this grammatical phenomenon appears. Nonverbal ways of the native speakers' dialogues, various gestures, and signs also can be illustrated by video information. For *illustrative viewing* there are some algorithms of training technology. They differ in pre-viewing work. Fig. (1) shows the algorithm for *illustrative viewing*.

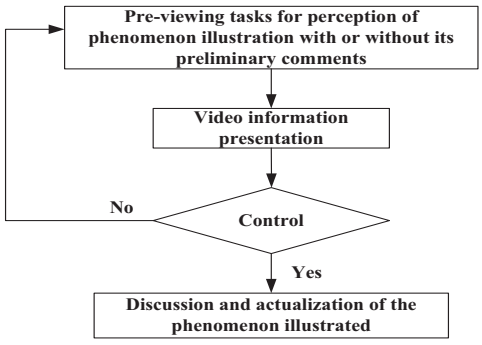


Fig. 1. Algorithm of foreign languages training process based on video information (*illustrative viewing)*

Let us consider the block diagram presented on Fig. 1. The algorithms consist of 2 stages uniting 4 blocks. The distinction between algorithm's variants consists in the following. In the first variant of algorithm students are offered to fix independently the certain phenomenon in the course of viewing, and then to pass to its discussion. For example, to fix, what sights have been shown during walk on a city. In the second variant of algorithm the description of the phenomena and objects which they will see is offered to students, comments are given, and then the video information is shown. For example, sights are listed, their characteristics are given, photos and drawings are shown, the background is told, and then it is offered to look at all it in video information. The first kind of algorithm actually is a version of search viewing, but in this case we pay attention to illustrative function of viewing.

Fact-finding viewing. When realizing a *fact-finding viewing* students' attention has to be attracted to video information as a whole without concentration on certain facts. The basic educational problem which faces students consists in finding out the basic information and problems as a result of viewing. The information processing is made consistently and involuntarily and results in construction of complex images.

Studying viewing. This kind of viewing provides full and exact understanding of all video information as much as possible, and its critical processing. There are some variants of training technology algorithms based on studying viewing.

Critical viewing. During viewing of this kind the video information actually plays a role of a situational basis for the organization of the further discussion. Tasks for critical perception are formulated, and then *critical viewing* is carried out. It is followed by the problems understanding control. In this way we have a situational basis for discussion which follows.

Critical viewing is a version of *illustrative viewing* as the illustration of a problem which will be discussed is actually carried out. We define it as a separate kind of viewing to underline the critical relation to a material which is subject for viewing. It is an indispensable condition of the further discussion organization. *Critical viewing* promotes formation and development of the critical thinking which is considered to be very important for a personality development.

Search viewing is focused on work with video information of cultural or special professional character. This kind of viewing is aimed at searching some special information.

The quality and efficiency of education and knowledge processing by students based on audiovisual technologies have been evaluated in experimental research.

III. RESULTS

Experimental research has been carried out in two groups of students: control and experimental. Both groups were studying foreign languages. The period of studies, the quantity of students, the starting level of foreign language communicative competence, the program contents were identical. The groups were working with the identical books. But experimental group was proposed to use video information in education process. Control group has never used video materials in studies.

For the communicative competence level estimation we use calculation of mastering factor of teaching material [20]. Mastering factor is considered to be a unit of measure for a communicative competence level. It is measured by the following

$$Q = \frac{m}{k}, \quad (1)$$

where Q corresponds to mastering factor; m means the number of the right answers; k is the number of possible answers.

Diagram on Fig.2 illustrates a comparison among results of the analysis of mastering factor for control and experimental groups for three years of training. The measurements have been based on complex tests composed by the author.

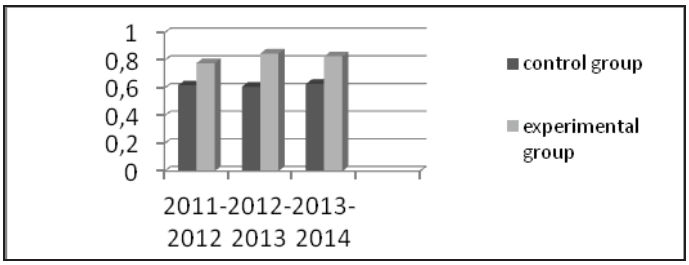


Fig. 2. Diagram of dynamics of mastering factor for control and experimental groups

Students' communicative competence level is considered to be an indirect indicator of students' satisfaction of the foreign language teaching technique based on video tools. To compare student's activity results in control and experimental groups in the subjective plan we have calculated the index of satisfaction by a foreign language teaching technique in these groups. The students' satisfaction by foreign language learning represents the subjective plan of activity. The indexes of students' satisfaction and importance factor have been defined by means of questionnaires developed by authors. Fig.3 shows the dynamics of students' index of satisfaction in the conditions of traditional and video based training.

The effectiveness and the quality of education are evaluated by means of two parameters: change of level of competence and students' satisfaction of training methods.

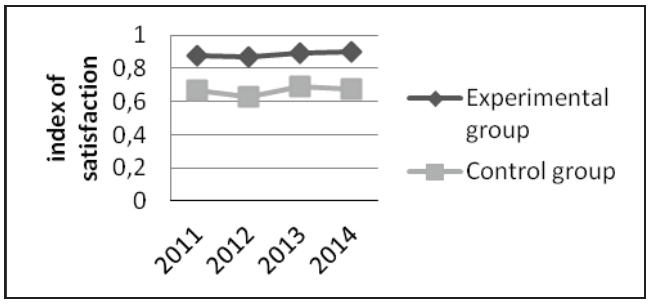


Fig. 3. Diagram of dependence of indicators of students' index of satisfaction in the conditions of traditional and video based training from a year of group formation

IV. Discussion

The analysis of the data of training results allows to make the following conclusions:

1. According to the final language tests results the mastering factor after 2 years of training testifies that the offered program, both in control, and in experimental groups is acquired
2. The communicative competence level of experimental groups is higher than in control groups. Average mastering factor is equal to 0,78 - 0,85 in experimental groups, to 0,61 - 0,63 in control groups. At the end of the first year of training mastering factor is higher in experimental groups, than in control on the average on 0,2; at the end of the II year of training - on 0,23.
3. Use of audiovisual technologies in foreign languages training leads to quantitative increase mastering factor that is expressed in qualitative increase of level of students' communicative competence in foreign languages.

Above descriptions show that the methods of exploiting video information in foreign languages training could vary greatly in many aspects. Moreover, some techniques can be easily integrated into other methods and forms of education process.

While comparing the index of students' satisfaction with a technique of foreign languages teaching in high school we should note the following:

- the students studying foreign languages in high school are satisfied by technologies of teaching; it has been proved by the fact that the average index of satisfaction is positive in general, see Fig.3;
- the general index of students' satisfaction with a technique of foreign languages teaching tends to increase during all the period of studies in experimental group; in control group we see an insignificant decrease of students' satisfaction;
- the index of students' satisfaction with a technique of foreign languages teaching in experimental group is higher than in control group, see Fig. 3; it means that the education technologies based on video are more interesting and more useful for students.

Variants of pedagogical algorithms of technology realization are applicable for work with any kind of video information taking into account possibility of entering of some changes connected with depth of perception of a material. From the point of view of sequence of actions they consider psychological features of video information perception, and also methodical features of the organization of foreign language educational process. The application of the developed algorithms will promote the most rational organization of educational process on the basis of the maximum use of all advantages of audio-visual synthesis and taking into account all stages of communicative competence formation. Training process will be even more effective if in realization of audiovisual technologies of training style characteristics of students about which it was told above are considered.

V. Conclusion

A central idea of this paper is to bring the common sense of humanistic thinking to the area of foreign language teaching based on video work in accordance with some simple principles:

- the learner is considered to be a center of education process as actor, doer, creator, and interpreter of video; all education process is organized for learner and is aimed at formation and development of foreign language communicative competence;
- the teacher offers the learner video resources of different

types in various forms and stages; they are selected by teacher and processed for being used in education process;

- the operational part of audiovisual technologies is realized by pedagogicas algorithms; their selection is defined by purposes of education process;
- since learners (and teachers) are very different from one another, they need to be provided with a wide range of activities and stimuli;
- all text is produced by the students - when viewing they produce their own very particular, internal version of what is on the screen; when videoing they make all the decisions both behind and in front of the lens;
- even course book material can have life breathed into it by letting the students take it over and re-fashion it in their own special way;
- experimental research has shown the effectiveness of audiovisual technologies use in education process.

VI. Current and Future Developments

This paper has presented a review of the state of foreign languages training audiovisual technologies and recent patents in this field. We briefly discussed the problem of using video in training process. Successively, we focused our attention on different types and forms of video information: video films, video information and self-made video. Moreover, this study allowed us to underline the main features of different types of video used in foreign languages training process. The experience of teaching with video meets expectations and clearly confirms the application and the use of video in education .

Our goal for the future is greater use and further development of audiovisual technologies in foreign languages training. Further development of the humanistic environment and its effective co-existence with existing traditional tutorials requires different approaches and collaboration.

Our past experiences and current trends in audiovisual technologies clearly determine and will determine our future directions and activities in this area.

The current situation in foreign languages teaching [10-13] should prompt us to conduct more researches on audiovisual technologies, developing new algorithms and tasks to increase the effectiveness of foreign languages teaching.

References

[1] Sh.M-Kh. Arsaliev, “Methodology of modern ethnopedagogics”. - M: GELIOS ARV, 2013. - 320 p. ISBN 978-5-85438-224-3.

[2] Sh.M-Kh. Arsaliev, V.I.Pisarenko, “Modeling of ethnopedagogical technologies in a context of a modern scientific picture of the world”. The Bulletin of the Chuvash state pedagogical university of I.J. Jakovlev. 1 (85). 2015. - pp.79-84.

[3] V. Pisarenko, “Audiovisual Technologies of Foreign Languages Teaching in Technical High School”. Moscow: Izdatelskiy dom MPA-Press, 2004.178 p.

[4] V. Pisarenko, “Video Resources Use in Modern Technologies of Foreign Languages Teaching” In IV Russian scientific and practical conference “Modernization of System of Professional Education Based on Regulated Evolutioning”, V 6. Ch.1. Yuzhno-Uralsk State University. – Chelyabinsk: «Obrazovanie», 2005, pp. 65-69.

[5] V. Pisarenko “Communicative Problems in Foreign languages Teaching Based on Audiovisual Technologies”, Izvestiya TRTU. Taganrog: Izd-vo TRTU, 2005. № 5 .pp. 121 – 127.

[6] O.I. Barmenkova, “Videolessons in System of Teaching Foreign Speech”. “Foreign Languages at School” 1999, № 3. 20-25 pp.

[7] V. Pisarenko, “Pedagogic Algoritm of Video Materials Use” In “Perspektivnyie informatsionnyie tehnologii i sistemyi”. Vol. 1(13). Taganrog. 2003. pp.77–83.

[8] I. Platonova, “Television in Foreign Languages Teaching”. Minsk: Vyisshaya shkola, 1980. – 88 p.

[9] E. Polat, “Telecommunications at School”. Informatika i obrazovanie. 1993. № 1. pp. 55–57.

[10] V. Pisarenko, “Problem of Effectiveness of Foreign Languages Lessons in Educational Space of Technical High School”. In International Conference “Intellektualnyie sistemyi” Moscow, Fizmatlit, 2005.pp.436-447.

[11] V. Pisarenko, “Problems of Informatization and Formalization in Foreign Languages Teaching in Technical High School”, Otkryitoe obrazovanie. 2004. № 4 (45).pp.20 – 29.

[12] V. Pisarenko, “Technical Tutorials Use in Foreign Languages Teaching”, Izvestiya vyisshih uchebnyih zavedeniy. Severo-Kavkazskiy region. 2006. № 2 (134).pp. 104-108.

[13] I. Salistra, “Methods of Foreign Languages Teaching”. Moscow: Vyisshaya shkola, 1966. 252 p.

Development of intelligent learning system based on the ontological approach

A.N. Afanasyev, N.N. Voit, D.S. Kanev
«Computer facilities» department
Ulyanovsk state technical university
Ulyanovsk, Russia
{a.afanasev, n.voit}@ulstu.ru, dima.kanev@gmail.com

***Abstract*—The paper has analyzed the current methods for developing automated adaptive learning systems. The authors have proposed the adaptive learning system based on the author's method of the synthesis of a learning path and the set of main models (models of subject domain, designer' profile, test) for training designers. The paper has presented the problems relating to the development of learning system architecture.**

***Index Terms*— Automated learning system, model of a subject domain, model of a learner, personalized learning process.**

I. Introduction

Today the promising direction in the computer-based learning is the usage of adaptive and multimedia technologies.

The main purpose of adaptive systems is to realize learning management based on user's personal characteristics. The adaptive methods allow making learning process shorter and more efficient through creating a comfortable learning environment for trainees. For this, the authors propose to change the pace of learning process, number of hours, the sequence of training materials, assignments, and content.

The mathematical support of adaptive system includes the model of a trainee's profile and model of a subject domain that are a basis for creating a personalized learning path. Using the tests for determining the trainee's current level of knowledge the learning system forms the trainee's profile. The learning system should provide both with theoretical and practical training on the computer-aided design (CAD) system and make an assessment of trainee's current level of competences and skills. It should teach trainees to carry out design activities.

Despite advances in adaptive learning system development the problems of learning systems' integration into the CAD system haven't been completely solved yet.

The study is concerned with an adaptive learning system based on the author's method of learning path synthesis.

II. The methods and tools for adaptive management of learning process in automated learning systems

The set of methods and tools for adaptive management of learning process [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] have been studied by authors. The common drawbacks with these methods and tools consist in:

- the lack of dynamic content of domain model doesn't allow the learning system to automatically include hybrid knowledge and enterprise's experience units into training process;
- effort in filling up with training and practical materials.

The analysis has shown that automated learning systems don't take into consideration the training specifics of design activity; there is no integration of the profile of designer's competences with his (her) design activity and with the enterprise design solution database.

III. Domain model of mechanical engineering design

The domain model is based on the ontology and consists of three levels: the scheme level, presentation level and practical level.

The scheme level describes the domain knowledge structure and relations among the knowledge units. At this stage the order of studying the elements is determined.

The presentation level is a set of training materials and reference books in hypertext form. The reference books are additional materials presented as hyperlinks for optional review. There are three types of supplementary training materials: reference books, glossary, and similar training materials.

The practical level is a level for designer's skill testing: the ability to perform design activities based on the acquired knowledge of automated design and design practice. This level is represented by a set of design solutions.

In subject domain there are the following classes:

- knowledge atoms;
- terms/concepts are the atoms grouped by any principle;
- training materials are the materials for studying. One of the types of training materials can be reference books.
- the purposes of training group atoms that should be studied;
- competences group the atoms that are included into this competence. Competences are a subclass of the purposes of training.

The domain model is given by:

$$O = (PSL, T, R, F, Ax). \quad (1)$$

where PSL is a set of design solutions, T is terms of application areas described by ontology.

A set of terms is given by:

$$T = \{C, In\}. \quad (2)$$

where C = {A, P, D, GOAL, COMP} is a set of ontology classes (A is knowledge atoms, P is terms/concepts, D is training materials, GOAL is the training objectives, COMP is competences), the "Training material" class has a slot of "Reference book" with a range of true or false values.

In is a set of objects of ontology classes.

R is a set of relations among ontology objects:

$$R = \{R_{learn}, R_{part}, R_{next}\}. \quad (3)$$

where R_{learn} – is a "be_studied_in" binary relation with «connected_to» semantics and connecting objects of ontology classes («Atom», «Concept») to the objects of «TrainingMaterial» class,

R_{part} – is a "be_a_part_of" binary relation with «part_of» semantics and connecting objects of ontology classes («Atom», «Concept») to the objects of «Concept», «Thetraining objectives» classes,

R_{next} – is "be_studied_after" binary relation with «after_of» semantics and connecting objects of ontology classes («Atom», «Concept») to the objects of «Concept» and «Atom» classes.

A set of interpretive functions is given as:

$$F = \{Fatom\_op, Fpsl\_a, Fedu, Fdefine, Fsimilar, T\}. \quad (4)$$

where Fatom_op: A → {Operation} – is a function of displaying an object of the «Atom» class on a set of operations of design solution,

Fpsl_a: PSL → {A} – is a function of displaying design solutions on a set of objects of «Atom» class,

Fedu: {A} → {D} – is a function of creating an ordered set of training materials for studying the certain knowledge atoms,

Fdefine: P → {D} –is a function of searching for training materials describing a certain concept,

Fsimilar: D → {D} –is a function of searching for similar training materials,

T: D → N – is didactic complexity of material.

A set of axioms is given by:

$$Ax = \{AxAHP, AxAHD, AxPAfP, AxPAfA, AxAAfP\}. \quad (5)$$

where AxAHP – «atoms are part of the terms/concepts», if the Y atom is a part_of the X term/concept, which is a part of the Z term/concept, the Y atom is a part of the Z term/ concept, in the form of Semantic Web Rule Language (SWRL): Term/Concept(?x) ∧ Atom(?y) ∧ Term/Concept(?z) ∧ be_a part of (?y, ?x) ∧ is_a part of (?y, ?z) → be_a_part of (?y, ?z).

AxAHL – «atoms are a part_of the training objectives», if the Y atom is a part_of the X term/concept, which is a part of the training objective of Z, the Y atom is a part of the training objective of Z, in the form of SWRL: Term/Concept(?x) ∧ Atom(?y) ∧ the_training_objective (?z) ∧ be_a_part of (?y, ?x) ∧ be_a_part of (?y, ?z) → be_a_ part of (?y, ?z).

AxAHD – «atoms are a part_of training materials», if the Y atom is a part_of the X term/concept, which is studied in training material of Z, the Y atom is studied in training materials of Z, in the form of SWRL: Term/Concept(?x) ∧ Atom(?y) ∧ Training Material(?z) ∧ be_a_part of (?y, ?x) ∧ be_ studied_in(?y, ?z) → be_ studied_in (?y, ?z).

AxPAfP – «atoms are studied after atoms», if the Y atom is a part of the X term/concept, which is studied after the Z term/concept, and the C atom is a part of Z, the Y atom is studied after C, in the form of SWRL: Atom(?y) ∧ Term/Concept(?x) ∧ Term/Concept(?z) ∧ Atom(?c) ∧ be_ a_part_of (?y, ?x) ∧ be_a_part_of (?c, ?z) ∧ be_ studied_after(?x, ?z) → be studied after(?y, ?c).

AxPAfA – «terms/concepts are studied after atoms», if the Y atom isa part of the X concept, which is studied after C atom, and the Y atom is studied after C, in the form of SWRL: Atom(?y) ∧ Term/Concept(?x) ∧ Atom(?c) ∧ is a part_of (?y, ?x) ∧ be_ studied_after(?x, ?c) →be_studied_after(?y, ?c).

AxAAfP – «atoms are studied after terms/concepts», if the Y atom is studied after the X term/concept and the C atom is a part of X, the Y atom is studied after C in the form of SWRL: Atom(?y) ∧ Term/Concept(?x) ∧ Atom(?c) ∧ is a part_of (?c, ?x) ∧ be_studied_after(?y, ?x)→ be_studied_after(?y, ?c).

IV. The Designer's Profile Model

To assess the designer's knowledge, the authors have chosen the overlay model which represents the knowledge as an overlay on the developed domain model. Each knowledge atom is corresponded to a level of trainee's knowledge of the certain atom and quantity of control measurement of this atom. Each knowledge atom is also corresponded to a level of skills and quantity of skill control measurement. The levels of knowledge and skills are changed after testing.

The designer's profile model is given by:

$$U = (UA, P, C, P\_AB, C\_AB, ACT\_G, ACT\_C). \quad (6)$$

where UA – is a subset of knowledge atoms that should be studied,

P = A → [0 .. 1] –is a knowledge level of the A atom,

C = A → N – is a number of control measurement of knowledge of the A atom,

P_AB = A → [0 .. 1] – is a skill level of the A atom,

C_AB = A → N – is the number of control measurement of the A atom skills,

ACT_G = A → N – is the total number of actions when using the reference skill of the A atom,

ACT_C = A → N – is the total number of actions in the current use of the A atom skill.

V. Design Task Model

To check the designer's skill level, the practical tasks on developing the design solutions are used. The system automatically compares the solution with the reference design solution and corrects the skills level and the number of control measurements of certain knowledge atoms.

The design task model is given by:

$$PTZ = (pq, std\_start, std\_finish). \quad (7)$$

where pq is a description of a task, std_start – is an initial design solution, std_finish – is a reference design solution.

The model of the design task solution is given by:

$$UPTZ = (U, PTZ, psl, Ap, An). \quad (8)$$

where U – is a designer's profile, PTZ – is a practical task, psl – is a design solution, Ap – is a subset of knowledge atoms for which the skills of designer are confirmed, An – is a subset of knowledge atoms for which the skills of designer are not confirmed.

After the analysis of the design solution the designer's profile is corrected. A set of knowledge atoms, for which the trainee's skills are confirmed, change the skill level and the number of control measurements by the formula:

$$C\_AB_{after} = C\_AB_{before} + 1,\ P_{ABafter} = (P_{ABbefore} \cdot C_{ABbefore} + 1)/(C_{ABbefore} + 1). \quad (9)$$

For which the student's skills are not confirmed:

$$C\_AB_{after} = C\_AB_{before} + 1,\ P_{ABafter} = (P_{ABbefore} \cdot C_{ABbefore} - 1)/(C_{ABbefore} + 1). \quad (10)$$

VI. THE SCENARIO MODEL

The scenario consists of stages containing a set of training materials, tests and design tasks. For each scenario a subset of knowledge atoms that the trainee should study is defined.

The scenario model is the following:

$$ST = (A, P, D, TZ, PTZ, PERIOD, Fd\_d, Fphase\_tz, Fphase\_ptz, Fphase\_d, Fphase\_first, Fnext\_phase). \quad (11)$$

where A – is a set of knowledge atoms, P – is a set of concepts, D – is a set of training material, TZ – is a set of tests, PTZ – is a set of practical tasks, PHASE – is a set of scenario stages, $Fd\_d = D \rightarrow D$ – is a function of training materials sequence, $Fphase\_d = PHASE \rightarrow D^*$ – is a function of belonging a training materials subset to the certain scenario stage, $Fphase\_tz = PHASE \rightarrow TZ^*$ – is a function of belonging a test subset to the certain scenario stage, $Fphase\_ptz = PHASE \rightarrow PTZ^*$ – is a function of belonging a practical tasks subset to the certain scenario stage, $Fnext\_phase = PHASE \rightarrow PHASE$ – is a sequence function of scenario stages, $Fphase\_first = PHASE \rightarrow D$ – is a function of the first training material at the certain scenario stage.

VII. LEARNING PROCESS MODELING (MODELING THE LEARNING PROCESS)

The algorithm used to generate a personalized scenario in detail with mathematical expressions is given in [12]. The detailed description of this algorithm consists of the following steps.

1. The choice of the purpose of education.
2. Download of the basic training scenario.
3. The designer consistently studies the training material in the framework of the stage, including additional information assets of an industrial enterprise.
4. The results of testing correct the level of designer's knowledge.
5. Testing the skills based on design tasks.
6. Assessment of the level of knowledge and skills. If the level is satisfactory, the designer moves on to the next stage of the scenario, on the second step. The learning process is concerned to be over if there is no next step. In case of unsatisfactory level of knowledge and skills a personalized learning scenario is formed on the seventh step.
7. The choice of the knowledge atoms requiring further study.
8. Generation of a personalized learning scenario. The selection of training materials.
9. The formation of a set of tests and design tasks.
10. Control testing.

VIII. THE ARCHITECTURE

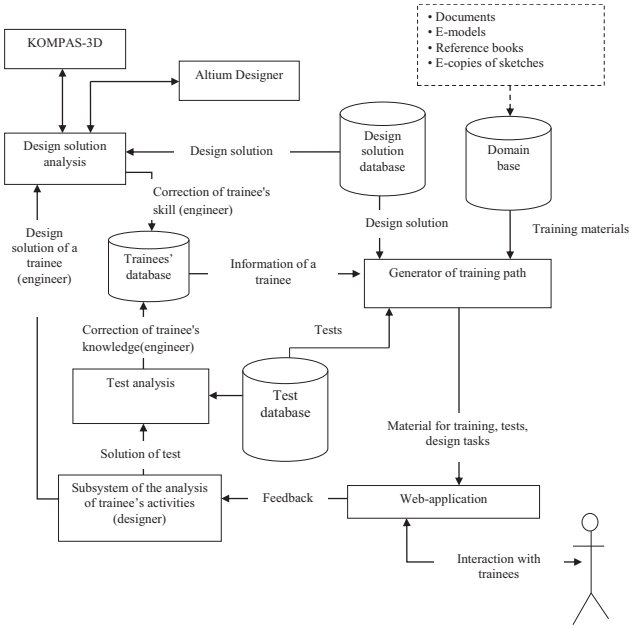


Fig. 1. Learning system architecture

The learning system architecture is shown in Fig. 1, a client-server architecture has been chosen as its basis. The customer implementation is a web application through which the user interacts with the learning environment. The learning system is based on the operating system Windows 7. The databases of design solutions, domains, trained engineers are stored on the MySQL 5.1 server. The generator of the learning path, test analysis, subsystem of the analysis of trained engineer's activities, web application is developed on PHP 5.3. The design solutions analysis is on C # 2.0. KOMPAS-3D. It is a CAD system of "Ascon". Altium Designer is a comprehensive computer-aided design system for radio-electronic tools developed by the Australian company called Altium.

IX. Conclusion

The automated training system using the new ontological domain model and a method of creating a personalized learning scenario have been developed. The ontological domain model of mechanical design is characterized by using the hierarchical and ordinal relations as well as by the integrating with design solutions. A method of creating a personalized learning scenario is characterized by the use of dynamic tools of interaction learning space models (trainee, test, test execution, practical task, practical task solution, scenario) with the ontological domain model and provides the formation of dynamic learning paths and reduction of training time.

Acknowledgement

The authors acknowledge that this paper was supported by the Russian Fund of Basic Research RFFI-15-07-08268a.

References

[1] O. A. Shabalina, “The learning process model and its interpretation in the training computer game,” Bulletin of SSTU, vol. 2 (70). Issue 1, pp.158–167, 2013.

[2] V. B. Kudryavtsev, P. A. Aliseichik, K. Vashik, “Learning process modeling,” Fundamental and Applied Mathematics, vol. 15, no. 05, pp. 111–169, 2009.

[3] V. Kasyanov, “WAPE - adaptive support system of distance learning of programming,” The International Conference "Computational and Informational Technologies in Science and Education", vol. 1, pp. 606–615, 2006.

[4] A. G. Dorrer, “Modeling of adaptive interactive training course,” Modern problems of science and education, vol. 5, pp. 52–59, 2007.

[5] M. J. Huang, H. S. Huang, M. Y. Chen, “Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach,” Expert Systems with Applications, vol. 33, pp. 551–564, 2007.

[6] M. K. Khribi, M. Jemni, O. Nasraoui, “Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval,” Educational Technology & Society, vol. 12 (4), pp. 30–42, 2009.

[7] V. Esichaikul, S. Lamnoi, C. Bechter, "Student Modelling in Adaptive E-Learning Systems," Knowledge Management & E-Learning: An International Journal, vol. 3, pp. 342–355, 2011.

[8] Yasir Eltigani Ali Mustafa, Sami Mohamed Sharif,” An approach to Adaptive E-learning Hypermedia System based on Learning Styles (AEHS-LS): Implementation and evaluation,” International Journal of Library and Information Science, vol. 3 no. 1, pp. 15–28, 2011.

[9] L. Mahnane, M. Tayeb Laskri, P. Trigano, “A Model of Adaptive e-learning Hypermedia System based on Thinking and Learning Styles”, IJMUE, vol. 8, no. 3, pp. 339–350, 2013.

[10] E. Ploum, Authoring of adaptation in the GRAPPLE project. Master Thesis. Eindhoven University of Technology, 2009.

[11] P. De Bra, D. Smits, K. van der Sluijs, A. I. Cristea, M. Hendrix, “GRAPPLE: Personalization and adaptation in learning management systems,” World Conf. edmedia, pp. 3029–3038, 2010.

[12] A. N. Afanasyev, N. N.Voit, D. S. Kanev, “Mathematical modeling of engineering education in the complex infocommunication of educational systems,” Radio engineering, vol. 7, pp. 133–136, 2014.

Loran-C Cycle Identification and ECD Estimation in Presence of white Gaussian Noise

Ali Kafaei Zad Tehrani
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran
alikafaee@yahoo.com

Ali Pourmohammad
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran
pourmohammad@aut.ac.ir

***Abstract* – A Loran-C system is a hyperbolic navigation system which works based on the TOA methods. The cycle identification and ECD estimation accuracies have considerable effects on the Loran-C system receivers localization accuracy so it is important to estimate the reference point and ECD as precisely as possible. White Gaussian noise degrades the performance of cycle identification and ECD estimation. Also, the error in cycle identification or ECD estimation degrades the performance of another one. In this paper two proposed methods for cycle identification and ECD estimation are evaluated in presence of white Gaussian noise and it will be described the relations between the cycle identification and ECD estimation. The simulations results show that the proposed methods are robust in presence of white Gaussian noise.**

Index Terms – Loran-C, Cycle Identification, ECD Estimation, White Gaussian Noise

I. Introduction

A Loran-C system is a hyperbolic navigation system based on the TOA methods. A Loran-C system consists of one Master station and at least two secondary stations which are organized into a chain. The master station transmits 9 pulses. The first eight pulses are 1 millisecond apart from together and the ninth pulse spaced 2 milliseconds after the eighth pulse. The secondary stations transmit a series of eight pulses, each spaced 1 millisecond apart from together too [1]. Once the last secondary has transmitted, the master transmits again, and the cycle is repeated. The period consists of two series of pulses for each station and it is called the group repetition interval (GRI) which is an important characteristic of the chain. The location of a receiver is computed based on the master and secondary stations time of arrival differences. To obtain the time of arrival, the reference point of each pulse must be detected. The detection accuracy has considerable effect on the estimation of the receiver's location. The process of obtaining of the reference point is called cycle identification and usually the 3rd zero crossing point is chosen to be the reference point [1]. Another important characteristic is the envelope to cycle difference (ECD) which indicates that the envelope shifts in time which it is illustrated in Fig.1. ECD can cause error in cycle identification on condition that it to be more than 2.5 µsec [2]. The cycle identification is a challenging task in presence of white Gaussian noise which usually occurred in applications. Most of the receivers use several GRIs to obtain a low noise signal. In this paper it is focused on obtaining the reference points and ECD without using several GRIs. The proposed methods can also be used for the times that a signal is obtained from several GRIs.

The reminder of this paper is organized in the following manner. In the section II some cycle identification methods are described and the proposed method is introduced. In the section III some ECD estimation methods are described and it is introduced the proposed method. Then, in the section IV the inference of ECD on cycle identification is analysed and in the section V the simulations results will be given for different SNRs, eventually in the section VI it will be concluded.

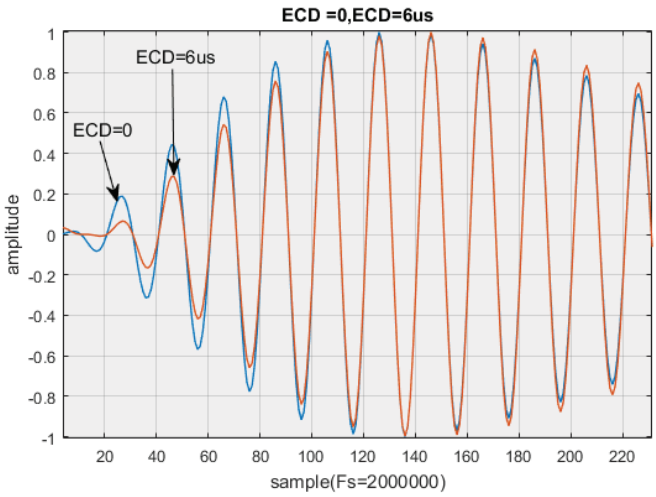


Fig 1. ECD = 0 (blue) and ECD = 6 µs (red)

II. Cycle Identification

Cycle identification is the process of detecting 3rd zero crossing point of a Loran-C signal. This process has considerable effect on the accuracy of a receiver localization. In [3] the delay and sum method (HCPR) is used which uses addition of the signal and delayed version of it to obtain a suitable signal to find the 3rd zero crossing point. This method is described in section A.

A. Delay and sum method

The delay and sum method is also known as the Half Cycle Peak Ratio (HCPR) method. If y(t) is the received signal and t is in μsec, the synthetic signal z(t) is created as:

$$z(t) = y(t) + \alpha \times y(t-5)\,. \quad (1)$$

where α is defined in as:

$$\alpha = \left(\frac{30+2.5}{30-2.5}\right) e^{\frac{-10}{65}} = 1.1975 \quad (2)$$

Fig. 2 shows *z(t)* and the zero crossing points which are marked with '+'. As it is illustrated in Fig. 2, there is a phase reversal at the 3rd zero crossing point which can easily be detected by sampling z(t-5) and z(t+5) of supposed zero crossing point until z(t-5) and z(t+5) have opposite signs [3]. For more information about the delay and sum method see [3] and [1]. To investigate other methods, see [5] please.

B. *The proposed method for cycle identification*

In this paper it is proposed a method for cycle identification which is robust in presence of white Gaussian noise. If y(t) is a LORAN-C single pulse, the match filter of the y(t) is defined as:

$$g(t) = y(T-t) \quad (3)$$

where T is the maximum value of t. The LORAN-C received signal is filtered with g(t) and Hilbert transform is used to obtain the envelope of the signal. To remove spurious peaks caused by the noise, a hamming filter is used and peak of the obtained signal shows the beginning of suitable signal. The 3rd zero crossing point is 30 μs after the beginning of this signal. The block diagram of the proposed method is shown in Fig 3 and Q(t) signal is shown in Fig 4.

III. ECD Estimation

A. *ECD Introduction*

ECD is the varied Loran-C signal's envelope. Eq. 4 illustrates Loran-C signal in presence of ECD as:

$$y(t) = A(t-T)^2 e^{-2\frac{(t-T)}{65}} \sin(0.2\pi t + PC) \quad (4)$$

where T is ECD value and PC is the value of the phase coding. In normal conditions, the value of ECD is zero. When the signal travels, the velocity of the propagation changes slightly with frequency and some components of the Loran-C signal travels faster than the others which results in ECD [3].

Many researches have been researched about ECD. In [4] Peterson and Dewalt used the collected data of some flights to estimate the relations between conductivity and ECD. They obtained a curve which shows that increasing in conductivity decreases ECD in nonlinear form. This Paper focuses on the obtaining value of ECD directly from the signal. This paper does not intend to model ECD, for ECD modeling see [4] and [2] please.

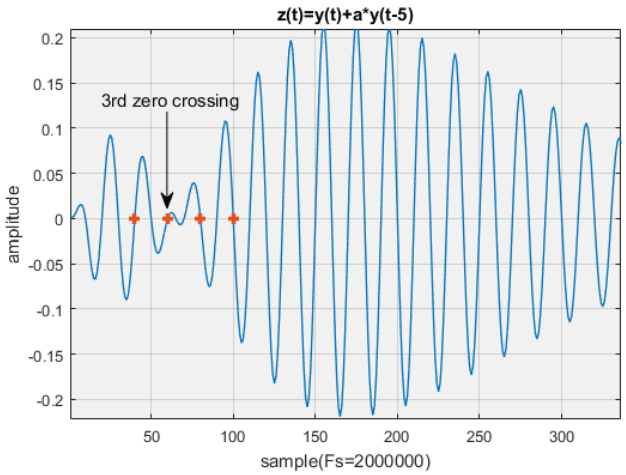


Fig 2. The Synthetic signal *z(t)* and four zero crossing point

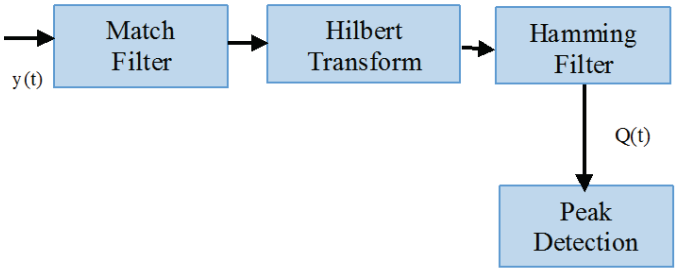


Fig 3. The block diagram of the proposed method

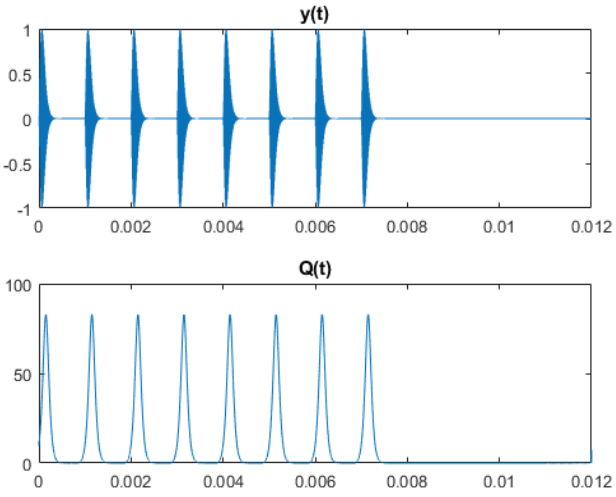


Fig 4. y(t) and Q(t) of Fig 3.

B. *ECD Estimation Based on LSE*

This method is described in [6]. The first envelope of the signal is obtained using Hilbert transform after applying the minimization algorithm like the steepest descend or the Newton-Raphson algorithm which are used to find the minimum of the cost function as:

$$J(T) = \sum_{i=1}^{N} \left(A(t_i - T)^2 e^{-2\frac{(t_i-T)}{65}} - y(s_i)\right)^2 \quad (5)$$

where $y(s_i)$ is a sample from the given signal and N is the number of samples which are used for obtaining the cost function. The cost function of Eq. 5 is a convex function which means that it has only one minimum.

C. The Proposed Method for ECD Estimation

The first signal is detected by the match filter, then the envelope is obtained by Hilbert transform. In the next step, two different points of the signal (from single envelope pulse) are sampled according to Eq. 6 and Eq. 7 as:

$$A(t_1 - T)^2 e^{-2\frac{(t_1-T)}{65}} = C_1 \quad (6)$$

$$A(t_2 - T)^2 e^{-2\frac{(t_2-T)}{65}} = C_2 \quad (7)$$

In Eq. 6 and Eq. 7, only *T* and *A* are unknown (the phase coding is known). If Eq. 6 is divided by Eq. 7, in the obtained equation only *T* is unknown and after few mathematical manipulations, it will be obtained Eq. 8 as:

$$T^2(H_2 - H_1) + T(-2t_1H_2 + 2t_2H_1) + t_1^2H_2 - t_2^2H_1 = 0 \quad (8)$$

$$H1 = \frac{C1}{C2} \quad (9)$$

$$H_2 = \exp(-\frac{2}{65}(t_1 - t_2)) \quad (10)$$

Solving Eq. 8 results two answers, the nearest result to zero is the value of ECD. This method has low computational complexity and does not need the use of optimization algorithms. The values of C1 and C2 should be enough high then it can be robust in presence of white Gaussian noise.

IV. Inference of ECD on Cycle Identification

The cycle identification proposed method uses the envelope to obtain the reference point location. When ECD is not zero, the location of the reference point is changed and the error will be occurred. Alongside with that ECD estimation needs exact timing of the signal, so that error in the reference detection leads to the error in ECD estimation. To solve this problem, it is need to find the exact location of the reference point. The location of the exact reference point is obtained as.

$$real\ reference = dref + ECD \quad (11)$$

In Eq. (11), *dref* is the detected reference point and the real reference point is the exact reference point. If ECD is between 2.5 μsec and -2.5 μsec, to reduce effect of ECD on the reference point detection, the nearest zero crossing point to the result of the proposed method for cycle identification will be the exact point of the reference point.

V. Simulations and The Results

To evaluate the performance of the proposed methods and compare them with the other methods in presence of white Gaussian noise, the results are simulated for different SNRs. Because of random nature of the noise, the results are evaluated 100 times and the average of the results are written in the paper. Table I show the values of the different parameters.

The results of the proposed method and delay and sum method for cycle identification are given in TABLE II. It is important to note that the error which is caused by the sampling frequency is not considered in the results given in TABLE II and TABLE III. TABLE II is the result of cycle identification on condition that ECD is zero. TABLE II shows that the proposed method works well especially for low SNR and the improvement is considerable. For SNR above 50 dB the both methods work properly. For evaluation of ECD, same procedure has been done and the results are shown in accordance with the known reference point from TABLE III. Although the amplitude of the sample is considered in computation of ECD, the proposed method is robust in presence of white Gaussian noise. The proposed method for estimation of ECD does not have computational complexity and it is suitable for implementation.

TABLE I. The Values Of The Parameters

number	parameters and values	
	parameter	*value*
1	Type of Loran-C signal	simulation
2	Sampling frequency	2000000
3	Filter type	FIR
4	Length of filter	57
5	Noise type	white Gaussian
6	Noise combination method	additive

TABLE II. Cycle Identification Evaluation[1]

column	Cycle identification evaluation		
	SNR(dB)	*Delay and sum(μs)*	*Proposed(μs)*
1	30	1.2	0.83
2	40	0.8	0
3	50	0.35	0
4	60	0	0

TABLE III. ECD Evaluation

column	ECD estimation error		
	SNR(dB)	*LSE(μs)*	*Proposed(μs)*
1	30	0.3	0.058
2	40	0.1	0.017
3	50	0.03	0.004
4	60	0.0057	0.0067

[1] Error is evaluated without consideration of sampling frequency error

VI. Conclusion

A Loran-C system is a hyperbolic navigation system which works based on the TOA methods. The cycle identification and ECD estimation accuracies have considerable effects on the Loran-C system receivers localization accuracy so it is important to estimate the reference point and ECD as precisely as possible. White Gaussian noise degrades the performance of cycle identification and ECD estimation. Also, the error in cycle identification or ECD estimation degrades the performance of another one. In this paper two proposed methods for cycle identification and ECD estimation evaluated in presence of white Gaussian noise and it described the relations between the cycle identification and ECD estimation. The simulations results showed that the proposed methods are robust in presence of white Gaussian noise. Also, it described that the occurred error in ECD or cycle identification can degrades another one's accuracy. The result of TABLE II and TABLE III showed that the proposed methods outperform to the compared methods. Another point is that the proposed methods are stable in presence of white Gaussian noise which makes them suitable to use in low SNR conditions. In this paper the sky waves and inferences issues were not taken into account. The proposed methods are not robust in presence of the sky waves. In future, it will be analyzed the methods which can be robust in presence of the sky waves and inferences.

References

[1] A.J. Fisher, "Loran-C cycle identification in hard-limiting receivers," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 290-297, Jan 2000

[2] P. Williams, D. last, "Modelling Loran-C Envelope-to-Cycle Differences in Mountainous Terrain," *32nd Annual Meeting, International Loran Association*, 2003

[3] L. Tetley and D. Calcut, "Electronic aids to navigation," Edward Arnold, London, 1986

[4] B.B. Peterson, K.M. Dewalt, "Analysis of Envelope-to-cycle Difference(ECD) in far field," *Proc. 21st Ann. Tech Symp*., pp 89-95, Wild Goose Association, 1993

[5] S. Li and F. Huang, "Algorithm of Cycle-Identification for Loran-C Signal Based on DLL," *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on*, Shanghai, China, 2012, pp. 1-3.

[6] N. Kouguchi, M. Sato and N. Morinaga, "Measurement technique for Loran-C pulse wave distortion measures and performance in an environment of noise," *Position Location and Navigation Symposium, 1992. Record. 500 Years After Columbus - Navigation Challenges of Tomorrow. IEEE PLANS '92., IEEE*, Monterey, CA, 1992, pp. 82-87.

Simulation of Virtual Environment for the Development of Teaching Laboratories

Kseniia Kuzmina
Department of Natural Sciences
ITMO University
St. Petersburg, Russia
kakuzmina@corp.ifmo.ru

Anzhela Andreeva
Department of Natural Sciences
ITMO University
St. Petersburg, Russia
avandreeva-fel@yandex.ru

***Abstract*— The article presents first step of the modeling of the virtual environment and the description of hardware and software tools for its development. This environment simulates a laboratory for the study of the characteristics of multimode optical fibers, which in the future will provide interaction with the biophysical objects. This three-dimensional model will allow the students of different technical specialities to study the automated systems laboratory course productively.**

***Index Terms*— virtual research laboratory, virtual environments, project-based learning**

I. Introduction

Improving the quality of highly qualified specialists training is an urgent problem because of the constant increase in the number of high-tech equipment in various sectors of manufacture. Therefore, it is necessary to improve methods of teaching students in higher educational institutions of technical and natural scientific profile. Thus, the use of virtual reality technology opens new possibilities in the organization of students learning [1].

The existing development of the virtual laboratories have a number of significant drawbacks. Many of them model only simple physical processes, and do not provide the "immersion" of the student in the 3D interactive environment of the phenomenon under study, with the possibility of natural interaction with objects (artificial and/or real) in a virtual environment. The second problem is the narrow focus of laboratory-based work. These problems can be solved by using the proposed virtual reality technology that allow creating a virtual environment simulating a laboratory and providing interaction with the biophysical objects. The software used for the development of this virtual laboratory are described below.

II. Software And Hardware

Currently, the area of optoelectronics as fiber-optic sensors is being actively developed [2]. In this work, the task was to create the 3d model scientific laboratory for the study of the characteristics of multimode optical fibers used in the construction of fiber-optic pressure sensors. Subtasks were:

- Design of interior of the laboratory for a specific theme (from a framework to a full room in the virtual environment).
- Setting up materials (virtual materials are set according to properties of real materials – transparency, reflection, light diffusion, roughness, topography, etc.).
- Adjusting the direction, power, lighting spectrum.

Primarily, the development of the virtual laboratory includes the creation of the room in which will be virtual hardware. As well as the actual equipment. At the initial step, three-dimensional mathematical model of the scene and the objects in it was established.

For modeling the software package Autodesk 3ds Max was used. This is a full-featured professional software system for creating and editing three-dimensional graphics and animation [3]. It contains the latest tools for artists and professionals in the field of multimedia. 3ds Max has extensive tools to create various shapes and complexity of three-dimensional computer models of real or fantastic objects of the environment, using a variety of techniques and mechanisms, including the following:

- Polygonal modeling that includes Editable mesh (editable area) and Editable poly (editable polygon) is the most common method of modeling is used to create complex models and low polygonal models. Generally, modelling complex objects converted to Editable poly starts with building a parametric object "Box", and therefore, the conventional method of modeling is called "Box modeling".
- Simulation based on heterogeneous rational B-splines (NURBS) (it should be noted that NURBS modeling in 3ds Max is so primitive that nobody uses this method).
- Modelling based on the so-called "mesh chunks" or Bezier surfaces (Editable patch) — suitable for modeling solids of revolution.
- Modeling using built-in libraries of standard parametric objects (primitives) and modifiers.
- Modelling based on splines (Spline) with subsequent application of the Surface modifier — a primitive analogue of the NURBS which is convenient for creating objects with complex flowing shapes that are difficult to create by methods of polygon modeling.

- modelling based on splines with application modifiers Extrude, Lathe, Bevel Profile, or create based on spline objects called Loft. This method is widely used for architectural modeling.

The result was a simulated laboratory room identical to the real lab. The laboratory is used to perform the experiment for the study of optical fiber characteristics. The result is shown in Fig. 1.

Fig. 1. A three-dimensional model of the laboratory

It is important to note that the laboratory has a particular theme, so each object in the three-dimensional scene was modeled separately. For example, an optical table, which is very rigid and flat base, which accommodates equipment for experiments, precise measurements and precise assembly of components of complex technical devices and assemblies (Fig. 2).

In the model the table structure and materials were reflected with high precision.

In the near future the lighting will be designed and realistic renderings of the laboratory will be created. And the camera will be set up (it revisualizing object that displays the scene view. The location and direction of view is determined by the position and orientation of the camera) for a more realistic overview of the laboratory for the observer.

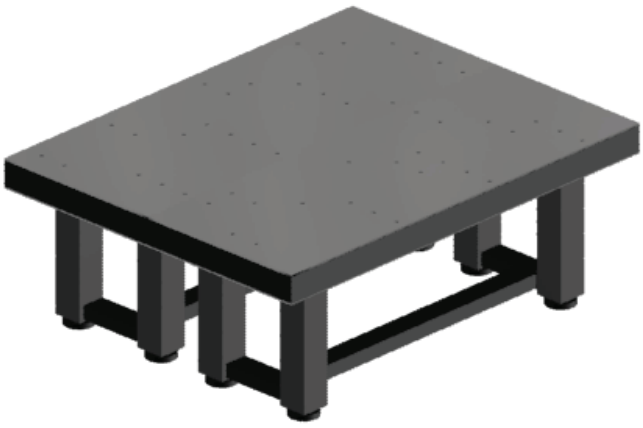


Fig. 2. Rendering of three-dimensional model of the optical table.

III. Conclusion

The organization of any research activities associated with the receipt and analysis of the results of the studies. This is a complex and laborious process that requires a lot of manpower and the unique expensive equipment. The problem of using this unique equipment can make the research difficult or even impossible at all. The use of a virtual environment that simulates the algorithm of the laboratory work is identical to the execution on real equipment for training of students of higher educational institutions will allow to solve this problem. This solution can improve the quality of training.

As a result of work tasks in hand was performed. The ultimate goal is the creation of a full complex of research laboratories in various fields of science and technology, working in virtual reality. The effect of immersion will be provided with the use of manipulators as virtual reality headsets in conjunction with the motion controller. The proposed environment provides the most realistic implementation of the experiment, since the use of additional tools provides an immersive experience.

References

[1] A. V. Trukhin, “On the use of virtual laboratories in education,” Open and distance education, vol. 4, 2002.

[2] A. A. Vetrov, S. S. Komissarov and A. N. Sergushichev, “Fiber-optic end-face interferometer - a universal element of building displacement sensors,” Optical Journal, vol. 1, pp. 3-6, 2008

[3] A. G. Gorelik Tutorial 3ds Max 2012. SPb, BHV-Peterburg, 2012, 544 p.

The Impact of Pulse Shaping Filters on OFDM-Based Visible Light Communications

Omer Narmanlioglu
Dept. of Electrical Electronics Eng.
Ozyegin University
Istanbul, Turkey
E-mail: omer.narmanlioglu@ozyegin.edu.tr

Refik Caglar Kizilirmak
Dept. of Electrical Electronics Eng.
Nazarbayev University
Astana, Kazakhstan
E-mail: refik.kizilirmak@nu.edu.kz

Murat Uysal
Dept. of Electrical Electronics Eng.
Ozyegin University
Istanbul, Turkey
E-mail: murat.uysal@ozyegin.edu.tr

***Abstract*—This paper discusses the impact of pulse shaping filter on the performance of orthogonal frequency division multiplexing (OFDM) based visible light communication (VLC) systems. Particularly, direct current biased optical OFDM (DCO-OFDM) with square-root-raised-cosine (SRRC) pulse shaping filters is investigated. The bit-error-rate (BER) performance of the system is evaluated for different excess bandwidth parameters (roll-off factors) and compared with the performance with rectangular pulse shaping filter through Monte Carlo simulations.**

***Index Terms*—Visible light communication, pulse shaping filter**

I. Introduction

Visible light communication (VLC) has recently attracted considerable attention due to its ability to convert every light-emitting-diode (LED) based luminary into a wireless data transmitter [1]. It is also considered as complementary technology [2] to radio frequency (RF) based wireless systems due to not only providing low-cost and energy efficient solutions but also its spectrum ($430 - 770$ THz) is not regulated and does not interfere with RF signals. VLC is expected to find place in many applications such as for indoor short-range communications, vehicular networks, aircraft cabins, hospitals, underground mines and underwater communications.

Intensity modulation and direct detection (IM/DD) schemes are used for VLC due to the non-coherent characteristics of LEDs. In IM/DD, the information is encoded by modulating the intensity of light and a simple photodetector (PD) at the receiver reconstructs the emitted information waveform. The main constraints on IM/DD schemes are that the transmitted signal must be real and positive valued. Therefore, VLC was initially started with simple pulse modulation techniques such as on-off keying (OOK) and pulse position modulation (PPM) [3] where the data rate was mainly limited by the bandwidth of the LEDs. Recent studies [4]–[8] have adopted orthogonal frequency division multiplexing (OFDM) which provides reliable transmission and achieves high data rate over wideband channels. Due to the constraints of IM/DD transmission, conventional OFDM structure, that has been in use for RF systems, cannot be directly applied to VLC. In order to overcome these constraints, different methods such as direct current biased optical-OFDM (DCO-OFDM) [4], asymmetrically clipped O-OFDM (ACO-OFDM) [5], flip (or unipolar) OFDM [6], [7] and enhanced unipolar OFDM [8] have been proposed. These methods differ in their spectral and power efficiencies. In this work, we consider DCO-OFDM due to its simplicity and considerably higher spectral efficiency.

Optical OFDM suffers from high peak to average power ratio (PAPR). The amplitude levels that exceed the dynamic range of the LEDs are usually clipped and the transmission is distorted. In the literature, several works investigate the performances of different OFDM based systems with clipping distortion [9]–[12]. None of the works, however, considers the effect of pulse shaping filter. In this work, we investigate the impact of pulse shaping filter in the context of VLC, particularly when the OFDM waveform is clipped. We consider rectangular pulse shaping filter as a benchmark and compare its bit-error-rate (BER) performance with square root raised cosine (SRRC) pulse shaping filters considering different excess bandwidth parameters (roll-off factors). We perform our analysis with a realistic indoor channel model that has been adopted as a IEEE $802.15.7$r1 VLC reference channel model (Scenario III) [13].

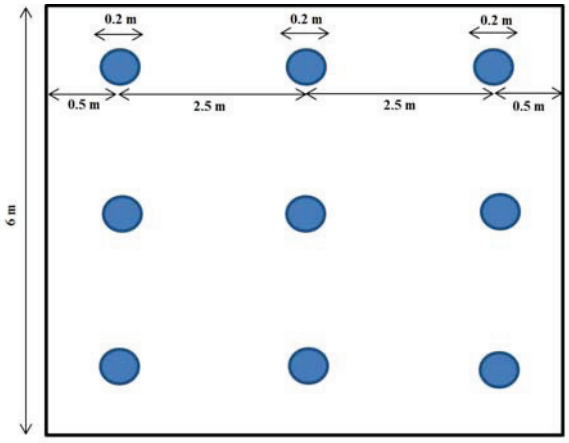
The remainder of this work is organized as follows. In Section II, we briefly summarize the IEEE $802.15.7$r1 channel model. In Section III, we present the transmission model of DCO-OFDM based VLC. The BER performance results for different pulse shaping filters are presented in Section IV. Finally, we conclude the paper in Section V.

Notation: $||.||^2$, $(.)^*$ and $[.]^{\mathrm{T}}$ denote Euclidean distance, complex conjugate and transpose, respectively. $\otimes$ is convolution operator and $F\{.\}$ denotes Fourier transform.

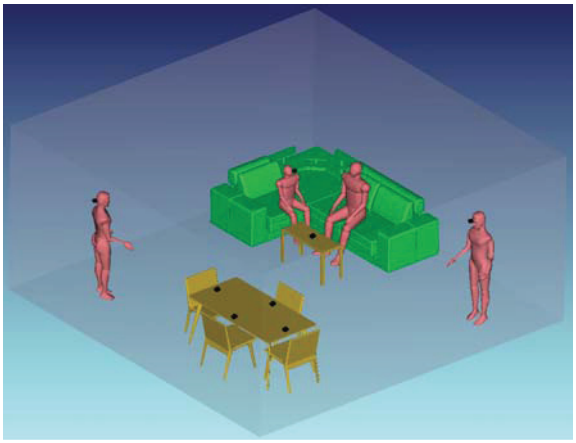
II. Channel Model

We use the reference channel model (see Fig. 1) adopted by IEEE $802.15.7$r1 Task Group "Short Range Optical Wireless Communication" [13] where a living room environment with the dimensions of 6 m x 6 m x 3 m is considered. The environment includes table, chairs, couch, coffee table and human bodies. The locations of 9 LEDs, which work at the same power level and transmit the same information, are shown in Fig. 1 (a). In the model, 8 test points (see Fig. 1 (b)) are chosen which are categorized into four groups. The groups are: (a) PD with the location of D1 on the coffee table at a

height of 0.6 m with with $45°$ rotation, (b) PDs with the locations of D2 and D3 that are next to the wall at a height of 1.7 m with $45°$ rotation, (c) PDs with the locations of D4-D7 on the table at a height of 0.9 m and (d) PD with the location of D8 on the top of couch at height of 1.1 m with $45°$ rotation. All PDs are presented by black color in Fig. 1 (b) and the coordinates are D1: $(0.6, -1, 0.6)$, D2: $(-2, -2, 1.7)$, D3: $(2, 2, 1.7)$, D4: $(-1.8, 1.5, 0.9)$, D5: $(-0.2, 1.5, 0.9)$, D6: $(-1, 1.85, 0.9)$, D7: $(-1, 1.15, 0.9)$ and D8: $(1.7, -1.8, 1.1)$. All the related parameters of the model are given in Table I.



(a)



(b)

Figure 1: Indoor room environment [13].

We represent the optical channel impulse response (CIR) for each PD by $c_i(t)$ where $i \in \{1, ..., 8\}$. The CIR for each PD is obtained by ray tracing simulations using Zemax®. In this work, we consider D1 as our destination terminal and Fig. 2 presents the obtained CIR with 1 nsec resolution. It can be noticed that the root mean square (RMS) value of delay spread is 9.64 nsec and the channel DC gain is 1.82 x 10^{-4}.

III. System Model

The block diagram of DCO-OFDM transmission is depicted in Fig. 3.

In DCO-OFDM, binary information is first mapped to complex symbols $s_1\ s_2\ ...\ s_{N/2-1}$ based on the deployed

TABLE I: Living Room Parameters.

Materials	Walls: Plaster, Ceiling: Plaster Floor: Pinewood
Objects	Table with 4 chairs, Couch Coffee Table, 4 human bodies
Object Specifications	Tables: Wooden Chairs: Wooden matched with table Couch: Cotton, Coffee table: Glass Human body: Shoes: Black gloss paint Head & Hands: Absorbing Clothes: Cotton
LED Specifications	Brand: CR6 − 800L Cree Inc. Half viewing angle: $40°$
Receiver area	1 cm^2

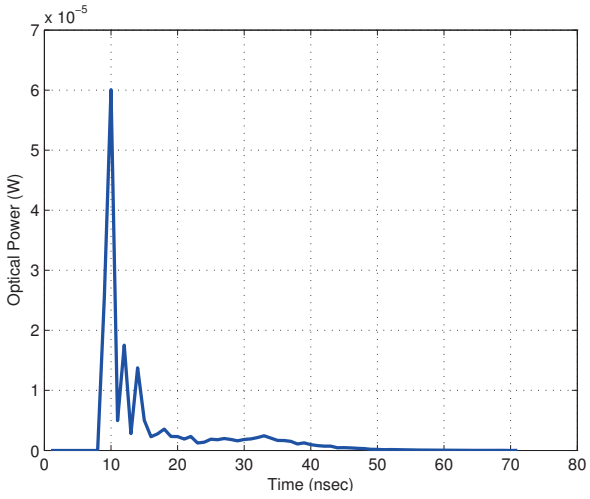


Figure 2: Optical channel impulse response (CIR) for D1.

constellation scheme such as $M-$order phase shift keying (PSK) or quadrature amplitude modulation (QAM) where N is the number of subcarriers. To ensure that the output of inverse discrete Fourier transform (IDFT) is real-valued, Hermitian symmetry is imposed on the transmitted signal which yields

$$\mathbf{X_S} = \left[0\ s_1\ s_2\ s_3\ ...\ s_{N/2-1}\ 0\ s^*_{N/2-1}\ ...\ s^*_3\ s^*_2\ s^*_1\right]^{\mathrm{T}}. \quad (1)$$

The output of $N-$ IDFT can be written as

$$x_S[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_S[k] e^{j\frac{2\pi nk}{N}}. \quad (2)$$

A cyclic prefix whose length, denoted by N_{CP}, is greater than or equal to the delay spread of the channel is appended to beginning of each OFDM frame and the parallel streams are converted to serial. Before the digital-to-analog conversion with the rate of T_{DAC}, the discrete signals are upsampled with a factor of T_U and passed through transmit pulse shaping filter which is denoted by $g_T[n]$. A DC bias (V_{BIAS}) is added to shift the band-limited signal into operation range of LED which is limited by turn-on-voltage (V_{TOV}) and maximum-allowed voltage (V_{MAX}). Fig. 4 illustrates the DCO-OFDM waveform within the dynamic range of the LED.

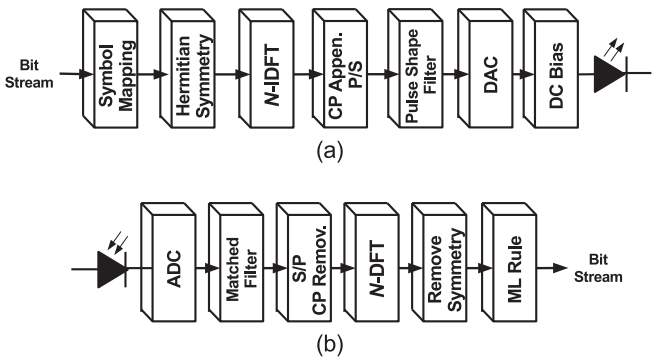


Figure 3: Block diagram of (a) transmitter and (b) receiver sides of DCO-OFDM transmission.

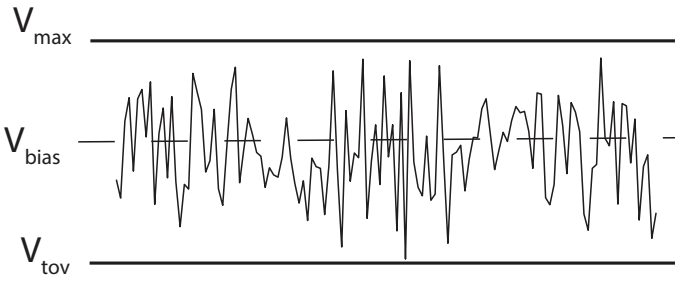


Figure 4: DCO-OFDM waveform.

The amplitude levels of time domain continuous waveform $(x_S(t))$ beyond the dynamic range will be clipped such as

$$\bar{x}_S(t) = \begin{cases} V_{MAX} & \text{if } x_S(t) \geq V_{MAX} \\ x_S(t) & \text{if } x_S(t) \in (V_{TOV}, V_{MAX}) \\ V_{TOV} & \text{if } x_S(t) \leq V_{TOV} \end{cases} \quad (3)$$

The time domain received signal by PD can be written as

$$y(t) = R\bar{x}_S(t) \otimes c(t) + v(t) \quad (4)$$

where R is the responsivity of PD (A/W) and $v(t)$ is additive white Gaussian noise (AWGN) with zero mean and σ_N^2 variance calculated by $N_0 W$ where N_0 is noise power spectral density and W denotes the system bandwidth. We can also write (4) as

$$y(t) = R(x_S(t) + d(t)) \otimes c(t) + v(t) \quad (5)$$

where $d(t)$ denotes the random distortion noise term with zero mean [14]. Its distribution and power depend on the selected pulse shaping filter and transmitted signal power.

After analog-to-digital conversion with the rate of T_{ADC}, the received sequence is passed through matched filter $(g_R[n])$ and sampled with T_U. Note that the symbol rate or sampling interval (T_S) is equal to T_{ADC}/T_U under the assumption of $T_{DAC} = T_{ADC}$. The signal is converted to parallel streams and cyclic prefix is removed. The signal on the k^{th} subcarrier at the output of $N-$ DFT can be written as

$$Y[k] = R(X_S[k]G_T[k] + D[k])C[k]G_R[k] + V[k]G_R[k] \quad (6)$$

where $Y[k] = F\{y(t)\}$, $X_S[k] = F\{x_S(t)\}$, $D[k] = F\{d(t)\}$, $G_T[k] = F\{g_T(t)\}$, $C[k] = F\{c(t)\}$, $G_R[k] = F\{g_R(t)\}$ and $V[k] = F\{v(t)\}$. Let $H[k]$ denote the frequency domain bandlimited electrical channel impulse which can be written as $H[k] = G_T[k]C[k]G_R[k]$. The output of *Maximum Likelihood* (ML) decision rule is then written as

$$\hat{X}[k] = \arg\min_{x \in \Phi} \left[\|Y[k] - RH[k]x\|^2 \right] \quad (7)$$

where Φ denotes the symbols in deployed constellation scheme with the average symbol energy of E_X. Then, the subcarrier-based SNR can be written as

$$\text{SNR}[k] = \frac{R^2 E_X \left|H[k]\right|^2}{\sigma_N^2 + R^2 \sigma_D^2 \left|C[k]G_R[k]\right|^2} \quad (8)$$

where σ_D^2 denotes clipping distortion term and depends on pulse shaping filter selection.

IV. Numerical Results

In this section, we present the effect of pulse shaping filter selection on the BER performance of the VLC system. All simulation parameters are defined in Table II. It is assumed that perfect channel state information is available at the receiver side.

TABLE II: Simulation Parameters.

Optical-to-electrical conversion loss (R)	1.0 A/W
Turn-on voltage (V_{TOV})	2.75 V
Maximum-allowed voltage (V_{MAX})	4.00 V
Bias voltage (V_{BIAS})	3.375 V
Upsample factor (T_U)	50
Number of subcarrier (N)	1024
Length of cyclic prefix (N_{CP})	24
Modulation	$16-$QAM
Sampling interval (T_S)	50 nsec
Noise power (σ_N^2)	-57 dBm

We consider square-root-raised-cosine (SRRC) pulse shaping filter with different roll-off factors (α). The time domain responses of considered pulse shaping filters are presented in Fig. 5 and their frequency responses are available in Fig. 6. It can be seen that SRRC filter with $\alpha = 0.1$ has the lowest peak value, however, it has high side lobes. In contrast, SRRC with unit α has the highest value and relatively, its side lobes are lower. In terms of bandwidth efficiency, SRRC with $\alpha = 0.1$ requires the lowest bandwidth. On the other hand, rectangular filter has an infinite bandwidth which leads inefficient frequency spectrum usage.

In Fig. 7, we present the BER performance of the DCO-OFDM based VLC system with the channel model presented in Section II. When the signal power level is low (up to 30 dBm average OFDM signal power), there is no clipping distortion and the BERs for different pulse shaping filters are almost the same. When the signal energy is further increased ($>$ 30 dBm), the SRRC pulse shaping filter with $\alpha = 0.1$ outperforms the others since it has the lowest peak value and accordingly suffers less from clipping. The SRRC pulse shaping filter with α = 1.0, on the other hand, that has the highest peak value gives the worst performance. Furthermore,

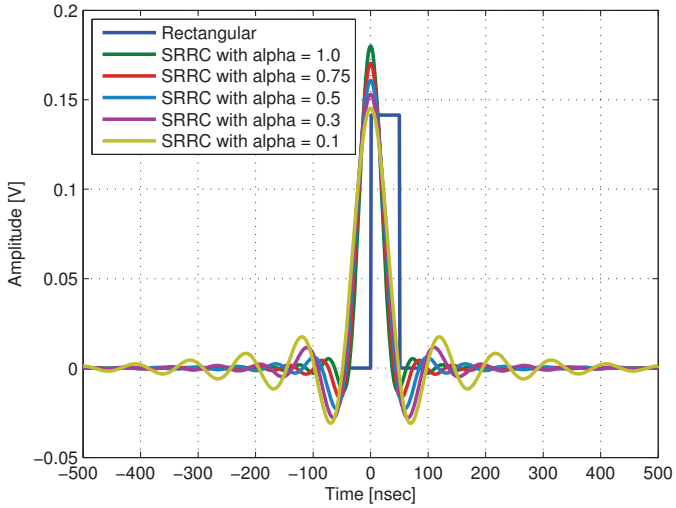


Figure 5: Time domain responses of considered pulse shaping filters.

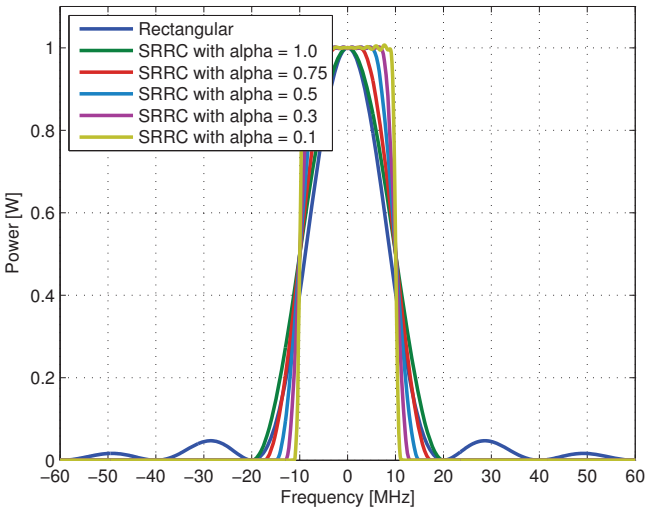


Figure 6: Frequency responses of considered pulse shaping filters.

it is observed that increasing the excess parameter leads not only inefficient bandwidth usage but also higher clipping distortion.

V. Conclusion

In this paper, we have investigated the impact of pulse shaping filter on the performance of DCO-OFDM based VLC system under a realistic indoor environment. We compare the BER performances of SRRC filters with different excess parameter values, especially in the clipping region of LEDs, through Monte Carlo simulations. Our results reveal that selection of excess bandwidth parameter determines not only bandwidth usage but also leads different clipping distortion.

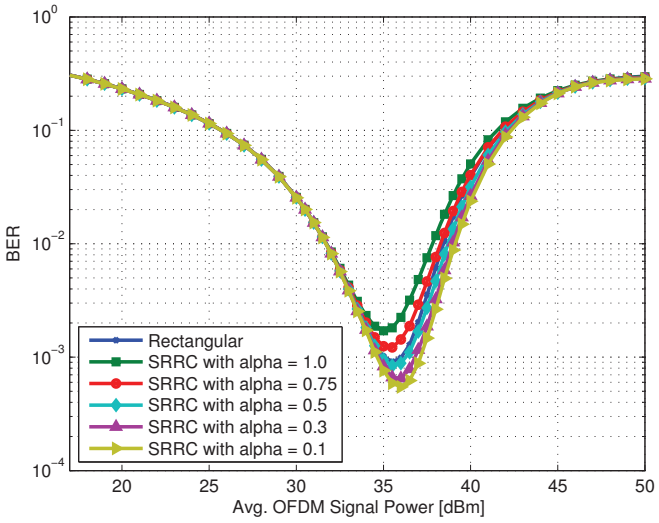


Figure 7: BER performances of different pulse shaping filter.

References

[1] S. Arnon, J. Barry, G. Karagiannidis, R. Schober, and M. Uysal, *Advanced optical wireless communication systems*. Cambridge university press, 2012.

[2] S. Wu, H. Wang, and C.-H. Youn, "Visible light communications for 5G wireless networking systems: from fixed to mobile communications," *Network, IEEE*, vol. 28, no. 6, pp. 41–45, 2014.

[3] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 100–107, Feb 2004.

[4] J. Armstrong, "OFDM for Optical Communications," *Journal of Lightwave Technology*, vol. 27, pp. 189–204, Feb 2009.

[5] J. Armstrong and A. J. Lowery, "Power efficient optical OFDM," *Electronics Letters*, vol. 42, pp. 370–372, March 2006.

[6] N. Fernando, Y. Hong, and E. Viterbo, "Flip-OFDM for Unipolar Communication Systems," *IEEE Transactions on Communications*, vol. 60, pp. 3726–3733, December 2012.

[7] D. Tsonev, S. Sinanovic, and H. Haas, "Novel Unipolar Orthogonal Frequency Division Multiplexing (U-OFDM) for Optical Wireless," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1–5, May 2012.

[8] D. Tsonev, S. Videv, and H. Haas, "Unlocking Spectral Efficiency in Intensity Modulation and Direct Detection Systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1758–1770, Sept 2015.

[9] R. Mesleh, H. Elgala, and H. Haas, "On the Performance of Different OFDM Based Optical Wireless Communication Systems," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, pp. 620–628, August 2011.

[10] S. Dimitrov and H. Haas, "Optimum Signal Shaping in OFDM-Based Optical Wireless Communication Systems," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pp. 1–5, Sept 2012.

[11] J. Armstrong and B. J. C. Schmidt, "Comparison of Asymmetrically Clipped Optical OFDM and DC-Biased Optical OFDM in AWGN," *IEEE Communications Letters*, vol. 12, pp. 343–345, May 2008.

[12] S. Dimitrov, S. Sinanovic, and H. Haas, "A comparison of OFDM-based modulation schemes for OWC with clipping distortion," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 787–791, Dec 2011.

[13] M. Uysal, T. Baykas, F. Miramirkhani, N. Serafimovski, and V. Jungnickel, *TG7r1 channel model document for high-rate PD communications*. doc: IEEE 802.15-15/0685r0, September 2015. Available: https://mentor.ieee.org/802.15/dcn/15/15-15-0746-01-007a-tg7r1-channel-model-document-for-high-rate-pd-communications.pdf.

[14] R. C. Kizilirmak, O. Narmanlioglu, and M. Uysal, "Relay-Assisted OFDM-Based Visible Light Communications," *IEEE Transactions on Communications*, vol. 63, pp. 3765–3778, Oct 2015.

The Influence and Role of CAD and Computer Mathematics in Engineering Education

S.T.Huseynov
Faculty of Air Transport, Dept. of Flying Apparatus and Aviation Engines,
National Academy of Aviation (NAA) of Azerbaijan and
Faculty of Mathematics & Informatics, Dept. of Informatics,
Ganja State University (GSU),
Ganja city, Azerbaijan Republic,
huseynovsahib@gmail.com

***Abstract* - The paper mainly has devoted to the important fields of modern engineering education in the aerospace fields, particularly in such disciplines as structural mechanics and application of computer modeling there. The author shares his experience in teaching and scientific research on some important issues about the role of modern software, which are concern to two different fields: CAD and computer algebra. In the paper it has been showed relevant original graphical results of computer modeling in aerospace structural mechanics via Ansys CAD software and number of examples with original cods in Maple computer mathematics system for analytical researches in theory of mechanical equilibrium.**

***Keywords* – CAD applications, computer mathematics, computer modeling, engineering education.**

I. INTRODUCTION

The CAD already on stretch more than thirty years is the most major and very important instrument of research for the engineers and researchers, which are busy in projection of and high tech machines and apparatus in various areas of the modern technics, [1]. It so happened that the first CAD system increasingly used in aerospace technology, and especially in the structural mechanics. Therefore, the preparation of the University in the field of specialists (engineers and researchers) is always the basic requirements have been brought to the teaching methods the appropriate CAD software. I myself can confirm that in the 80s of the last century (I was still a student of MSTU named after N.E.Bauman) in the education system of the former USSR while considerable attention was also paid to the mathematical training of students. Along with the study skills of working with CAD systems there were given the big attention for training of future engineers and researchers in programming methods too, though it was obvious very big gap with the US in computer technology. When I use the term mathematical training, really I mean not only the general theory of mathematical analysis and etc. (of very important parts of higher mathematics). But first of all, I mean the very important mathematical apparatus, which is the main mathematical tool of any modern CAD software even still these days - this is a finite element method (FEM), [2]. It's no secret that this the numerical discrete mathematical method which is a main computations tool of all CAD systems, it has applied in so many (if not all) engineering areas - including even in microelectronics. But, if not strange, due to the rapid development of computer technology (I mean the latest advances in hardware and software of modern PC) in the above areas of engineering education I personally watched some negative trends. I have a feeling that in a system of engineering education this negative trait in varying degrees in the widespread form exists in so many other engineering universities. When I have conducted a series of lectures and laboratory workshops on structural mechanics with the appropriate discipline for CAD, I came to the conclusion that the students which are absolutely unaware with mathematical foundations of the FEM (or they know this method very superficially) they may, with the help of an experienced teacher quite successfully to cope with the decision of quite complicated technical problems. Besides, worse of all, that the average student does not even have the necessary prior knowledge and techniques in the field of programming. To eliminate (or at least for reducing) these, in my opinion, negative phenomena I believe that it is necessary to add or increase computer mathematics application in engineering disciplines, for instance in case of solving of specific applied engineering problems. The goal of the present article consists to show strength, the capacity of these two different modeling tools and their essential features in engineering education: the methodical use of CAD and computer mathematics in solving of specific technical problems.

II. APPLICATION OF CAD SOFTWARE IN STRUCTURAL MECHANICS

Below are the results of calculation in the Ansys software (this software is classified as CAD) a rectangular wing aircraft flying at subsonic speed. In Ansys Workbench (this is an application of Ansys, where there's a possibility of so-called visual programming) the students solved this problem with my help on a practical training on the use of CAD in the structural mechanics, [3] and [4]. In the fig. 1 it is shown the geometric model of the object, which was built in a special module under name Geometry (this is Ansys Workbench

application). Although during first time of research the design of a rectangular aircraft wing may seem simple, but in fact this design in terms of structural mechanics is complicated enough. As it can be seen from fig. 1, the structure of this construction consists of 4 elements (it is known from aerospace engineering): we can see here stringers, spars, ribs, and wing skin. I have found that about 60 - 70% of the total work time students spend in the CAD environment Ansys Workbench is in the construction of this geometric model of the object (aircraft's wing structure with internal strengthening elements).

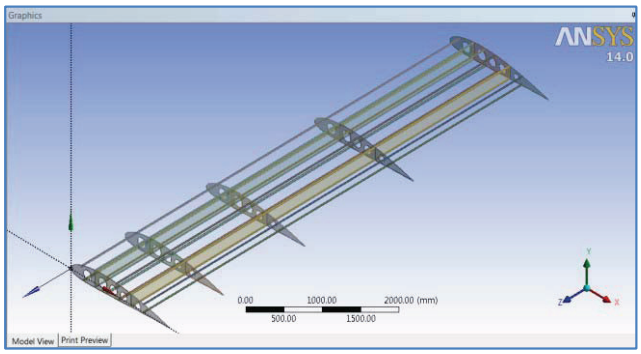


Figure 1. 3D graph of the plane wing with only inside strengthening constructive elements.

In the fig. 1 the wing's surface hasn't been shown (this was done for good visibility of structure of the all construction). But in the fig. 2 we can see the surface of the wing - now it is seem in semi-transparent form for a good visibility of internal reinforcement elements.

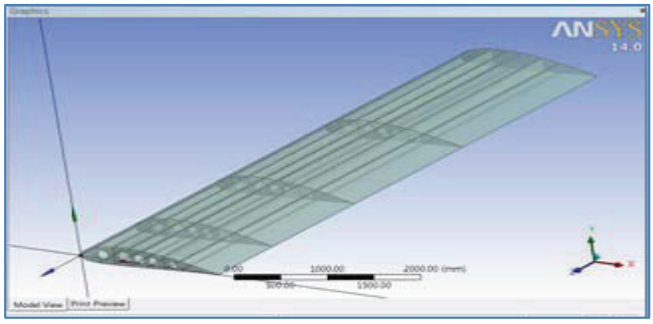


Figure 2. 3D graph of the plane's wing with outside and inside constructive elements of strengthening.

In the fig. 3 we can see result of automated discretization of the air volume, which is flowing surrounding of the wing's rigid body (we have chosen geometrical model of this air volume in form of rectangular parallelepiped). This procedure is carried out in the Fluid Flow (Fluent) module for modeling aerodynamic processes using the finite element method. More in detail, but in a 2D graph this discretized part of air which is flowing surrounding of the airfoil of wing is shown in fig. 4.

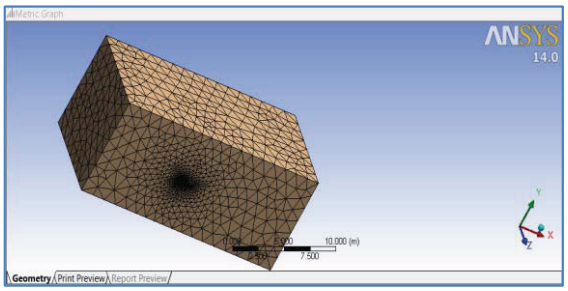


Figure 3. 3D graph of the automatically discretized (by FEM in Fluent Flow (Fluent) module in Ansys Workbench) air environment enclosed plane wing, which is geometrically has been modelled as the rectangular parallelepiped.

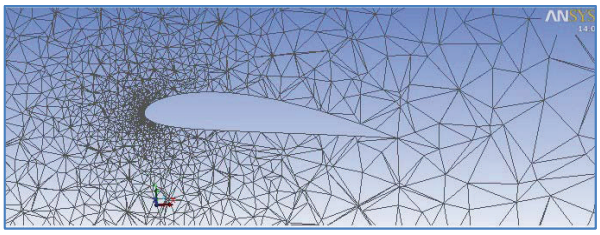


Figure 4. 2D graph of the discretized part of air which is flowing surrounding of the airfoil of wing.

Some of the results of aerodynamic calculations presented below in Fig. 5 and 6. In the fig. 5 we can see a 2D graph of the redistribution of the field of aerodynamic pressure on the airplane's airfoil in the cross-section at the attachment to the fuselage. 2D graph of the vector field of flow of velocity of the air around the airfoil in the same region of cross-section part of wing is shown on the fig. 6.

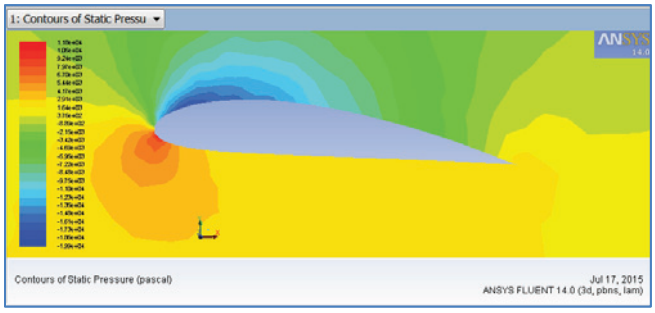


Figure 5. 2D graph of reallocation of a field of aerodynamic pressure on a plane wing airfoil section in a region at fixing with a fuselage.

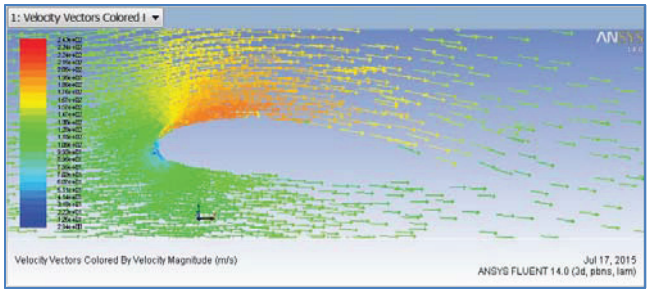


Figure 5. 2D graph of a vector field of velocity of the air stream, which is flowing around an airfoil section of the plane in a region at fixing with a fuselage.

After the end of the simulation of aerodynamics in module Fluid Flow (Fluent), in automatic mode, data is transmitted (the boundary conditions, etc.) to the module under name Static Structural, where according to the setting, which should be done preliminary by user, automatically will be solved the strength problem. In the fig. 7 shows a part of an airplane wing, which was fully discretized (with all the internal parts of reinforcement) according to automatic mode in an module Static Structural with using FEM analysis for modeling of structural mechanics problems (obtaining the stress-strain state of researched the construction).

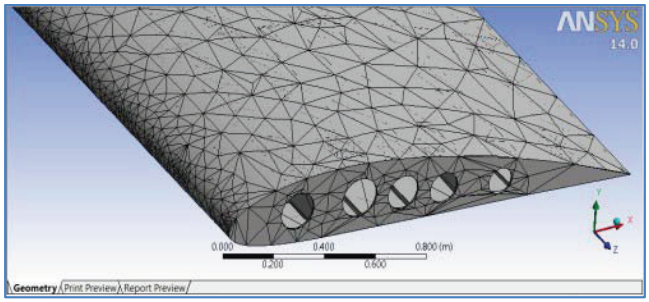


Figure 6. 3D graph of the part of a plane wing, which is completely discretized (with all internal parts of constructive durability elements) by FEM meshing procedure in an automatic regime.

The user (student) must previously to set the module Static Structural: firstly to select the appropriate approach for the FEM computations, and secondly to obtain (calculate) the most interesting options and the results of the simulation in graphical output. Below, in the fig. 7 is shown in 3D image form of the total deformation of the wing with the color palette according to the range of values of deformations along the wing surface.

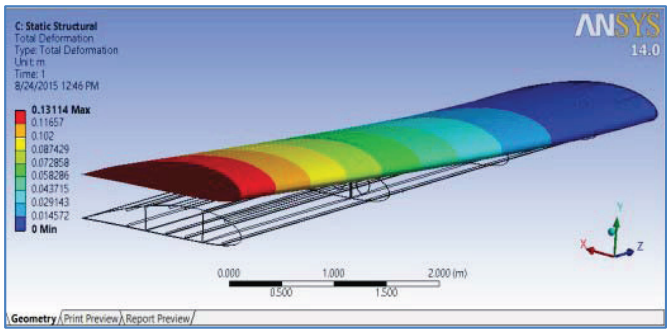


Figure 7. 3D graph of a plane wing's total deformation with color palette of computed values on the wing surface.

The computed equivalent stresses along wing surface are shown in the fig. 8 (by color palette of these values on the wing surface).

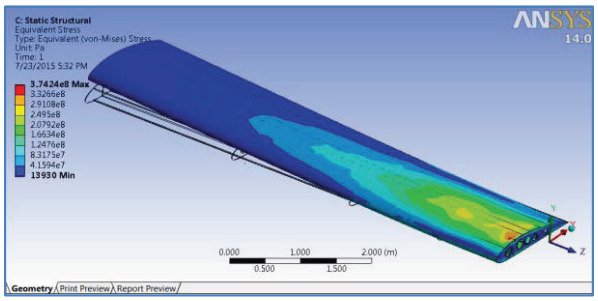


Figure 8. 3D graph of a plane wing's equivalent stress with color palette of computed values on the wing surface.

I think that it is possible to express it in simple terms, our problem considered in the category of multiphysics, you can imagine how a certain formula in the form *aeromechanics* = *aerodynamics* + *static structural*. To implement solution of this rather complex simulation problem in Ansys Workbench, it is necessary to make connection and adjustment of the above described modules for modeling of aerodynamics and the mechanical strength of the wing of the aircraft according to the diagram in fig. 9.

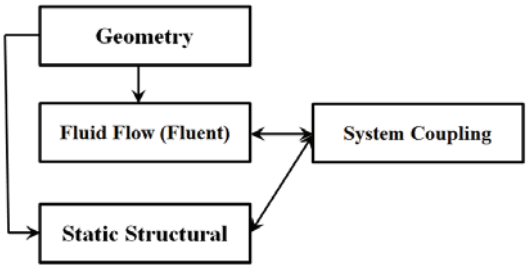


Figure 9. Diagram for modeling of multiphysics problem a plane wing's aeromechanics.

To get interesting results of simulations for researchers in the CAD environment of the Ansys Workbench (Ansys software application), of course, we have great opportunities, but under the theme and scope of this article, many other very important results aren't possible to show here. I think important to remind, that the considered problem actually is quite complex and time consuming, both in terms of mathematical content and in terms of continuum mechanics. In fact, about 10 or 15 years ago, computer simulation of such problems were available only for very few advanced researchers with access to the computers which have sufficiently performance and the appropriate software, and these conditions were not available to most technical universities.

III. Application of Analitical Methods for Solving of Equilibrium Problems in Structural Mecanics via Computer Algebra Computations

In this chapter we will try on the basis of linear formulation to solve with using analytical method the problem about equilibrium of the rod, which is loaded by longitudinal distributed load. We will adhere in accordance

with analytical methods, which are the author suggested in [6], fig. 10. In the problem is necessary to determine the critical load at which the rod loses stability.

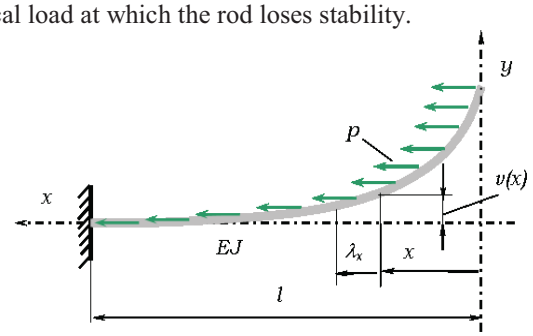


Figure 10. Buckling of the rod under longwise distributed load

To solve this problem analytically, we will use of advantages of computer algebra with symbolic programming in Maple. According to [1] we have the following linear ordinary differential equation (LODE) with one variable coefficient and let us write this equation directly via symbolic programming in Maple, [5]:

```
Rod_Eq_1:=diff(v(xi),xi,xi)+a^2*xi*v(xi)=0;
```

$$Rod\_Eq\_1 := \frac{d^2}{d\xi^2}v(\xi) + a^2\xi v(\xi) = 0. \qquad (1)$$

Let's denote $w(\xi) = \frac{d}{d\xi}v(\xi)$ - as angle displacement of the rod cross section; $v(\xi)$ - as transverse linear displacement of the rod; $\xi = x/l$ - as the dimensionless the lengthwise coordinate of the rod (or lengthwise axis of the rod); l - length of the rod; $a^2 = \frac{pl^3}{EJ}$ - constant coefficient; p – the lengthwise distributed load; E - is Yung's elasticity module of material; J - is moment of inertia of cross section of the rod.

Let's to solve ours LODE (1) analytically in Maple (in obtained symbolic solution (2) we can see Airy function):

```
Ans_Eq_2:=dsolve(Rod_Eq_2);
```

$$Ans\_Eq\_2 := w(\xi) = C_1 AiryAi\left(-(a^2)^{1/3}\xi\right) - C_2 AiryBi\left(-(a^2)^{1/3}\xi\right). \qquad (2)$$

Where, C_1 and C_2 are the integration constants.

Now we can find the first derivative of obtained analytical function $w(\xi)$ (really this derivative is the third derivative of transversal displacement of the rod).

```
nu(xi):=diff(w(xi),xi);
```

$$v(\xi) := -C_1(a^2)^{1/3}AiryAi\left(1, -(a^2)^{1/3}\xi\right) - C_2(a^2)^{1/3}AiryBi\left(1, -(a^2)^{1/3}\xi\right). \qquad (3)$$

Our boundary conditions we can present in the following form: $M := -EJ\left(\frac{d}{dx^2}v(x)\right) = 0$, at the point $x = 0$ and $\frac{d}{dx}v(x) = 0$ at the point $x = 0$. Or in dimensionless coordinates $\xi = x/l$ we have $w(\xi) = \frac{d}{d\xi}v(\xi)$ - the angle displacement of the rod's cross section; $v(\xi)$ – the transverse linear displacement of the rod. Thus, we can rewrite above described boundary conditions in the following form:

at the point $\xi = 0$ the boundary condition $\frac{d}{d\xi}w(\xi) = 0$, and at the point $\xi = 0$ the boundary condition $w(\xi) = 0$.

```
Eq1_BVP1:=eval(w(xi),xi=1)=0;
```

$$Eq1\_BVP1 := C_1 AiryAi\left(-(a^2)^{1/3}\right) + C_2 AiryBi\left(-(a^2)^{1/3}\right) = 0. \qquad (4)$$

```
Eq1_BVP2:=simplify(eval(nu(xi),xi=0))=0;
```

$$Eq1\_BVP2 := \frac{1}{2}\frac{(a^2)^{1/3}(C_1 3^{1/6} - C_2 3^{2/3})\Gamma\left(\frac{2}{3}\right)}{\pi} = 0. \qquad (5)$$

```
Ans_Int_Const:=solvefor[C1,C2](Eq1_BVP1, Eq1_BVP2);
```

$$Ans\_Int\_Const := \{C_1 = 0, C_2 = 0\}.$$

We have two alternate solutions:

1) $C_1 = 0$ and $C_2 = 0$, but we can't agree with all of them (with the solutions) because of physical and geometrical properties of the problem (this is so called trivial solution);

2) C_1 and C_2 aren't all of them equals to zero, but we know from the theory of linear algebra, that linear algebraic equations system (4) and (5) should have determinant with nonzero value. That is why we should to build a following determinant:

```
with(LinearAlgebra):
V:=Matrix([[AiryAi(-z),AiryBi(-z)],[3^(1/6),-3^(2/3)]]);
```

$$V := \begin{bmatrix} AiryAi(-z) & AiryBi(-z) \\ 3^{1/6} & 3^{2/3} \end{bmatrix}. \qquad (6)$$

Where we denote following expression $z = -(a^2)^{1/3}$.

We should equalize determinant (6) to zero.

```
Eq1:=Determinant(V)=0;
```

$$Eq1 := -AiryAi(-z)3^{2/3} - AiryBi(-z)3^{1/6} = 0. \qquad (7)$$

Thus, we have obtained transcendental equation (7). Let's solve this transcendental equation by using powerful features of Maple. To estimate probable range of numerical solution, we could plot this function, fig. 11.

```
plot(Eq2, z=0..10, linestyle=0, color=blue, thickness=4, axesfont = ["HELVETICA", "ARIAL", 20], title="Transcendental equation solution", titlefont = ["ROMAN", 24]);
```

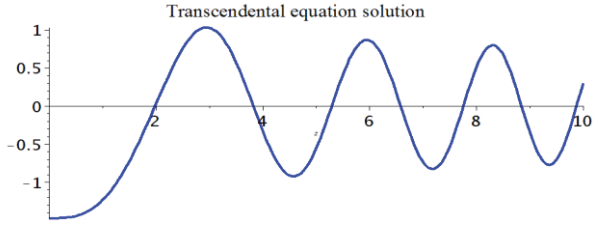


Figure 11. 2D graph of the transcendental function.

Now we can solve numerically in Maple the equation within convenient for us diapason. In other terms we should choose minimal root among of these numerically solved roots. `z[min]:=fsolve(Eq1=0, z, 0..3);`

$$z_{min} = 1.986352707 \qquad (8)$$

We have obtained minimal value of root of the equation. Now, at last, we can find critical value of lengthwise distributed load p due to result (8) and expressions

$$a^2 = \frac{pl^3}{EJ}, \quad z_{min} = -(a^2)^{\frac{1}{3}} .$$

To symbolically carry out it, we should use the following Maple algorithm:

```
a:=(p*l^3/E/J)^(1/2);
```

$$a := \sqrt{\frac{pl^3}{EJ}}$$

```
Eq3:=Eval(z[min])
```

$$Eq3 := 1.986352707\,(pl^3/EJ)^{1/3}$$

```
p[cr_exact]:=solvefor[p](Eq3);
```

$$p_{cr\_exact} := p = 7.837347434(EJ/l^3). \qquad (9)$$

The analytical result (9) equal to the result in [1] and according to the author this result is exact solution of the problem.

CONCLUSIONS

The main purpose of this article, as it was mentioned in the introduction - was to show how and for what purpose in the learning process of students of engineering universities in the fields of mathematical modeling should be applied, these two, very different in their characteristics, software: software with CAD environment and software in computer math. Problems which often arise on this subject in the educational process in some of the engineering universities, has been discussed in the same section of this article. In the first chapter of the article was set out a methodical approach to the use of Ansys software in the study of the quite complex task in the field of aeromechanics (synthesis of aerodynamics and structural mechanics) - definition of internal stress-strain state of rectangular wing of the aircraft, which is flying at subsonic speed. In the second chapter of this article was given analytical approach for solutions of another task in Maple (software of computer math) - the study of stability of rod under longitudinal compression. But a second, more an easy task, also belongs to structural mechanics. Hopefully, the attentive reader will understand that the first problem can't be solved using the software Maple, which is belong to the category of computer mathematics. On the other hand, here not exists an important reason to solve the second problem with using only Ansys software. But in this context it should be noted that although in theory this approach is possible, but is conjugate with very great difficulties. Ansys this is CAD system, which is adapted only for numerical simulation on the basis of FEM analysis, but Maple belong to another category, although it has quite powerful tools for numerical simulation, but this program has perfect advantages, mostly just for analytical calculations. Thus, the Ansys is a very powerful CAD system and, as we have seen it, this software does not have the versatility to research or study of all engineering problems (and it can't be limited to such applied engineering science as structural mechanics). Therefore, the neglect of some students, that analytical methods obsolete and they have not of no use (and sometimes some teachers and specialists are expressing this position) - this is absolutely not true. Because there is other, more rational approach, and I think this way is more objective. In the engineering education should be to maximize the opportunities in both directions: CAD systems (software such as Ansys), using a discrete numerical analysis of FEM and analytical calculations with using computer mathematics (for example, similar to using Maple).

ACKNOWLEDGMENT

I would like to thank to the prof. A.S.Samedov - vice rector of National Academy of Aviation of Azerbaijan, who really is inspirer of my scientific and pedagogical activity in NAA of Azerbaijan. And besides I wish also to thank the all referees which carefully have read the original manuscript, and in result they gave me theirs valuable comments, advises and suggestions. I believe, that theirs critical approach will improve the presentation of this paper.

REFERENCES

[1] S.T.Huseynov, N.V.Ibadov, A.A.Aslanov, "Interactive Teaching Methods of Mathematical Physics by the methods of Computer Visualization", 2012 6th International AICT Conference proceedings, pg. 405 – 409, Tbilisi, Georgia, 2012.

[2] O.S.Zienkiewicz, R.L.Taylor, "The Finite Element Method", Volume 1, 5th edition, Butterworth-Heienemann publication, 708 p., 2000

[3] H.Lee, "Finite Element Simulation with Ansys Workbench 14", http://myweb.ncku.edu.tw/~hhlee/Myweb_at_NCKU/Books.html

[4] İ.Çayiroğlu. "Aircraft Wing Analysis. Ansys Workbench 14" https://www.youtube.com/watch?v=qahFy0d43-s4;

[5] Kocsis, "Application of Maple ODE analyzer in the investigation of differential equations in higher engineering education", An International Journal for Engineering and Information Sciences, Vol. 2, pp. 177-183, Budapest, 2007, available: www.akademiai.com.

[6] A.S.Volmir, "Stability of Deformable Systems", in Russian, M.: "Nauka", 1967, p. 133 - 136.

Web application as support for learning methodology in informatics education

Peter Kuna
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
pkuna@ukf.sk

Martin Vozár
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
mvozar@ukf.sk

***Abstract* — In this paper, the authors focus on the using of web applications in informatics education. Applications are part of a new model of learning in Discrete Mathematics course in the study program in Applied Informatics. The knowledge that students acquire in the subject are key. Learning and acquiring new knowledge is one of the priority skills for personal development of every person and thereby directly affecting the other direction. In teaching, we should lay stress on that students not only memorized knowledge but understand them and then they would easily be able to use in practice.**

Keywords—learning methodolog; web application; informatics education

I. INTRODUCTION

The proposed methodology is intended to support and stimulate students' interest and to show them that math is necessary for their further studies. The first step in solving was to determine the merits of the various thematic units that are taught in Discrete Mathematics. These units have subsequently confronted with thematic units at other universities that have similar focus. Current teaching the runs support of established e-learning course. Based on the results of the authors of the report proposed a new structure for e-course and adapt content Discrete Mathematics requirements arising from the survey in the study program in Applied Informatics. During the design e-course experiences were utilized with the formation of branched tutorials. Linear training programs prescribe firm and binding sequence of steps in one line for all students. The curriculum takes over the small portions of information, but preferably only one in each step. Branched (nonlinear) tutorial allows in solving problems procedure varied [1]. Alternatives of the program ultimately lead to successful by tackling the problem. Every student passes through its own way, while length of the individual parts of the e-course corresponds to his personality, skills and talents. The attractiveness of materials also contributes newly designed web applications in which a student can directly test their theoretical knowledge. Based on the experiment, we verify the accuracy of us proposed methodology. The experiment was attended by 107 students. To evaluate the procedures and the used instruments we will use the method pre-test [2]. Upon completion of the course, learners have completed post-test. The results of both tests are statistically analysed and evaluated. We expect that this designed and created educational practices will have a significant impact on the level of acquired knowledge of students from different subjects. This would thus acquire knowledge had become longer-lasting. It should result in the updating of skills of students of informatics curricula.

II. REQUIREMENTS SPECIFICATION

LGraph is us developed a web application based on client-server architecture. Requirements for application LGraph us define methods MoSCoW. MoSCoW is a kind of numerical assignment and it is mentioned by DSDM Consortium [3], Hatton [4] [5], Tudor and Walter [6]. MoSCoW currently 17 incorporates into the software development methodology DSDM (Dynamic Systems Development Method). The idea of MoScoW is that it groups all requirements into four priority groups “MUST have”, “SHOULD have”, “COULD have”, and “WON’T have”.

- "Must Have", which means that this type of requirement must be met in the project, otherwise the project is deemed to have failed,
- „SHOULD have“ is a type of requirements that should be met in the project,
- „COULD have“ also means that this type of requirement should be met in the project, but the requirement of this type has a lower priority than a "Should Have",
- „WON‘T have“ requirements are not met in the project but can be implemented in the future versions.

III. LOGICAL ARCHITECTURE

LGraph is an application consisting of a presentation, application and data layers.

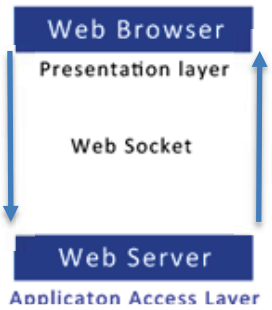


Fig. 1. Logical Architecture LGraph

A. Presentation layer

Presentation layer of application LGraph will represent only one document written in HTML, to which are attached two cascading style sheets and six script written in JavaScript. Emphasis will be placed on the simplicity of the user interface, so the page itself is divided into three parts only. The upper part will serve as a container for the main application menu, the bottom of which is their dimensions identical to the top, will be used to display short text, which provides information on the application itself. The third main part is further divided into two parts, on the toolbar, and container for HTML canvas, that space where the user will work with diagram graph. For the application and user interaction will serve modal dialog boxes that the user will be given an error or information messages or the user will need to enter the input data. The creation of a variety of visual effects that will make interactive application will be used scripting language JavaScript. In order to increase efficiency and transparency, the code will be used in a project known application framework "JQuery".

For applications and user interaction will serve modal dialog boxes that the user will be given an error or information messages or the user will need to enter the input data.

The main application menu will be placed in the upper right of the page and will be implemented as vertical pop-up menu. The menu will consist of three items, which will enable the creation of a chart, graph algorithm selection and display of help.

Toolbar that will allow the user to work with the diagram chart and play animations of graph algorithms will be placed over the container, in which will be located HTML canvas.

The main, and as regards the dimensions of the largest part of the user interface will form the container in which it will be located HTML element called "Canvas". In this area the user will be allowed, after selecting a suitable instrument to implement the proposal or modification of diagram chart to then be able to view an animation of the selected graph algorithm. The elements embedded in this space are objects created with the assistance of the application of the 'Raphael', which is used to facilitate work with vector graphics in the web browser. To each such object will dynamically, according to the needs that arise during the course of the application instance assigned Event Manager, which will be functions mostly having to worry about change in the appearance of these objects. Change the look of these objects in the occurrence of various events will make the application more interactive but mainly will play a key role while playing, which will demonstrate the different steps of graph algorithms.

B. Application layer

LGraph is a web application based on client-server architecture. The client will be implemented as a thin client, which will not know the logic of the application located on the server.

On the client side will be located six scripts, of which the four scripts embedded code that will be in charge of functionality and user interface management server communication. The other two scripts are freely available application framework that offered functionality will be used in the above four scripts.

Client-side script:

- "jquery-1.8.3.js" is now often used, freely available application framework that offers many features such as selectors, displaying animation, event management, AJAX and many others. Application will use mainly selectors and event management,
- "Raphael.js" is more freely available application framework designed to facilitate the work with vector graphics in the web browser, the functionality offered by the framework will be used to draw the vertices and edges of the graph chart and change their attributes,
- "websocket.js" script which will be implemented the coating object over WebSocket object, an instance of this object wrapping will offer a method for sending data to be parameters of this method on the server and will also implement a function which, when a message arrives from the server, based on type of report, sends management features designed to process this type of report. This script will also contain a function that will be responsible for creating an instance of the coating object and thus establish a WebSocket server connection,
- "menu.js" is a short script that will implement the functionality of the main application menu,
- "graph.js" will include the implementation of three objects:

  1. the coating object "Vertex", which will represent the top of the charts and graphs will be responsible for the parameters received in the constructor for creating an instance of the object type Raphael and display it in the canvas and at the same time offering functions enabling change the appearance of the object and functions implementing its behavior in the occurrence of various events. These so-called "event handlers" will be dynamically as appropriate object type raphael added or taken away, what will be possible by calling functions wrapping object.
  2. the second object implemented in this script will be also the wrapping object and will be named "Edge". The functionality instance of this object will be similar to the object "Vertex" unless they have to be adjusted to represent the edges of the graph chart.
  3. the last object will be called "Graph" and will represent a data structure to store a reference to an object instance of type Vertex and Edge. The methods of this object will be allow adding and removing instances of objects of the two types from the data structure and at the same time will be also be involved in the addition and removal from the canvas. This data structure will not be responsible for verifying whether or not it is possible to perform the operation selected, it is only a kind of container for the graph vertices and edges, but not the full implementation of the data structure, serving to represent a particular type of graph,

- "application.js" script which will include objects and functions assigned to matching "event handlers" each object according to the needs that arise during the application, showing step of animations, currency dimensional objects constituting the user interface. In addition, it will also include features that will be in charge of carrying out the necessary actions before formulate the rule which will be sent to the server and functions which will be the type of news that arrives from the server management sold to subsequently implement appropriate action.

The server will be deployed freely available server Apache Tomcat container that implements the Java Servlet and JavaServer Pages specifications and provides an environment for running Java code based purely on the Java platform. Code implemented on the server will represent applications and data application layer. This code is organized into three packages that still contain their sub package.

The package titled graph.web will consist of four packages:

- "servlets" will consist of classes that will be responsible for creating WebSocket connections with clients and create an object that will be the responsibility of selling the management object classes contained in the package handlers, upon arrival of the request from the client,
- "interfaces", will include interfaces that will be implemented by classes in the package handlers,
- "jsonmessages" will include classes that will be used by classes in the package handlers to deserialize strings in JSON format,
- "handlers", will include the implementation of the class whose instance on the type of request from a client before another instance of the management, the implementation of which is in the package and is designed to process this type of request from a client.

The main mission of the classes in these packages graph.web of package will be establish a connection with the client and on arrival requirements, create an instance of the object, which is intended for processing that type requirements and that this requirement creates an adequate response. The whole process is shown in the diagram in Fig 2.

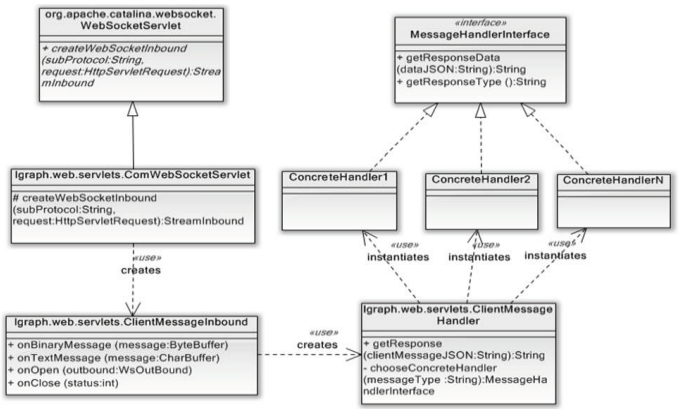


Fig. 2. The comunication process of the classes in package graph.web

After the arrival of the requirements for establishing a connection is created instance of the class "ComWebSocketServlet", which is inherited from class "WebSocketServlet", which is an implementation of WebSocket specification provided by Tomcat. Subsequently, creates an object of type "ClientMessageInbound" whose single attribute will be object of class "ClientMessageHandler". Class "ClientMessageInbound" will be implement the method "onTextMessage", which in its body will use the object class "ClientMessageHandler", namely the method of "GetResponse". If the connection is established successfully, the client can send the server requirements. Upon arrival requirements, which will be always text string in JSON format, will be called the method "onTextMessage" object class "ClientMessageInbound" in which the body to be called the method of the object "GetResponse" belonging to the class "ClientMessageHandler", which receives as a parameter the text string in JSON format. Method "GetResponse" of the object "ClientMessageHandler" after deserialization JSON string generated from the type of requirements instance one of the classes implementing the interface "MessageHandlerInterface" and give it the data necessary to construct a JSON string, which will be actually answer that clients expect. After receipt of the reply in the form of a text string in the method body "onTextMessage" object class "ClientMessageInbound" sends the reply to the client.

Classes in the package handlers will implement the interface "Message Handler Interface" that defines methods "getResponseData" and "getResponseType". Each such class in the method body "getResponseData" will work with object instances of classes contained in package of package lgraph.graph subpackage. During the execution of the body of the method object is created in one of the classes contained in package "jsonmessages", which will converted to string in JSON format, which in this form will contain data for the client. Then, determine the type of responses produced over yet code execution method "getResponseData".

The second package called lgraph.graph, will consist of five packages:

- "factory", will consist of one class which instance will be a factory for creating objects of classes contained in package "implementations",
- "interfaces", will include interfaces that will implement the factory and abstract classes from which will inherit classes in package "implementations",
- "datastructures" will consist of classes that will represent the data structures used by classes in packages "implementations" and "algorithms",
- "implementations", will consist of two classes will inherit from the abstract class in a package "interfaces", which will be an abstract representation of the data structure of the graph, the two classes will be concrete implementation of data structures, graph and digraph,
- "algorithms", will consist of classes that will implement the behavior of graph algorithms.

These packages under package of jgraph.graph will be subject to the exceptions contained classes in packages "factory" and "algorithms" create a Data Access Layer. Classes included in the package "algorithms" will implement graph algorithms and also will offer methods to obtain references to objects whose attributes will be data structures containing the steps carried out the algorithm.

The last package called "lgraph.test" will have no subpackage. This package will be space for classes that need to be established for the purposes of testing the functionality of other classes in the project.

C. *Data layer*

Data Access Layer will be located on the server. When connected to the clients and the creation of an instance of class "ClientMessageHandler" will also create an instance of the class "UserWorkspace" will throughout the connection with the client to serve as a container for references pointing to instances of objects that store data with which the client works. Object of class "UserWorkspace" will therefore be as one of its attributes to store a reference to the object of one of graph structures, which create runtime instances of the client application requests. The class of which instances will form data layer applications will be included in the package "lgraph.graph.datastructures" and "lgraph.graph.implementations". The package "lgraph.graph.interfaces" will present the definition of abstract class from which they inherit classes in package "lgraph.graph.implementations".

IV. Actual or Anticipated Outcomes

The environment in which to display the user interface of is a web browser. The entire user interface consists of a single HTML document. When designing the emphasis was on its simplicity and intuitiveness to enable the user to control the application did not lead to a chaotic and quickly learn how to work with it. The actual environment is divided into three parts: the main application menu, toolbar and desktop. Desktop forms the main part of the user interface. In essence, is a desktop canvas element, which is part of the HTML5.

Application Design, we also tried to create so that in future, its possible extension achievable without major complications. For high priority application we consider that allows the user to design a chart of graph or digraph according to its own ideas. The user has the opportunity to view an interactive animation of the selected graph algorithm on several different diagrams, which can help them to understand the philosophy of the algorithm.

A. *Application environment LGraph*

The entire user interface consists of a single HTML document. When designing the emphasis was on its simplicity and intuitiveness, so that when you first use the application to control the user did not lead to a chaotic and quickly learn how to work with it. The main menu of the application is implemented as vertical pop-up menu.

The application LGraph allows creating a new chart, change how will be transferred any of its vertices move from one place to another desktop. Another functionality of this application is that user can choose which graph algorithm would like to apply to his designated diagram.

The application Graph offers a choice of four graph algorithms. Browse in depth, width, Dijkstra's algorithm (fig 3), Kruskal's algorithm. Assuming the user is proposed diagram for wishing to apply algorithms and features of the diagram meet the conditions for applying the selected graph algorithm on this diagram, the user will be informed through a dialog box, which inputs awarded, so that the implementation of graph algorithms successful. If all the conditions are met and the user entered correct entries will be able to run the animation. Animation can be suspended for an arbitrary period of time and then in her resume. During playback the animation is not allowed any user to manipulate the diagram, which suggested.

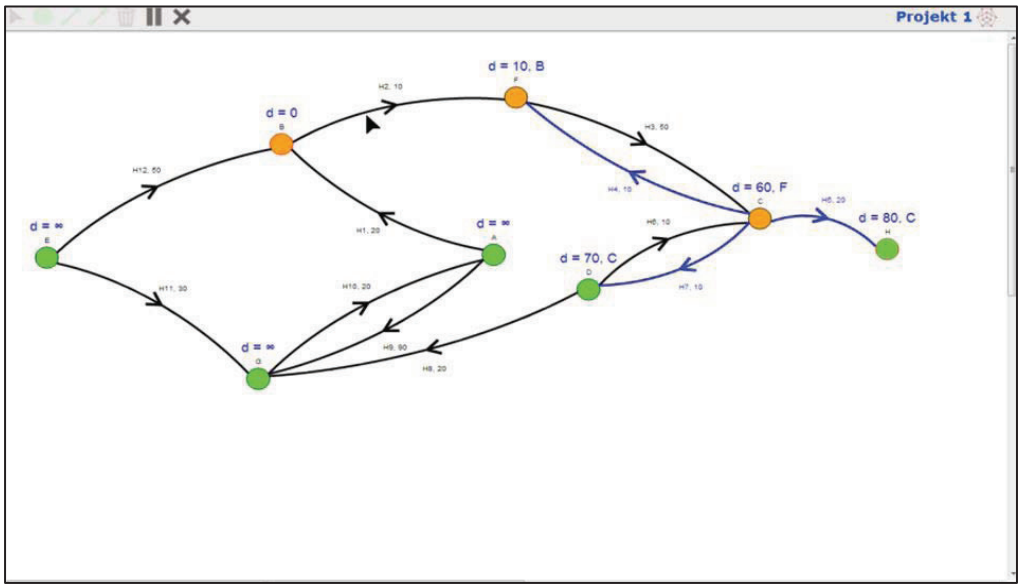


Fig. 3. Demonstration of Dijkstra's algorithm

V. DISCUSION

Experiment is divided into several consecutive steps [7], [8]. The actual implementation of the experiment was conducted at Constantine the Philosopher University in Nitra. In winter semester students completed the subject DM1 and DM2 in summer semester. For the teachers were prepared complete materials (input and output tests for students questionnaire) while inform them about the conditions of the experiment (not to disclose to the students that are involved in the experiment; the final test; the experimental class prepared to use e-course (form could choose discretion and availability)).

For the development of the pre-test students had one lesson (90 minutes), at the beginning we dealt with the necessary instructions for drafting questions and organizational issues. This has reduced the time to prepare questions for 80 minutes playing time. Pre-test consisted of the 15 questions. The correct answers students should mark or circle from the options available to them.

Posttest 1 students graduated after waking all thematic units from subject DM1 within the examination period at the end of the winter semester the first year of study. Posttest 2 students completed again after waking all thematic units from subject DM2 at the end of the summer semester of the first year. The aim posttest 1 and posttest 2 was to verify that students can use various subjects discussed in thematic units DM1 and DM2.

The lowest percentage students achieved in the pre-test and vice versa the highest in the second posttest. Statistically significant differences were demonstrated between pre-test and posttest. Graph average and confidence interval (Figure Test) visualizes point and interval estimation of the mean. We can see that the lowest percentage students achieved the pre-test and the highest percentage was in the second posttest.

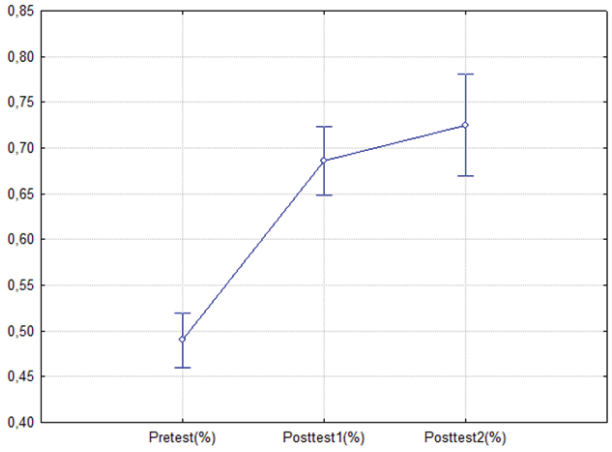


Fig. 4. Categorized graph - point, interval estimate of average; y-axis is the score, the x axis represents the type of test.

VI. CONCLUSION

For verification of knowledge was used with repeated measures experiment, that is, we tested not only the differences between independent samples but also between related samples. The same experiment was used by authors in works [9] and [10]. The experiment was conducted using pre-test and two post-tests. Pre-test students completed in the first year. Post-test 1 completed at the end of the first year and post-test 2 at the end of the second year of. In the experiment, a total of 107 of students participated. The lowest percentage was reached in the pre-test and vice versa the highest in the second post-test. Statistically significant differences were demonstrated between pre-test and post-test. This confirmed the natural assumption that completing the pre-test and two post-tests increasing the level of knowledge.

ACKNOWLEDGMENT

This paper is published thanks to the financial support of the project UGA VII/9/2016, Named: Educational applications as part of the support for teaching mathematical subjects.

REFERENCES

[1] Kuna, P., Turčáni, M., Applying the principles of e-course methodology creation for the subject Discrete Mathematics in the information learning environment In. Interactive Collaborative Learning/Engineering Pedagogy 2012, ICL 2012 : Proceedings from International Conference, ISBN 978-1-4673-2426-7, pp.1-4 [CD-ROM].

[2] Vozár, M., - Kuna, P., Innovative education with e-learning support in European project In: ICETA 2012 : IEEE 10th International Conference on Emerging eLearning Technologies and Applications, November 8 – 9, 2012, The High Tatras, Slovakia. - Košice : Technical University, 2012. - ISBN 978-146735122-5, pp. 395-399.

[3] DSDM Consortium, DSDM public version 4.2. Retrieved 6 Jan, 2009, from www.dsdm.org

[4] Hatton, S., Early prioritisation of goals. In Advances in conceptual modeling – Foundations and applications, 2008 pp. 235-244.

[5] Hatton, S., Choosing the right prioritisation method. 19th Australian Conference on Software Engineering, 2008, pp.517 - 526.

[6] Tudor, D., & Walter, G. A., Using an agile approach in a large, traditional organisation. Proceedings of AGILE 2006 Conference (AGILE'06), 2006, pp.367-373.

[7] Kapusta, J., Munk, M., Turčáni, M., Experimental comparison of adaptive links annotation technique with adaptive direct guidance technique. In WEBIST 2009 - Proceedings of the 5th International Conference on Web Information Systems and Technologies, 2009, pp. 250-255.

[8] Hvorecký, J., Drlík, M., Munk, M., Enhancing database querying skills by choosing a more appropriate interface. In 2010 IEEE Education Engineering Conference, EDUCON 2010, art. no. 5492434, 2010, pp. 1897-1905.

[9] Kapusta, J., Munk, M., Turčáni, M., Evaluation of adaptive techniques dependent on educational content. In 4th International Conference on Application of Information and Communication Technologies, AICT2010, 2010, art. no. 5611791.

[10] Hvorecký, J., Drlík, M., Munk, M., The effect of visual query languages on the improvement of information retrieval skills. In Procedia - Social and Behavioral Sciences, 2 (2), 2010, pp. 717-723.

Construction of individual educational trajectory of students based on e-learning

Rinat R. Ibatullin
Kazan Federal University
Elabuga, Republic of Tatarstan, Russia
Rinat.Ibatullin@kpfu.ru

Ellina S. Anisimova
Kazan Federal University
Elabuga, Republic of Tatarstan, Russia
ESAnisimova@kpfu.ru

***Abstract*— The most important task of the modern education system is implementation of individual educational trajectory of each student. In this regard, adaptation of the educational process to personality development of each student becomes extremely necessary. One of the ways of tackling the ensuing problem is e-learning as a model of organization of interaction between the teacher, student and information technologies. Currently, there are developed plenty of distance learning courses personifying learning process. This article describes the use of e-learning courses on the example of such disciplines as "Economic Informatics" and "Numerical Methods", their influence on the construction of individual educational trajectory of students; as well the paper analyses the impact of e-learning courses on the quality of students' academic achievement.**

***Index Terms*— individual educational trajectory, distance learning, e-learning course.**

I. Introduction

In modern education the student's personality becomes of primary significance. Therefore, it is important to adapt learning process in educational institutions to the student's personality development peculiarities. "The Concept of Modernization of Russian Education" states that the most important task of education is formation of spirituality and culture, initiative, independence, tolerance, the ability of self-education and self-development, successful socialization and active adaptation in the labor market."

Therefore, the priority of modern education is implementation of individual educational trajectories based on personality, interests and needs of students.

Individual educational trajectory is a personal way of fulfilling each student's personal potential in education [1].

Implementation of individual trajectory is determined by ongoing training activities. There are three directions of individual educational trajectory [2]:

1. Informative direction, i.e., formation of individual educational trajectory providing the student with an opportunity to choose the content of education and its level that best suit his or her abilities, needs and interests.

2. Activity-oriented direction, i.e., forming individual educational trajectory based on modern pedagogical and IT technologies.

3. Procedural direction, i.e., organizational aspects of the educational process.

Due to the active use of modern information technologies, the most relevant direction for implementation of individual trajectories is activity-oriented, carried out via information technologies.

Various types of information technologies can be used in educational activities, but the most popular one among them is e-learning technology. This article will consider the impact of e-learning on the construction of individual educational trajectory of students, as well as the peculiarities of their application in the educational process.

II. The influence of e-learning technologies on the construction of individual educational trajectory of students

E-learning technologies are understood as educational technologies implemented mainly using information and telecommunication technologies in indirect (distant) or partly mediated by interaction of the student and the teacher (Article 32 of the Law of the Russian Federation of July 10, 1992 № 3 266-1 «On Education ").

E-learning technologies are used by the educational institution for the purpose of giving students the possibility of mastering educational programs directly at the student's place of residence or location.

Nowadays institutions of higher education widely use e-learning courses in various disciplines [3, 4]. Of course, their use does not exclude laboratory and practical lessons, monitoring, intermediate and final certification by direct interaction of the teacher with the students.

Learning groups in higher education institutions can be heterogeneous. The following groups of students can be distinguished:

- students with a faster pace of development;
- students with poor health;
- students with low learning motivation and learning difficulties;
- gifted students.

E-learning helps each student of a particular group to design their own individual trajectory. For each category of students, the teacher can distribute tasks of a certain difficulty level. Depending on the student's success teacher can use

individual teaching methods, offer additional educational materials, links to information resources [5]. The fact that the student can take the course in accordance with his/her specific rate of learning the material at any convenient time proves to be very advantageous. The teacher sets deadlines to complete certain sections (units) and during that time (before testing) the student must study the material of the chapter and do all the necessary tasks. This is the way individual educational trajectory of the student is built up – in accordance with the student's rate and inclination to a given subject area.

The use of e-learning courses doubtlessly helps to activate students' self-sufficient work [6], to increase their interest in the discipline, to form skills of independent analysis of the material and problem statement and solution, abilities for self-education. Thus, e-learning contributes to the formation of undergraduate students' competencies, providing a competitive advantage as the young professionals.

Another important function of e-learning courses is the possibility of monitoring - the teacher can always monitor the activities of each student (his accomplishments, academic achievement) in real time mode. [7]

III. Peculiarities of e-learning courses use in education on the example of 'economic informatics' and 'numerical methods'

The authors of the article have developed such e-learning courses as "Theoretical foundations of the Information Society", "Economic Informatics", "Numerical Methods", "Economic and Mathematical models", "Mathematics", etc. on LMS Moodle platform. Obviously, the use of distance learning course in this or that discipline has certain peculiarities.

As an example, let us consider the use of e-learning course in "Economic informatics" discipline. This discipline is studied by the first year students in the direction of training 38.03.01 Economics (profile: accounting and auditing).

The main purpose of the discipline is acquaintance of students with software and hardware used by economists; forms and methods of their use in various spheres of activity. Students receive practical skills of work with modern hardware and software tools.

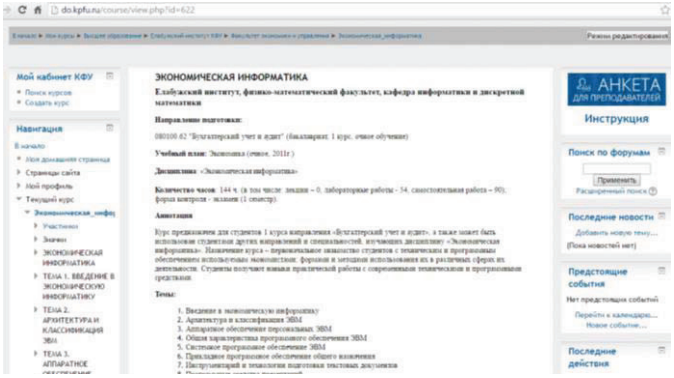


Fig. 1. "Economic Informatics" e-learning course homepage

The specific feature of this discipline is the absence of lectures; here, laboratory work and self-study are carried out only. In this regard, the use of e-learning course in the study of this discipline is well reasoned. After learning the unit of the discipline, students do a test (if it is a theoretical unit), or laboratory work and tasks for self-study (if it is a practical unit) and send them for checkup. In this case, they can send not only text documents, but also graphics, presentations, videos, etc. Students are allowed to do the tasks at any convenient time before the deadline specified by the teacher. In the study of each unit, students are provided guidelines, material for study, a glossary of relevant terms, a checklist of questions and tasks for laboratory work. The teacher checks the laboratory work of students and grades them.

It should be noted that in case of any problems aroused, the student may seek the assistance of the teacher. For that purpose, the ensuing e-learning course uses a tool named "Forum".

Tasks for self-study are creative in nature and every student performs them individually in accordance with his/her preferences regarding the topic and level of training. The tasks offer such kinds of activity as development of the business plan of a company (unit "Tools and Techniques of Text Documents Preparation "), development of the advertising the company ("Presentations Software"), designing the company database ("Database Management Systems"), etc.

The e-learning course allows the student to get acquainted with the proposed material in advance (laboratory works, tasks for self-study) and to begin their performance. This course does not exclude classroom teaching in which the student can consult the teacher on the topics of discipline in case they have any difficulties.

Also the e-learning course provides the student with the possibility of having videoconferences, using Skype, when the teacher conducts the lesson online. It is very convenient in case a student is unable to attend classes for health reasons.

When studying of the discipline is completed, the authors conducted a study to determine the impact of the e-learning course on the academic performance of students. For this purpose, the students were divided into two groups.

The first group had studied the discipline in the traditional form performing laboratory work in the classroom, and individual assignments for self-study, but without the use of a distance (e-learning) course, while the second group had studied the discipline with the use of a distance course, classroom lessons had been used to assist students in arising difficulties.

After completion of the discipline in both groups, the students did a test that contained questions on all units studied, and a practical assignment.

In Fig. 2 we can see a comparative analysis of the quality of academic performance of the two groups in all the studied units. Fig. 2 shows the number of students in each group who have done at least 70% of the tasks of the unit correctly.

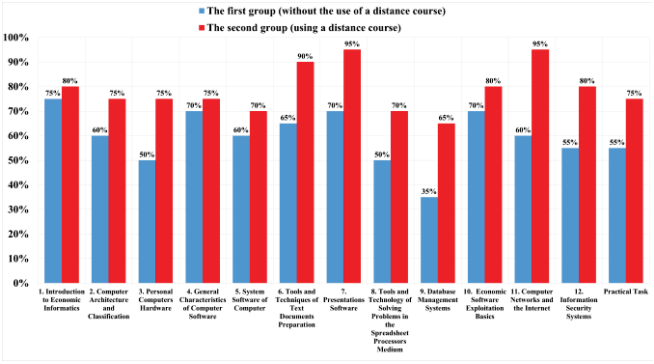


Fig. 2. A comparative analysis of the quality of academic performance in two groups – students who studied the discipline without the e-learning course, and with it, in the study of the discipline "Economic Informatics"

The students were offered to do practical tasks in text documents processing, presentations, developing databases, working with spreadsheets. They were evaluated in percentage terms (0-100%) according to the degree of their performance.

In Fig. 2, you can see that the quality of academic performance of the group, who studied the discipline with the use of an e-learning course, in all units exceeds the quality of the group, who studied the discipline in the traditional form. This is due to the greater autonomy of students, who studied the discipline distantly, and personification of the learning process when using an e-learning course.

The discipline "Numerical methods" is related to solution of applied mathematical problems. For its studying, the authors also have developed an e-learning course. It is intended for second-year students of the direction of training 09.03.03 – Applied Informatics, and can also be used by students of other specialties, studying the discipline "Numerical Methods". The course provides basic information about classical numerical methods for solution of various applied problems: direct and iterative methods for solving systems of linear algebraic equations; of nonlinear algebraic and transcendental equations; interpolation; differentiation and integration; solution of differential equations.

It should be noted that this discipline has a mathematical focus and is quite difficult for students. However, the distant course contains sources of additional information, allowing to study a particular material in detail. In addition, there is a summary of the course, full and short variants of lectures, a glossary with basic terms on the discipline, methodical recommendations for teachers and students, the list of electronic resources, a news forum, a laboratory work forum, and examination questions.

In studying each unit of the discipline students are provided with lecture material, assignments for laboratory work and samples of their performance, a checklist, glossary, guidelines for the study of this topic.

Obviously, this course allows classroom activities, where after students' preliminary study of the material, the teacher can assist students in most difficult issues and tasks.

After finishing this discipline there has been conducted a study to determine the impact of e-learning course on the quality of students' academic performance.

Initially, the students were divided into two groups. The first group studied the discipline in a traditional form, without using an e-learning course: took a course of lectures and did laboratory work in the classroom. The second group on the contrary studied the discipline using an e-learning course, classroom sessions were used to assist students in most difficult issues.

Later on, the students took an examination which included a test containing questions on key topics, and a practical task. Fig. 3 shows a comparative analysis of the quality of academic performance between the two groups in all the studied units. The figure shows the number of students (in percentage terms) in each group who have performed at least 70% of the tasks in the unit correctly.

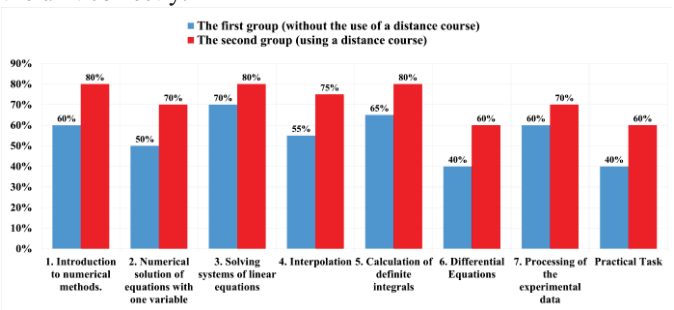


Fig. 3. A comparative analysis of the quality of academic performance in two groups – students who studied the discipline without the e-learning course, and with it, in the study of the discipline "Numerical Methods"

The practical task presupposed development of different programs:

• numerical solution of equations with one variable by various methods (by trials, bisection, chords, tangents, Newton, iteration, apportionment, combined method);

• numerical solution of linear equations system;

• interpolating by splines;

• integration;

• solution of differential equations by the Euler and Runge-Kutta methods;

• processing of experimental data using least-squares method, mean-value method.

The practical task was evaluated in percentage terms (0-100%) in accordance with the degree of its performance.

It can also be seen that the academic performance of the group, who studied the discipline using a distance course, in all units exceeds the performance quality of the group, who studied the discipline in the traditional form.

The students studying disciplines "Economic Informatics" and "Numerical Methods" using e- learning courses were surveyed to assess the efficiency of e-learning courses use in the learning process.

The results of the study are the following:

• 95% of students' state that they are ready to learn with the use of e-learning courses;

• 100% of students are completely satisfied with the possibility of permanent access to learning materials / tasks;

• 90% of students feel free in communication with classmates and teachers on the forums.

Thus, students are positive about the effectiveness of the use of e-learning courses in the learning process.

IV. Conclusion

Studying a discipline using a distance learning course helps to design an individual educational trajectory of each student in accordance with his/her capacity to learn, level and rate of training indeed. Individual rate and ability to work at convenient time for students, self-statement of potential problems and their solution – all of these have a positive impact on learning, contribute to formation of abilities for self-education that is one of the fundamental criteria for selection of staff by modern HR managers. An ability for self-education characterizes an employee as a person working for the result, striving to find the most effective and more cost-efficient ways of achieving the goal. Nowadays, people with high levels of abilities for self-education are more valuable and competitive in the labour market.

Thus, introduction of distance learning elements in the traditional educational process is one of the most important directions of education system modernization. Thanks to the support of the teacher who recognises the right of each trainee's personal learning objectives, students will be able to set a specific goal of their work on an individual educational program and to express it in the form of the intended result [8].

References

[1] A. V. Khutorskoy, "Didactic heuristics. Theory and technology of creative teaching", Moscow: MSU press, 2003, 258 p.

[2] S. A. Vdovin and I.M. Kungurova, The Nature and directions of individual educational trajectory implementation, Science, 2013. (Viewed on: 20.08.2015).

[3] A. A. Andreev, Didactic basis of distance learning, Moscow, 1999.

[4] A. A. Babaev, Informatization of learning courses maintenance, Problems of Economy and New Technologies of Teaching. Proceedings of the VIII International scientific-practical conference, vol. 1, Saint Petersburg: MBI, 2009, pp. 69-71.

[5] A. S. Belkin, “The Situation of Success. How to create it", Moscow: Education, 1991, 176 p.

[6] N. V. Volzhenina, "Organization of self-sufficient work of students in the learning process", Barnaul: Publishing house Alt. University press, 2008, p. 59.

[7] E. S. Polat, M. Yu. Buharkina, and M. V. Moiseyev, "Theory and practice of distance education", Moscow: Academy, 2004, 416 p.

[8] A. P. Trapezina, "Educational program – the student’s route", Saint Petersburg, 2000, 228 p.

Professional Training in Embedded Systems and its Promotion

Alena Hašková
Faculty of Education
Constantine the Philosopher University in Nitra
Nitra, Slovak Republic
ahaskova@ukf.sk

Dirk Van Merode
Technology and Design, Campus de Nayer
Thomas More Mechelen-Antwerpen University College
Sint-Katelijne-Waver, Belgium
dirk.vanmerode@thomasmore.be

***Abstract* — Tendency to eliminate man´s intervention in the controlling processes in different hardware applications as much as possible supports implementation of embedded systems in a broad scale of industry sectors. Consequently this leads to an increased demand for embedded system specialists at the labour market and at the same time to increased requirements on their competences. Issues related to curricula topics and professional training methodology in embedded systems at higher education institutions have been investigated in the international European Tempus project with a consortium of partners from the European Economy Area and target partners in Ukraine, Georgia and Armenia (DesIRE - *Development of Embedded System Courses with implementation of Innovative Virtual approaches for integration of Research, Education and Production in UA, GE and AM,* 2013 - 2016). In the paper there is presented a description of the project content together with an overview of the carried out activities and achieved main results.**

***Index Terms* — Embedded systems, tertiary education, training courses, curricula innovation, international co-operation.**

I. Introduction

To eliminate man´s intervention in the controlling processes in different hardware applications as much as possible currently so-called embedded systems are used. These systems are used in various sectors, starting from aerospace industry, through automobile industry, logistics, automation, medical devices up to consumer electronics such as cameras, mobile phones, microwaves, washing machines, television including satellite facilities. In principle they consist of combination of software and hardware components.

An embedded control system is specially designed and produced for a particular electronic device. Logically, a specially designed system cannot be taken off and simply used in other devices. Its usage in other electronic devices requires a completely new design and programming of the system, i.e. creation of a new embedded system.

Nowadays, two main branches of the embedded system development are recognised:

- adjustment of classical PC sets with high computing power and their use as embedded systems (embedded control systems for CNC control machines and production of assembly lines, industrial computers with PLC control functions, special HMI modules for industrial production),
- significant increase of computing power of hardware components primarily designed and developed for embedded systems (Raspbery PI 3, Arduino, Udoo X86).

Both development trends have extended possibilities of using embedded systems in practice. At the same time, they increased a demand for programmers able to unify operational systems as well as software development tools for these systems with programming aids developed for programming the desktop, internet, or mobile applications. Issues tackling this area of technology applications have been reflected in the four-year international project 544091-TEMPUS-1-2013-1-BE-TEMPUS-JPCR *Development of Embedded System Courses with implementation of Innovative Virtual approaches for integration of Research, Education and Production in Ukraine, Georgia and Armenia* (DesIRE, 2013 – 2016)

II. Background of the Project

Similarly to its title – *Development of embedded systems courses with the implementation of innovative virtual approaches focused on the integration of research, education and production in the Ukraine, Georgia and Armenia* – activities of the DESIRE project have been aimed at teaching courses (subjects) covering the embedded system topics at higher education institutions in Ukraine, Georgia and Armenia.

The main coordinator of the DESIRE project ([1], [2], [3]) is Thomas More Mechelen-Antwerpen University College (TMMA, Belgium). Other project partners are:

- Ilmenau University of Technology (IUT, Germany),
- Constantine the Philosopher University in Nitra (UKF, Slovakia),
- State Technical University in Zaporizhzhya (Ukraine),
- Crimean Engineering and Pedagogical University, Simferopol (Ukraine),
- Donbass State Engineering Academy, Kramatorsk
- (Ukraine),
- Boris Grinchenko Kyiv University (Ukraine),
- National Polytechnic University of Armenia, Yerevan (Armenia),

- Yerevan State University of Architecture and Construction (Armenia),
- Yerevan Telecommunication Research Institute (Armenia),
- Georgian Technical University, Tbilisi (Georgia),
- Ivane Javakhishvili Tbilisi State University, (Georgia).

Fig. 1. The DesIRE logo

The DesIRE project has got together all the involved partners to mutual co-operation supported by EU. The project has been aimed at modernization of tertiary education in countries neighbouring with the European Union (West Balkan Area, Area of East and Central Asia, North African Area and Area of Middle East). It should contribute to the intensification of synergy between HEIs and business enterprises in the partner countries and at the same time also to the extension of co-operation in the field of science, research and education among the countries involved.

The main aim of the project is to introduce a change from theoretically oriented teaching of technical subjects relevant to embedded system issues at the HEIs in the partner countries to practice-oriented and competence-based teaching ([10], [9], [6], [5], [8]).

III. Project Methodology

The main general aim of the project, a change from theoretically oriented teaching of some technical subjects to practice-oriented and competence-based teaching at the HEIs in the partner countries, has been intended to be achieved exploring good practice and expert activities of the partners from the European Economy Area. But because teaching courses in embedded systems design and production requires a very specific, qualitative and substantial knowledge growth in the target HEIs (which ensure inflow of high-skilled people to the labor market), to keep this knowledge relevant to the local and international labor market, the HEIs have had to strengthen their links with relevant branches of industry. This is the reason why the main, general aim has been concretized in following particular specific aims:

- to produce a comparative analysis of current curricula of the relevant subjects (courses) taught at the target HEIs and competences expected by the labor market in embedded systems,
- to develop relevant and up-to-date course study materials,
- to supply supporting hardware for a physical embedded systems laboratory, virtual and remote laboratories in the target HEIs, Ukraine, Georgia, and Armenia,
- to train teachers in both the EU and partner countries institution to develop their skills to train students to work with embedded systems required by the labour market,
- to run a pilot teaching period with the new material and to assess its results.

Achievement of the stated specific aims has been supported by activities structured into five project workpackages (WP):

WP1: *Analysis of current curricula and competences in Embedded Systems in the target countries*,
WP2: *Curricula modification and courses development*,
WP3: *Implementing a (virtual) learning environment in embedded systems design engineering*,
WP4: *Retrain university teachers in embedded systems design engineering and virtual learning platforms*,
WP5: *Pilot teaching/operation*.

The workpackage WP7 *Dissemination and Enterprise Collaboration* has served partially also as a supporting workpackage to the workpackages WP1 and WP2 and the other two project workpackages have been designed completely for the project operation activities (WP 6 *Quality Assurance and Quality Control*, WP8 *Management of the project*).

IV. Project Results

Although the project has not been finalized yet and some of its activities have been still under development and checking, most of the specific aims can already be referred as achieved and completed.

WP1 Analysis of current curricula and competences in Embedded Systems in the target countries

The aim of this workpackage was to elaborate an analysis of the current curricula of the subjects dealing with embedded systems engineering in the target higher education institutions (project partner institutions in UA, GE and AM). The analysis was performed based on content (what has been taught in the embedded domain) and competences (which of the competences are priority to be developed) criteria. At the same time surveys were carried out to find out requirements of stakeholders in embedded systems engineering in the partner countries and European Economic Area. For the purposes of these analyses as stakeholders were chosen local industry, teachers and students. Outputs of these activities were reports with recommendations on competences expected in embedded systems in labor market. Consequently it was agreed to renew and upgrade existing curricula using the newly developed teaching material in the target HEIs. An overview of the main points resulting from the agreements can be summarized as it follows [7]:

- All HEIs from all countries involved in the project decided to collaborate closely.

- Each partner HEI has the necessary human, education, information and logistic support for a high-quality training of students in the field of embedded systems.
- Analyses of the labor market show the importance of training of specialists in embedded systems.
- University students are familiar with contemporary approaches in education, however, they are rarely used, and, in general, virtual labs are better known than remote experiments.

WP2 Curricula modification and courses development

Following the findings and conclusions resulting from the activities in the WP1 and expected competences by stakeholders, within the WP2 the project partners from the European Economic Area had the task to develop relevant course materials for different subject topics in embedded systems engineering.

The project partner TMMA (Thomas More Mechelen-Antwerpen University College, Belgium) was responsible for formal theoretical and practical approach to different domains of embedded systems (embedded hardware, embedded software, CAD/CAM/CAE for embedded systems, digital signal processing, legislation on and certification of embedded end products, quality engineering). The intention was to develop modular courses which will allow external stakeholders to get certificates in the specific subjects.

The project partner IUT (Ilmenau University of Technology, Germany) was responsible for the development of course material in e-learning, virtualization and remote laboratories.

The project partner UKF (Constantine the Philosopher University in Nitra, Slovakia) was responsible for the quality of the methodology used in the curricula and the educational process and its main task was to provide course material on teaching technical studies and soft skills for engineers [4].

The task of the target HEIs in the partner countries was to adopt the developed teaching materials and incorporate them into the syllabi and curricula of relevant subject courses taught within different bachelor and master study programs.

Finally three main thematic modules with courses and teaching materials for related subjects were developed.

The first module *Hardware for Embedded Systems* consists of a mixture of ordinary classroom teaching and practical exercises using the new equipment. Courses in this module are:

- Microcontrollers,
- Digital Electronics,
- Digital System Design,
- Embedded Communication,
- Sensors, Actuators and Interfacing.

The second module *Software for Embedded Systems* consists as well of a mixture of theory and practical exercises structured into the following courses:

- C for Embedded Systems,
- Embedded Software Development,
- Embedded Operating Systems,
- GUI Development,
- Multicore Programming,
- Software Testing.

The third module *CAD/CAM/CAE for Embedded Systems* train students in ECAD electronic design, ALTIUM, MCAD structural design and Pro Engineer.

In addition to the three main thematic modules, seven additional courses were developed:

- Digital Signal Processing,
- Remote Labs and Virtualization,
- Legislation on and Certification of Embedded End Products,
- Management and Marketing of Embedded End Products,
- Quality Engineering,
- New Teaching Approaches in Engineering,
- Soft Skills for Engineers.

For the target HEIs in the partner countries to be able to plan the implementation of the new teaching materials, the project partners have at first developed syllabi for each course stating the topics, the amount of students` work, the aims, the learning outcomes, the prerequisites and so on [1].

WP3 Implementing a (virtual) learning environment in embedded systems design engineering

Within the WP3 the target HEIs have received supplies of the new equipment for training in embedded systems design engineering. The key role was hold by the project partner IUT (Ilmenau University of Technology, Germany). Its task was to plan, coordinate and implement a virtual learning environment for teaching and studying embedded systems design engineering, in particular by means of remote labs and virtualization.

List of materials and equipment to be ordered by the project coordinator (Thomas More Mechelen-Antwerpen University College, TMMA Belgium) was prepared in co-operation with the project partners IUT and UKF (list of materials and equipment for ESD was prepared mainly by TMMA, for remote labs mostly by IUT). Determining key performance parameters were stability and accessibility of the remote lab, active students and teacher usage. Other success indicators were the number of key competences which can be trained with the e-learning environment and the self-studying results of students.

As follows from the above-mentioned in frame of the WP3 the target HEIs were provided with the new equipment for both hands on training (so-called Embedded Systems Design Lab) and remote and virtual training (Goldi Remote Lab). The target HEIs received this up-to-date equipment and additional computers, network and server equipment, to support implementation of the newly developed teaching materials and innovated syllabi. In each target HEI 8 set-ups were supplied, with one PC and five different embedded systems. In addition, two HEIs invested in a low-entry robot system, to teach students principles of embedded programming, to make them enthusiastic about STEM (sciences, technology, engineering and mathematics). Other universities also invested in a 3D

printer, to support the work in CAD-systems and to completely close the loop to a functional prototype of a real product.

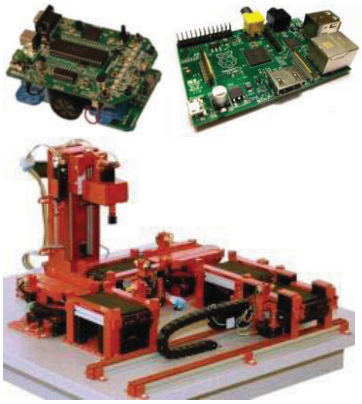


Fig. 2. Embedded systems delivered to the target HEIs

WP4 Retrain university teachers in embedded systems design engineering and virtual learning platforms

To make the academic staff in the target HEIs aware and able to implement the required changes in the learning environments of their home institutions was one of the specific aims of the project. To achieve this aim – to make the academic staff able to train students to work with embedded systems in accordance with the labor market`s requirements - several training activities for the university teachers both in the EU and target institutions in the partner countries on the use of the newly created courses on embedded system design and the virtual environment were planned.

Within the WP4 teachers of the target institutions from the partner countries were trained in the EU. Training on teaching methods was carried out at UKF, training on virtual learning systems and remote labs setup at IUT and training on embedded systems design at TMMA.

Training on teaching methods *New teaching approaches in Engineering* was held at Constantine the Philosopher University in Nitra (Slovakia) on September 14 – 21, 2014. The program of the training consisted of activities covering the following topics:

- The Bologna Process and European Higher Education Area,
- Interactive forms of learning supported by computer technology,
- E-learning methods in education,
- Information technologies in the remote experiment,
- Informatics minimum for users of remote experiments in LMS,
- Interactive forms of learning supported by computer technology,
- Use of remote labs and interactive forms of learning,
- Experiences with the use of the remote experiments in the university technology education,
- Demonstration of Flow Code Buggy and Lego Mindstorms.

Trainings in virtual learning systems and remote labs setup *The hardware for embedded systems* and *Remote and virtual labs* were held at Ilmenau University of Technology in Ilmenau (Germany) on April 13 – 25, 2015. The program of the whole training consisted of activities covering the following topics:

- Presentation of DesIRE lab components,
- Introduction to Quartus II (Altera),
- Hardware-oriented design with Quartus II,
- Introduction to Arduino and Arduino shields,
- Demonstration of Arduino applications,
- Usage of Arduino shields with other components,
- Introduction to the Remote Lab Management systems,
- Requirements for the design of online labs compliant with ISA,
- Examples of labs running on ISA,
- GOLDi components and installation,
- GOLDi – online experiments,
- GOLDi – usage of different web-control units,
- GOLDi – security and safety aspects in online labs,
- Overview of the main drawbacks concerning labs and RLMSs.

Training in embedded systems design was held at Thomas More University College in Sint-Katelijne-Waver on June 04 – 20, 2015. The program of the training consisted of activities covering the following topics:

- Formula Flowcode Buggy,
- Creo Workshop,
- Altium Workshop,
- Altium Academic Roadshow,
- C for Embedded Systems,
- Embedded Communication,
- Embedded Software,
- EmSys Research Group,
- Digital Signal Processing,
- Multicore Programming,
- Embedded Operating Systems.

Fig. 3. Fig. 5 – Photo from TMM summer course

WP5 Pilot teaching/operation

The final output of the whole project should be implementation of the new curricula and modules in the study programs in the target higher education institutions. To enable fast and efficient dissemination of the course material, master classes were planned for a contact group of relevant stakeholders (teaching staff, students, external stakeholders). The goal of the master classes, which are still being delivered (as the project has not been finished yet) has been presentation and demonstration of the new teaching infrastructure (ESD, e-learning, virtualization, remote laboratories) and to it related teaching processes and teaching materials. To facilitate the institutional transition to the new curricula and course modules, train trainers sessions have been organized in the target HEIs by the attendees of the trainings held in Slovakia, Germany and Belgium for efficient dissemination of the gained knowledge.

V. Conclusion

The biggest obstacle for smooth life of the project has been the political destabilization in Ukraine, which has had serious consequences on the project processes. The most serious impact of the destabilization on the project processes has been change in the legal status of the organizations located in the illegally occupied and annexed Crimea and Sevastopol area by the Russian Federation. These facts caused that Crimean Engineering and Pedagogical University, Simferopol (CEPU, Ukraine) was excluded from the project by the EACEA. This has been, of course, very regrettable not only for the excluded partner but also for the rest of the project partners and mainly for the project management (e.g. a part of the budget was frozen). Another consequence of the political destabilization in Ukraine with a serious impact on the project life are obstacles regarding running business in Ukraine (because of the unstable currency and difficult import regulations which are factors strongly influencing e.g. payments for mobility and staff costs and supplies of the relevant equipment and material to the Ukrainian target HEIs).

Despite of the above-mentioned obstacles the project management believes in successful completion of all project activities and the whole project, too.

References

[1] DesIRE Project website: http://tempus-desire.eu/

[2] DESIRE, "Informational bulletin". Avalaible at: http://zntu.edu.ua/uploads/tempus/DesIRE/20141015/Informational_Bulletin_1.pdf

[3] DESIRE, "Development of embedded system courses with implementation of innovative approaches for integration of research, education and production in UA, GE, AM". Flyer. Available at: http://tempus-desire.thomasmore.be/data/promotion/DESIRE_ flyer2.pdf

[4] A. Hašková, P. Kuna, E. Malá, and D. van Merode, "How to improve professional competences of technical subject teachers". Application of information and communication technologies - AICT2015, 9th IEEE International conference proceedings, Baku: Designing and publishing department of Qafqaz University, 2015, pp. 536 – 540.

[5] T. Kozík, M. Šimon, P. Kuna, P. Arras, and G. Tabunshchyk, "Internet security of remote experiments". Proceeding of the International scientific – practical conference. Kazan, 2015.

[6] T. Kozík, M. Šimon, and P. Kuna, "Remote experiment in terms of view of didactics of education". Didactics of informatics, Dydaktyka informatyki, Rzeszow: WUR, 2014, pp. 200-2011.

[7] D. Van Merode, G. Tabunshchyk, P. Arras, and K. Henke, "New teaching approaches in embedded system courses". International symposium on Ambient intelligence and embedded systems, 24 - 26 September, 2015, Oostende, Belgium.

[8] N. Morze, O. Buinytska, B. Hrytseliak, "Embedded systems, software for their design in the training of future teachers". Computing and information technology in educational institutions. 3/2014, 51, pp. 68-74.

[9] A. Parkhomenko, "Analysis of the specialized platforms application features in the embedded systems design". Proceedings of annual scientific-practical conference ZNTU Science Week-2014, Zaporizhzhya: ZNTU, 2014, pp. 121-122.

[10] G. V. Tabunshchyk, "New technologies in the training of embedded systems' specialists". Proceedings of the international scientific conference Internet-Education-Science-2014: IES-2014. Vinnitsa, 2014, pp. 248-250.

Impact Factor Penalized by Self-Citations

Rasim Alguliyev
Institute of Information Technology
Baku, Azerbaijan
r.alguliev@gmail.com

Ramiz Aliguliyev
Institute of Information Technology
Baku, Azerbaijan
r.aliguliyev@gmail.com

Nigar Ismayilova
Institute of Information Technology
Baku, Azerbaijan
nigar.ismayilova.iit@gmail.com

***Abstract*—This paper proposes a new version of impact factor. The proposed version penalizes JCR IF by journal self-citations. The penalized IF was experimented on 26 journals from Thomson Scientific JCR for 2013 and compared with the JCR IF. For these journals rank correlation between penalized IF and JCR IF is equal to 0.63.**

***Index Terms*—Impact Factor, weighted IF, journal self-citation, penalized IF**

I. Introduction

Currently the most popular indicator for journal evaluation is impact factor. Impact factor (IF) is a fraction, where numerator is the total number of citations in the current year to articles published in this journal in two previous years and denominator is the number of articles published in the two previous years [10]:

$$\mathrm{IF}_j^t = \frac{C_j^t}{A_j^{t-1} + A_j^{t-2}}, \quad (1)$$

where IF_j^t is the impact factor of the journal j in the year t; A_j^{t-1} and A_j^{t-2} are the total number of articles published in journal j in the two previous years $t-1$ and $t-2$; C_j^t is the total number of citations received by journal j in the year t.

In spite of impact factor's popularity there are many adverse effects of ranking journals by IF. Neuberger & Counsell [16] have counted several limitations regarding the use of impact factors, as manipulations by authors and editors, disambiguation with peer-review process, distinction between the qualities of journal articles, etc. Ways for boosting the prestige of a journal, in other words, "tricks" of engineering and IF manipulating–which increases nominator and decreases denominator in equation for IF calculation were pointed by Falagas & Alexiou [7]. Unweighted citations also have been noted as main disadvantage of IF, so in IF calculation the weights for all citations is one and it does not depend on impact of cited journal [17]. Different modifications of IF, weighted by cited journals' IF were proposed by Habibzadeh & Yadollahie [11], Sombatsompop et al. [23], Alguliyev et al. [1, 2], and Buela-Casal [5]. In [9] constructed mathematical model of generalized weighted impact factor.

Besides unweighted citations in calculation of impact factor not considered following factor: Weight fractions of self-citations – in certain cases, authors are forced to cite articles from the same journal. Although some of self-citations may be legitimate, they can distort the scientific literature and opinion of science policy-makers [3]. Fassoulaki et al. [8] found that high rate of self-citations actively influences on impact factor of journal. In this paper, penalized IF (PIF), considering weight fractions of self-citations is proposed.

II. Related Work

Self-citations of journal are citations to the journal by articles published in the same journal. Yu & Wang [26] analyzed that journal editors may distort value of impact factor by increasing the number of self-citations. In [18] has been stated that coerced journal self-citation is unethical and if unchecked it will continue to falsify the value of IF. Campanario [6] had analyzed journals selected from list of Journal Citation Reports –during 1998-2007 years and reported that, 54% of IF increases and 42 % of IF decreases depend on number of journal self-citations. It is has been found that in some journals number of self-citations is the main part of all citations. In these cases, self-citation has the potential to distort the real influence of the journal in given subject [25]. Demanding of authors by journal editors to use citations for articles published in this journal during previous years [4, 20, 22]. Gonzales & Campanario [14] noticed that journal editors publish many editorial materials with large number of self-citations and by this way they increase IF value. Also large amount of review materials published in journal and decreasing number of other articles increases IF value [21]. Krauss [12] offered to take self-citation rates into account for journal rankings to prevent further intentional increasing journal self-citations. Seglen [19] remarked selective journal self-citation as one of problems associated with the use of journal impact factors. Cases of coercive self-citation by peer reviewers also have been analyzed [24].

One of the most important values for determining the journal impact is distribution of citations among citing sources. Leydesdorff & Bornmann [13] offered calculating fractional citations for normalizing journal impact by cited sources. Fractional impact calculation of journal is important for defining differences between small and large sets of cited sources. The audience factor proposed by Zitt & Small [28] is a fractional indicator of journal citations which takes into account cited journal impact and affecting measures. It is a weighted impact factor which addressed to field-discrepancies by citing-side normalization, in contrast both with post-facto field-normalization and influence measures [27]. For solutions of the problem related by citation frequencies between different sciences in [15] proposed Scopus's Source Normalized Impact per Paper.

III. Proposed Impact Factor

Distortion cases of IF by self-citation have been noticed above. For more adequate ranking of scientific journals in this section we proposed a new impact factor penalized by self-citation (PIF) which decreases by increasing self-citations:

$$\mathrm{PIF}_j^t = \mathrm{IF}_j^t \times \log\left(\frac{C_j^t}{sC_j^t}\right), \quad (2)$$

where sC_j^t is the number of self-citations of journal j in the year t to the articles published in the two previous years $t-1$ and $t-2$.

IV. Experiments

The proposed PIF is calculated for 26 journals from Thomson Scientific JCR for 2013 and compared with the JCR IF. Table 1 describes bibliometric characteristics of the journals.

TABLE 1. Bibliometric Characteristics of Journals

#	Journal	Self-citations	Number of citations in 2013	Number of articles published in 2011 and 2012
1	Applied Numerical Mathematics	4	259	241
2	Scientometrics	146	800	497
3	Pattern Recognition	161	2391	513
4	Pattern Recognition Letters	57	1276	542
5	IEEE Transactions on Pattern Analysis and Machine Intelligence	82	3806	401
6	Neural Computation	26	448	226
7	Swarm Intelligence	6	62	27
8	Neural Processing Letters	7	130	77
9	Artificial Life	1	74	48
10	Cognitive Computation	11	112	89
11	Computer Speech And Language	6	168	68
12	Fuzzy Optimization and Decision Making	14	82	46
13	Genetic Programming and Evolvable Machines	8	91	45
14	International Journal of Applied Mathematics and Computer Science	45	255	136
15	Journal of Ambient Intelligence and Smart Environments	12	92	74
16	ACM Transactions on Applied Perception	4	78	40
17	ACM Transactions on Knowledge Discovery from Data	1	102	37
18	ACM Transactions On Information Systems	3	152	42
19	ACM Transactions on the Web	5	196	39
20	ACM Transactions on Sensor Networks	2	144	54
21	ACM Transactions On Software Engineering And Methodology	2	165	37
22	IEEE Transactions on Computational Intelligence and AI in Games	11	169	50
23	IEEE Transactions on Dependable and Secure Computing	2	395	143
24	IEEE Transactions on Autonomous Mental Development	15	107	54
25	World Wide Web: Internet and Web Information Systems	17	119	59
26	Journal of Informetrics	93	459	145

In experimental analysis was calculated penalized IF for selected journals and created new ranks were compared with their ranks by JCR IF's (Table 2).

TABLE 2. INDICATORS OF JOURNALS AND THEIR RANKS

#	Journal	2013 JCR IF		PIF	
		Value	*Rank*	*Value*	*Rank*
1	Applied Numerical Mathematics	1.036	25	1.876	13
2	Scientometrics	2.274	4	1.680	15
3	Pattern Recognition	2.584	3	3.028	3
4	Pattern Recognition Letters	1.062	23	1.434	17
5	IEEE Transactions on Pattern Analysis and Machine Intelligence	5.694	1	9.490	1
6	Neural Computation	1.694	8	2.094	12
7	Swarm Intelligence	1.833	6	1.859	14
8	Neural Processing Letters	1.237	16	1.570	16
9	Artificial Life	1.930	5	3.608	2
10	Cognitive Computation	1.100	20	1.109	23
11	Computer Speech And Language	1.812	7	2.622	6
12	Fuzzy Optimization and Decision Making	1.000	26	0.768	26
13	Genetic Programming and Evolvable Machines	1.065	22	1.125	22
14	International Journal of Applied Mathematics and Computer Science	1.390	13	1.047	24
15	Journal of Ambient Intelligence and Smart Environments	1.082	21	0.957	25
16	ACM Transactions on Applied Perception	1.051	24	1.356	20
17	ACM Transactions on Knowledge Discovery from Data	1.147	18	2.304	10
18	ACM Transactions On Information Systems	1.300	15	2.216	11
19	ACM Transactions on the Web	1.595	10	2.541	8
20	ACM Transactions on Sensor Networks	1.463	12	2.717	5
21	ACM Transactions On Software Engineering And Methodology	1.472	11	2.821	4
22	IEEE Transactions on Computational Intelligence and AI in Games	1.167	17	1.385	18
23	IEEE Transactions on Dependable and Secure Computing	1.137	19	2.610	7
24	IEEE Transactions on Autonomous Mental Development	1.348	14	1.150	21
25	World Wide Web: Internet and Web Information Systems	1.623	9	1.372	19
26	Journal of Informetrics	3.580	2	2.482	9

Rank correlation between proposed indicator and JCR IF is not so high, is equal to 0.63.

Let's compare journals from the same disciplines.

- Journal of Informetrics and Scientometrics the take 2nd and 4th positions according to their IFs calculated by Thomson Scientific. According to PIF ranking they have transferred to the 9th and 15th positions, respectively. The high penalty for self-citations of these journals associated by the few number of journals investigated scientometrics problems. On the other hand, all scientific disciplines can be analyzed by scientometrics methods; it explains the high positions by the distribution of citations.
- Swarm Intelligence (6th) and Artificial Life (5th) – the first journal have penalized by self-citations and have lost 8 positions, besides the second journal have moved forward 3 positions, because of insignificant number of self-citations.

ACKNOWLEDGEMENT

This work was supported by the Science Development Foundation under President of the Republic of Azerbaijan – **Grant № EIF-2014-9(24)-KETPL-14/02/1.**

REFERENCES

[1] R.M. Alguliyev, R.M. Aliguliyev, and N.T. Ismayilova, "Impact factor weighted by 5-year impact factor", Problems of Information Technology, no.2, pp.31-40, 2015.

[2] R.M. Alguliyev, R.M. Aliguliyev, and N.T. Ismayilova, "Weighted impact factor (WIF) for assessing the quality of scientific journals". arXiv preprint arXiv:1506.02783, 2015.

[3] T.F. Babor, K. Stenius, S. Savva, and J. O'Reilly, J. "Publishing addiction science: a guide for the perplexed", Wates Way, Brentwood, UK: Multi-Science Publishing, 2008, 229 pp.

[4] R.A. Brumback, "Impact factor wars: episode V – the empire strikes back", Journal of Child Neurology, vol.24, no.3, pp.260-262, 2009.

[5] G. Buela-Casal, "Assessing the quality of articles and scientific journals: proposal for weighted impact factor and a quality index", Psychology in Spain, vol.8, no.1, pp.60-76, 2004.

[6] J.M. Campanario, "Large increases and decreases in journal impact factors in only one year: the effect of journal self-citations" Journal of the American Society for Information Science and Technology, vol.62, no.2, pp.230-235, 2011.

[7] M.E. Falagas and V.G. Alexiou, "The top ten in journal impact factor manipulation" Archivum Immunologiae et Therapiae Experimentalis, vol.56, no.14, pp.223-226, 2008.

[8] A. Fassoulaki, A. Paraskeva, K. Papilas, and G. Karabinis, "Self-citations in six anaesthesia journals and their significance in determining the impact factor", British Journal of Anaesthesia, vol.84, no.2, pp.266-269, 2000.

[9] A. Ferrer-Sapena, E.A. Sánchez-Pérez, L.M. González, F. Peset, and R. Aleixandre-Benavent, "Mathematical properties of weighted impact factors based on measures of prestige of the citing journals", Scientometrics, vol.105, no.3, pp.2089-2108, 2015.

[10] E. Garfield, "The agony and the ecstasy — the history and meaning of the journal impact factor", International Congress on Peer Review and Biomedical Publication, 2005, pp.1-22.

[11] L. González and J.M. Campanario, "Structure of the impact factor of journals included in the Social Sciences Citation Index: citations from documents labeled "editorial material", Journal of the American Society for Information Science and Technology, 58(2), 252–262, 2007.

[12] F. Habibzadeh and M. Yadollahie, "Journal weighted impact factor: a proposal", Journal of Informetrics, vol.2, no.2, pp.164-172, 2008.

[13] J. Krauss, "Journal self-citation rates in ecological sciences", Scientometrics, vol.73, no.1, pp.79-89, 2007.

[14] L. Leydesdorff and L. Bornmann, "How fractional counting of citations affects the impact factor: normalization in terms of differences in citation potentials among fields of science" Journal of the American Society for Information Science and Technology, vol.62, no.2, pp.217–229, 2010.

[15] H.F. Moed, "The Source-Normalized Impact per Paper (SNIP) is a valid and sophisticated indicator of journal citation impact", Journal of the American Society for Information Science and Technology, vol.62, no.1, pp.211-213, 2011.

[16] J. Neuberger and C. Counsell, "Impact factors: uses and abuses", European Journal of Gastroenterology & Hepatology, vol14, no.3, pp.209-211, 2002.

[17] G. Pinski and F. Narin, "Citation influence for journal aggregates of scientific publications: theory, with application to literature of physics", Information Processing & Management, vol.12, no.5, pp.297-312, 1976.

[18] Jr, N.C. Romano, "Journal self-citation V: coercive journal self citation – manipulations to increase impact factors may do more harm than good in the long run", Communications of the Association for Information Systems, vol.25, pp.41-56, 2009.

[19] P.O. Seglen, "Why the impact factor of journals should not be used for evaluating research", British Medical Journal, vol.314, no.7079, pp.498–502, 1997.

[20] A. Sevinc, "Manipulating impact factor: an unethical issue or an Editor's choice?", Swiss Medical Weekly, vol.134, nos.127-128, p.410, 2004.

[21] M. Siebelt, T. Siebelt, P. Pilot, R.M. Bloem, M. Bhandari, and R.W. Poolman, "Citation analysis of orthopaedic literature; 18 major orthopaedic journals compared for Impact Factor and SCImago", BMC Musculoskeletal Disorders, vol.11, no.4, pp.1-7, 2010.

[22] R. Smith, "Journal accused of manipulating impact factor", British Medical Journal, vol.314, no.7079, p.463, 1997.

[23] N. Sombatsompop, T. Markpin, and N. Premkamolnetr, "A modified method for calculating the impact factors of journals in ISI Journal Citation Reports: polymer science category in 1997-2001", Scientometrics, vol.60, no.2, pp.217–235, 2004.

[24] B.D. Thombs, A.W. Levis, I. Razykov, A. Syamchandra, A.F. Leentjens, J.L. Levenson, and M.A. Lumley, "Potentially coercive self-citation by peer reviewers: a cross-sectional study. Journal of Psychosomatic Research, vol.78, no.1, pp.1–6. 2015.

[25] Thomson Scientific. Journal self-citation in the Journal Citation Report Science Edition: a citation study from Thomson ISI. 2002.

[26] G. Yu and L Wang, "The self-cited rate of scientific journals and the manipulation of their impact factors", Scientometrics, vol.73, no.3, pp.321-330, 2007.

[27] M. Zitt, "Citing-side normalization of journal impact: a robust variant of the Audience Factor", Journal of Informetrics, vol.4, pp.392-406, 2010.

[28] M. Zitt and H. Small, "Modifying the journal impact factor by fractional citation weighting: the audience factor", Journal of the American Society for Information Science and Technology, vol.59, no.11, pp.1856–1860, 2008.

Development of the Virtual Physical Laboratory Work "Franck-Hertz Experiment"

Madina Ipalakova, Yevgeniya Daineko, Mariya Brodyagina, Marina Yunnikova, Zhiger Bolatov, Yermek Yessenov
Department of Computer Engineering and Telecommunications
International Information Technology University
Almaty, Kazakhstan
m.ipalakova@gmail.com

***Abstract* – In the article the development of the virtual physical laboratory using the example of one of the laboratory works is presented. The choice of the development platform is justified; the structure of the application and the user interface is described.**

***Index Terms* – Virtual physical laboratory, virtual experiments, physics, XNA Framework.**

I. Introduction

Currently, the modernization of education is carried out very actively in Kazakhstan. This process, apart from transition of the universities to academic freedom and autonomy, introduction of polylingual education, updating the educational programs, includes the wide use of information and communication technology (ICT) within secondary and higher education. This implies development of the global information network, application of distance learning technologies, providing education institutions with ICT tools and others. Thereby, the majority of the universities in Kazakhstan implement the named technologies, which significantly change the nature of acquisition, development and transferring of knowledge.

Study and use of ICT in educational process allow students to get the skills and qualifications, which are essential in the modern society. Besides, wide use of ICT makes education more accessible and provides the foundation for the Life Long Learning (LLL-paradigm). Moreover, ICT is an effective tool for the development of new teaching methods and techniques, which improve the quality of education.

One of such new learning tools, which is used more and more often, is virtual laboratories [1]. They allow conducting different types of experiments and simulations, if a real laboratory is unavailable for some reasons [2]. Physics is one of many other sciences, where such virtual simulations are possible.

In the International Information Technology University (Almaty, Kazakhstan) at the Department of Computer Engineering and Telecommunications the authors develop a Virtual Physical Laboratory (VPL), which is designed to help the students to study physics by providing the experimental support for the theoretical learning [3]. At the moment, the Laboratory consists of six laboratory works from various sections of physics, according to the sample program for physics from the standard of higher education for the technical universities in Kazakhstan. The distinguishing feature of the laboratory is that the laboratory works are presented in three languages, namely Kazakh, Russian and English.

In this article the one of the virtual laboratory works, namely "Franck-Hertz Experiment", is presented.

II. Physical Description of the Experiment

The aim of the virtual laboratory work "Franck-Hertz Experiment" is to study the excitation of the inert gas atoms by electrons, to determine the excitation energy of inert gas atoms, the wavelength and the mass of the emitted photon.

Franck-Hertz Experiments, the aim of which was to measure the ionization potentials of the atoms, confirmed experimentally the Bohr postulates. For this work in 1925, they were awarded the physics Nobel Prize for establishing laws in collisions of electrons with atoms.

The experiments of Franck and Hertz were as follows. Through the test gas the electrons, accelerated by an electric field, were passed. At the collision with the atoms of the gas the electrons could move into new excited states with a definite value of the energy, more than the ground-state energy. Moreover, if the energy levels of the atom are discrete, the kinetic energy of the electrons should not be less than a certain minimum value that can excite the gas atoms.

In the experiments a tube, and three electrodes: cathode, grid and anode, were used (Fig. 1). The tube was filled with mercury vapor at a pressure of $p \approx 1$ mm Hg. Art.

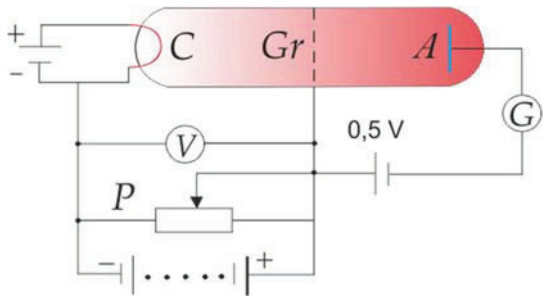


Fig. 1. Schematic diagram of measuring the current voltage characteristics of a gas-filled triode.

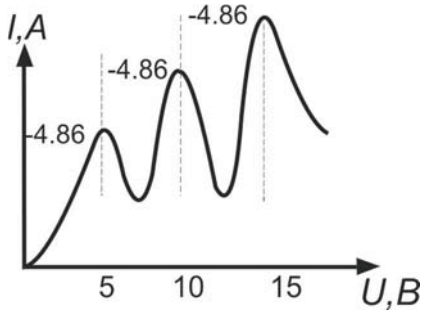


Fig. 2. The dependence of the current in the anode circuit on the voltage between the cathode and the grid.

The electrons are accelerated by the potential difference *U* between the cathode and the grid. This potential difference can be changed using the potentiometer *P*. Between the grid and the anode there is a retarding field 0.5 *V* (the method of retarding potential).

The dependence of the current through the galvanometer *G* on the potential difference between the cathode and the grid *U* was measured. During the experiment the dependence, shown in the Fig. 2, was obtained. Here, $U = 4.86$ *V* corresponds to the first excitation potential.

According to the Bohr's theory, each of the mercury atoms can get a well-defined energy, turning into one of the excited states. Therefore, if in the atoms the stationary states really exist, the electrons, colliding with mercury atoms, should lose energy discretely, by defined portions equal to the difference between the energy of the corresponding stationary states of the atom.

Thus, the experiment has shown that the electrons transfer their energy to the atoms of mercury in portions, and 4.86 *eV* is the smallest possible portion, of which may be absorbed by the mercury atoms in the ground state. Therefore, Bohr idea of the existence of atoms in stationary states brilliantly stood the test of experiment.

III. Development of the Virtual Laboratory Work

A. Technological Basis

The choice of the development platform was based on the review of the existing game engines, such as XNA Framework, Unity 3D Engine and Unreal Development Kit (UDK). The main criteria in selecting a game platform to create the virtual laboratory work were the support of 3D and the speed of project development. After the analysis the choice was made in favor of the XNA Framework.

XNA Framework is a set of tools with a managed environment (.NET), which is designed for rapid development and simplified control of computer games for Windows, Windows Phone and Xbox 360. XNA offers a set of tools that is designed to avoid a repetitive pattern code «Code Reuse» and join the different aspects of game development in the same environment. A special feature of XNA is the ability to easily convert code between Windows and Xbox 360 applications.

XNA Framework is divided into three main sections: graphics engine, model of game applications, and conveyor of the content. In order to use all the parts, it is necessary to connect the libraries of the Microsoft.XNA.Framework package written on C#. XNA Framework provides good support for 3D and many different types of images, as well as small memory consumption for the image based games development. Great performance during the work with the elements is caused by a built-in GPU- acceleration. The application development is carried out with C# language, using Visual C# Express development environment. The advantages of C# programming language are:

- speed of project development;
- cross-platform;
- code performance and resource requirements;
- support of libraries;
- convenience of debugging and assembly.

Another important factor in choosing the XNA Framework is the full support for classes and object-oriented programming, and full access to the .NET base classes' library.

B. The Structure and the Implementation of the Laboratory Work

From the software point of view a virtual laboratory work is a project collected using C# language, which is run firstly. The project consists of a set of parts, but generally they can be divided into two groups. They are game resources and program code. The game resources of the project are the graphical, musical or other resources that are necessary to make a scene realistic. Classes and procedures written in the C# language, which are responsible for interaction between the resources, belong to program resources. The term class refers to a certain program unit, where data, their structure and possible functions are strictly defined. Classes are combined in software libraries. In case of work with XNA Framework there is an opportunity to connect third-party libraries, and also to program and use the own ones.

Scenes make the basis of the project. A scene processes and includes all resources of the project. During development of the project in two-dimensional space 3D realism is achieved using art tools.

«Franck-Hertz Experiment» laboratory work includes six user classes. The structure and the relationships between the classes are shown in the Fig. 3.

The start of the application is carried out through the Program class. ScreenForm is the main class, which interacts with the Launcher class. The Launcher class connects the elements from the following classes: ModelGame, Camera and ModelFrankAndHertz. ModelFrankAndHertz class is responsible for the experiment itself. This class is used to launch the form and draw all the main elements. Camera and ModelGame classes are required for the creation of realistic environment in the experiment. The ModelGame class processes all graphics representation in the laboratory work using vector graphics and data arrays. This class contains all-important components of game, graphic device, parameters of a window and the dispatcher of contents.

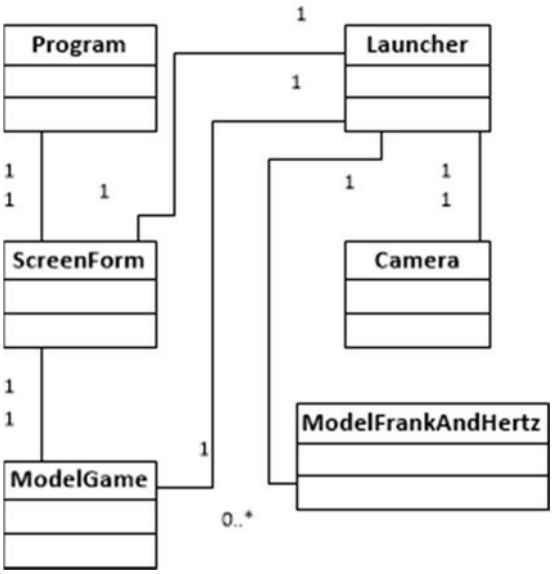


Fig. 3. Class diagram of «Franck-Hertz Experiment» laboratory work.

C. User Interaction

Execution of the virtual laboratory work represents implementation of a certain sequence of actions, which includes input of initial data, start of the experiment and registration of the results. «Franck-Hertz Experiment» virtual laboratory work contains five main steps for a successful completion. Information about the content of the work and the reference about implementation are attached in the form of methodology of the laboratory work. Figure 4 shows the sequence of steps of user interaction with the «Franck-Hertz Experiment» laboratory work.

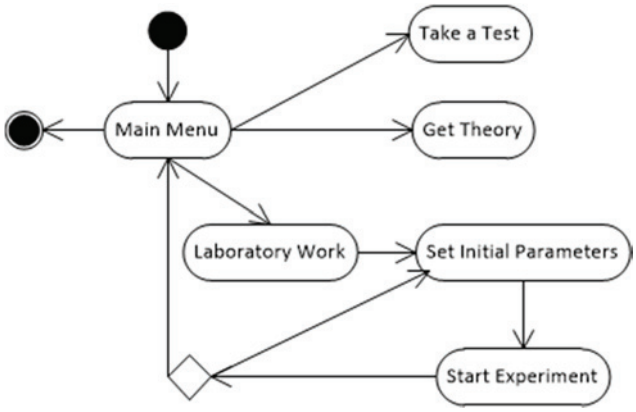


Fig. 4. Activity diagram of «Franck-Hertz Experiment».

After initiation of the virtual physical laboratory the main window is opened, in which user has access to the following sections: «Test», «Theory» and «Laboratory Work». Sections «Test» and «Theory» contain test tasks to consolidate the knowledge and methodology of the experiment. The section «Laboratory Work» allows starting implementation of the experiment.

The user interface is used to control functional capabilities of the virtual laboratory work. The full reference on how to work with the application is available for students. The home page interface of the «Franck-Hertz Experiment» laboratory work is presented in the Fig. 5.

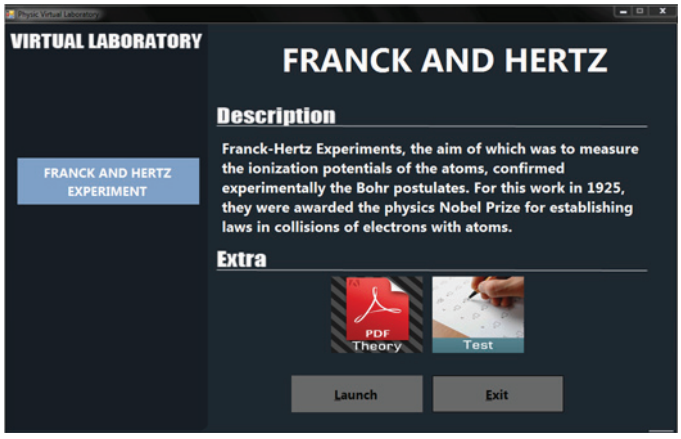


Fig. 5. Home page interface of «Franck-Hertz Experiment» laboratory work.

Figure 6 shows the interface of the running experiment. First, a user should select the type of the gas from the provided list: argon, helium, mercury, krypton, neon or xenon. Second, the user can configure the filament voltage (V) and enter data values of a marker. Finally, after all the necessary data are entered by the user, data table will be displayed with the following information: calculated peaks and minimum points. Also a graph will be shown. The sequence of steps can be repeated for the different values of V and marker, according to the given task.

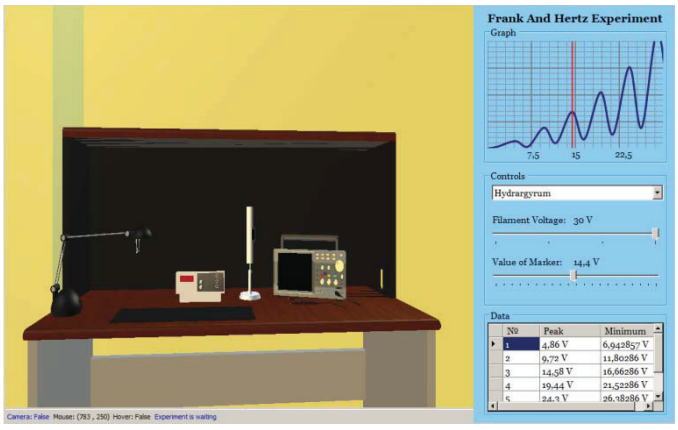


Fig. 6. Interface of the experiment of «Franck-Hertz Experiment» laboratory work.

IV. Conclusion

The development of ICT inevitably affects the way we teach and study. Virtual libraries are the good example of how ICT can facilitate and improve the learning process, if real laboratories are inaccessible. In the article the authors present the virtual laboratory work "Franck-Hertz Experiment", which is a part of the Virtual Physical Laboratory. The Laboratory is

introduced into the process of teaching physics in the International Information Technology University (Almaty, Kazakhstan). As the future work new laboratory works are planned to be developed and integrated into the VPL.

ACKNOWLEDGMENT

The project was accomplished with the financial support of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan according to the Program of grant funding of scientific research in 2015-2017, 2622/ГФ4, reg.№ 20150115РК01145.

REFERENCES

[1] E.A. Mohamed and S. Hasegawa, "An instructional design model and criteria for designing and developing online virtual labs," International Journal of Digital Information and Wireless Communications vol. 4(3), pp. 355-371, 2014.

[2] A.V. Trukhin, "Ob ispolzovanii virtualnyh laboratoriy v obrazovanii," Otkrytoe i Distantcionnoe Obrazovanie, vol. 4(8), 2002 (in Russian).

[3] Y. Daineko and V. Dmitriyev, "Software module "Virtual Physics Laboratory" in higher education," IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), 15-17 Oct. 2014, pp. 1-3.

Semi-automatic Annotation Tool for Sign Languages

Kairat Aitpayev, Shynggys Islam, Alfarabi Imashev
National Laboratory Astana
53 Kabanbay batyr ave., Astana, Kazakhstan
E-mail: {kairat.aitpayev, sislam, alfarabi.imashev}@nu.edu.kz

***Abstract*—The goal of this work is to automatically annotate manual and some non-manual features of sign language in video. To achieve this we examine two techniques one using depth camera Microsoft Kinect 2.0 and second using simple RGB mono camera. In this work, we describe strength and weaknesses of both approaches. Finally, we propose the semi-automatic web-based annotation tool based on second technique, which uses hand and face movement detection algorithms. Furthermore, proposed algorithm could be used not only for annotating clean training data, but also for automatic sign language recognition, as it is works in real time and quite robust to variability in intensity and background. Results are presented in our corpus[1] with free access.**

***Index Terms* — Sign language recognition, annotation tool, hand detection, hand tracking.**

I. Introduction

The objective of this research is to reduce time for annotation videos containing Sign Language (SL).

A sign consists of two types of features: Manual features contains gestures made with the hands, Non-manual features such as facial expressions or body posture, which can both effect to meaning of a manual sign. According to studies of both Liddel, Johnson [1] and Skokoe [2], manual features can be separated into 5 main visemes[2]: shape, placement, movement, orientation and arrangement. Non-manual features have more visemes like eyebrow position, lip shape, body shift, twisting shoulders, etc. According to this, annotation process needs much time and human resources. Whereas most of current research [3], [4], [5], [6], [7] in automatic SL recognition uses only manual features, our approach mainly focus on automatic annotation of them and in addition on some non-manual features like head movement. Therefore, our work mostly related with hand detection, tracking and classification.

Most of the state-of-the-art approaches detect the hands in sequence of frames and then classify the path that they take. This causes difficulties as the hands move quickly (and are often subject to motion blur), have high degrees of freedom (and therefore vary in appearance) and often occlude each other. In addition, tracking often employs skin color, which means that the face and hands can be easily confused and the clothing worn by the signer must be of a contrasting color and have sleeves. All of these issues effect on the success of tracking approaches.

The proposed hand detection and tracking approach can handle with specified limitations. Finally, we have developed web-based annotation tool, which embeds all developed algorithms. The full annotation process consist of manual annotation of frame ranges in video for each sign and succeeding automatic annotation of features.

II. Related Work

There are various approaches for hand detection in SL. Most accurate results could be achieved using data gloves to acquire 3D position and trajectory of the hands [8]. In some vision approaches used gloves were single colored, one for each hand, which makes it easy to detect and segment hands when they occlude each other [4], [9]. A more natural approach is to use a skin color model [11]. Depth used to simplify the issue of background segmentation and increase accuracy of hand shape classification. Old approaches use stereo camera pair [12], modern ones use depth cameras [13].

ELAN[3] is an annotation tool, which is commonly used for manual SL annotation. There are additional interface to it that can handle automatically generated annotations by sign language recognition [15]. The state-of-the-art in automatic SL annotation and in general automatic SL recognition can be considered the work of Buehler et al. [14].

III. Methodology

In this section we in detail describe annotation process and each part of proposed algorithm.

A. Annotation process and web-interface

We have developed our own annotation tool, which uses web-interface to involve more people to process. Tool allow to signers record videos using their own web-cameras, evaluate records for quality and annotate them.

For data collection, we use two approaches. First then signers record videos in their homes using their own computers and web-cameras. This allows us collecting big dataset with different conditions (quality of video, background, intensity). Second, we record so called clear videos in our studio with almost ideal conditions (black background and signers wear dark clothing with sleeves), see examples in Fig 1. For annotation process videos only with good qualities are allowed. Those are all videos recorded in our studio and videos with good quality according to the first approach.

[1] Kazakh sign language corpus - https://kslc.kz
[2] visemes much the same as phonemes in speech
[3] https://tla.mpi.nl/tools/tla-tools/elan/

Proposed annotation method is semi-automatic. Key frames, in which a sign starts and stops, are done manually. In parallel our algorithm annotating manual features: *a)* bounding boxes of hands for future shape classification; *b)* placement; *c)* movement. We are not annotating arrangement, because it could be easily calculated from placement and orientation, as it is task of hand classification.

Because of variety in recoding conditions, we propose two robust algorithms for automatic annotation of manual features.

B. Using Depth Cameras

As a depth camera, we used Microsoft Kinect 2.0 (Kinect). According to our previous experience [16] it can calculate depth pretty well. New version of Kinect has better accuracy, resolution and can recognize new features like facial expression and hand status (open or closed).

We assumed that joints of skeleton, which produced by Kinect could be used as points of hands, head and body position. Our preliminary experiment showed that hands joints are missing their positions when hands occlude each other, see Fig. 2. Body joints have the same problem as signs usually occurs in from of chest. However, body position can be calculated from shoulders joints, as body joint is exactly in the middle of them and hands usually do not overlap shoulders. Head joint as facial expression recognition have quite stable accuracy and used directly for automatic annotation.

For our final experiment, we used skin segmentation and HOG classification (algorithm described in details in next section) applied to color map produced from Kinect to detect hand. After hands were detected, we used coordinate mapping[4] to find average depth for each hand segmented according to a depth map.

We still have a problem with frame synchronization in video. As we are using almost all feature recognition of Kinect, it needs very high computational resources. At the same time in parallel, we are trying to record the video, because of overloaded system program sometimes skipping frames, so that finally we could have less frames in second than needed.

C. Using Mono Camera

The main problem of building semi-automatic annotation tool is to find hands at every frame of a video. The detection of hands is implemented using the method that combines skin color and motion detection algorithms [18]. In our algorithm, skin color elements are classified with the rules, which describe the skin cluster in the RGB color space [19]. This method illustrated better results compared to the skin probability

Fig. 1. Examples of recorded videos, a) videos recorded in studio with ideal conditions, b) in homes using standard webcam.

[4] Embedded function of Kinect SDK 2.0 to synchronized depth, color or body maps to with other.

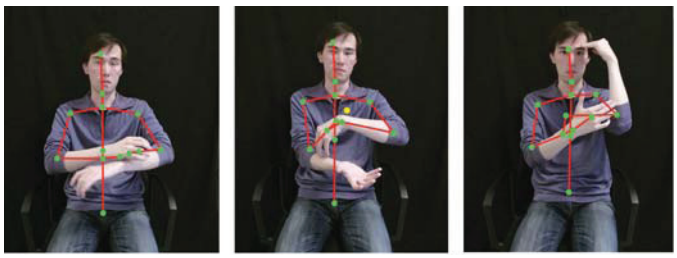


Fig. 2. Body points of Kinect with dismissed positions of joints while showing signs.

algorithm in [18]. In order to distinguish hands from face the motion detection algorithm is applied to images where pixels of skin color were labeled. The motion is detected by taking the difference of two consequent frames. Then, we cover hand candidate areas with the rectangles of fixed size. Rectangles can intersect, but none of them includes the center of another rectangle [18]. To prepare data for hand detection, we need preprocessing step that segment pixels of interest from background. The segmentation algorithm is done using graph cuts [20]. In this algorithm image pixels and relationship between the pixels will be considered as a graph nodes and edges, respectively. A minimal cost cut on the graph produces an effective segmentation [20]. After applying graph cut, we extract features of every rectangle window represented by HOG descriptor.

Implementing all the above-described steps for every frame of video is time consuming. Therefore, after the first detection of hands, we find interesting points and track them. Since, borders of hands contain many points of interest, we extend the hand areas using the OpenCV dilate function with kernel of 10. To find points of interest we use dense sample algorithm proposed in [21]. Also, we track points by optical flow algorithm implemented in OpenCV library. The main problem of tracking is lost points. In other words, some points remain outside of the hand region. To avoid this problem, every 10 frames, we check points. As soon as we find a small part of points that are static, we remove them from the trajectory list.

During the video time the configuration of hands changes multiple times. For this reason, at every new configuration we need to reinitialize points of interest. We propose an algorithm based on the stops of the interest points and variations of trajectory. If the number of points that stop is greater than the 70% of all points, then we reinitialize points in that frame. Moreover, between every frame we plot a graph of variances of trajectories and choose local peaks as frames of interest.

Finally, on that interesting frames we draw rectangles on detected hands, to identify the shape component of sign, see Fig. 3. The size of rectangles are equal to the size of head that is computed using detector that constructs a classifier by selecting a small number of important features using AdaBoost [22]. To identify the movement component of sign, we calculate the average speed of hand on the whole video. Then, we compare the speed between two frames of interest with this average speed and identify its type.

To find placement component of sign, we divide the image into grids and label them according to the position of the head [10] Therefore looking at the position of hands we identify its type of placement, Fig 4.

IV. Experiments

In this section, we evaluate the performance of our semi-automatic annotation tool and hand detection algorithm. We also determine the influence of different parameters to HOG descriptor.

A. Datasets

There are 8 fingerspelling or gesture datasets we have faced in the very beginning such as Jochen Triesch Static Hand Posture Database, Jochen Triesch Static Hand Posture Database II, Sebastien Marcel Static Hand Posture Database [23], Thomas Moeslund's Gesture Recognition Database [24] and Polish Sign Language Dataset [25] and Massey University's HumanHand [26]. In our experiment, we have used Polish sign language fingerspelling dataset, from which we took 220 positive and 1633 negative images of size 128*128 pixels.

For evaluation of semi-automatic annotation algorithm, we recorded 50 videos in ideal conditions (on dark homogeneous background).

B. Experimental Setup.

After several experiments with Polish sign language dataset, we have found out that the shadows of hands on the train data images had huge impact on HOG classifier perception. Therefore, we eliminated shadows manually in the whole dataset, which improved hand detection. Then, we tried the descriptor on these cleaned data with 9, 18 orientation bins, a cell size of 4, 8, 16, 32 pixels and block size of 1, 3, 5, 7 cells.

To compare results of semi-automatic annotation tool we have manually annotated 50 videos. On each video, we drew

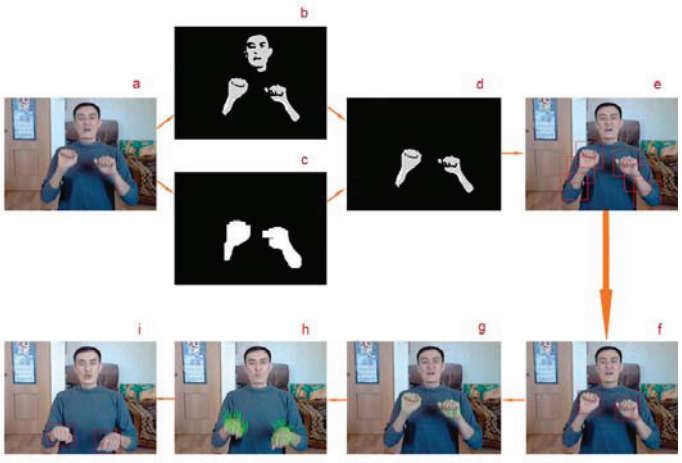


Fig. 3. Pipeline of hands tracking. a) original image, b) skin segmentation map, c) motion map, d) intersection of skin and motion maps, e) covered hand candidate areas with the rectangles of fixed size, f) founded hands using graph cut and HOG classification, g) getting points of interest, h) tracking the points of interest, i) founded hands using tracking algorithm.

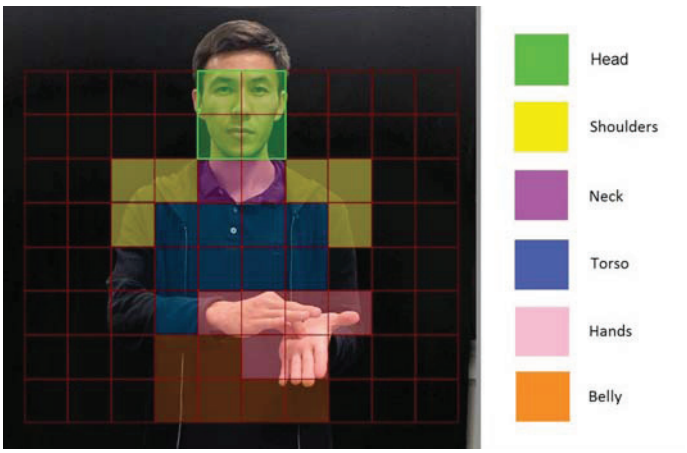


Fig. 4. Placement of hand on the body [10].

rectangles on hands to mark the shape, calculated the movement velocity and denoted the placement. Then, we run our algorithm on the same videos, which automatically determined these sign components. The accuracy of our algorithm for different sign components is evaluated as follows:

- Shape. We calculate the percentile of intersected area of manual and automatic annotation.
- Movement. We describe movement with three types of velocity such as fast, normal, and slow. Then, we take the percentage of matches as accuracy.
- Placement. For each frame of interest, we compare positions of hands determined by two types of annotation.

B. Results.

For hand detection, we got our best results of 93.2% with the following HOG descriptor parameters: 9 orientation bins, a cell size of 8 pixels, and block size of 1 cells.

Results of placement, movement, and orientation of hand for different types of camera presented in Table 1. Overall, the results for Mono Camera videos outperform the results for Kinect videos in all three components of signs.

V. Conclusions And Future Work

Usage of Kinect may look attractive at first time, but taking to consideration that joints are missing their positions when hands are moving in front of chest and head, annotation will be not accurate, as needed, which we could see from the results. In addition, usage of Kinect imposes limitations to recording process, as all videos should be recorded only in studio.

TABLE I. Comparison of Shape, Movement, and Placement on Videos from Kinect, Mono Camera.

visemes of sign	Types of camera	
	Kinect	***Mono***
Shape	83.2%	93.2%
Movement	76.5%	84.1%
Placement	91.3%	92.3%

On the other hand, second approach showed promising results and does not have limitations as previous one. This algorithm could be applied for already recorded videos from other datasets and integrated to existing annotation tools.

In future, we are planning to make our algorithm more robust to make it possible for wider application and add facial feature annotation by integrating algorithm of Zhu et al. [17].

ACKNOWLEDGMENT

This work was supported by the Ministry of Education of Science of the Republic of Kazakhstan under the research program number 349//049-2016.

REFERENCES

[1] S. K. Liddell and R. E. Johnson, "American sign language: The phonological base," Sign language studies 64.1, 1989, pp. 195-277.

[2] W.C. Stokoe, “Sign language structure: An outline of the visual communication systems of the american deaf,” Studies in Linguistics: Occasional Papers 8, 1960, pp. 3 – 37.

[3] P. Dreuw, T. Deselaers, D. Rybach, D. Keysers, and H. Ney, “Tracking using dynamic programming for appearance-based sign language recognition,” In Face and Gesture, 2006.

[4] T. Kadir, R. Bowden, E. J. Ong, and A. Zisserman, “Minimal training, large lexicon, unconstrained sign language recognition,” In Proc. BMVC., 2004.

[5] T. Starner, J. Weaver, and A. Pentland, “Real-time American sign language recognition using desk- and wearable computer-based video,” IEEE PAMI, 1998.

[6] C. Vogler and D. Metaxas, “ASL recognition based on a coupling between HMMs and 3D motion analysis,” In Proc. ICCV, 1998.

[7] C. Vogler and D. Metaxas, “Handshapes and movements: Multiple-channel American sign language recognition,” In Gesture-based Communication in HCI, 2003.

[8] F., Gaolin, W. Gao, and J. Ma, "Signer-independent sign language recognition based on SOFM/HMM," Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on. IEEE, 2001.

[9] T. Starner, A. Pentland, “Real-time american sign language recognition from video using hidden markov models,” Motion-Based Recognition, Springer Netherlands, 1997, pp. 227-243.

[10] H. Cooper, R. Bowde, “Large lexicon detection of sign language,” International Workshop on Human-Computer Interaction, Springer Berlin Heidelberg, 2007, pp. 88-97.

[11] G. Awad, J. Han, A. Sutherland, “A unified system for segmentation and tracking of face and hands in sign language recognition,” 18th International Conference on Pattern Recognition (ICPR'06), IEEE, 2006, pp. 239-242.

[12] S. Hong, N.A. Setiawan, C. Lee, “Real-time vision based gesture recognition for human-robot interaction,” International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, 2007 pp. 493-500.

[13] H. Cooper, E.J. Ong, N. Pugeault and R. Bowden, “Sign language recognition using sub-units,” Journal of Machine Learning Research, 2012, pp. 2205-2231.

[14] P. Buehler, A. Zisserman, M. Everingham, “Learning sign language by watching TV (using weakly aligned subtitles),” Computer Vision and Pattern Recognition, 2009 (CVPR 2009), IEEE Conference on. – IEEE, 2009, pp. 2961-2968.

[15] P. Dreuw, H. Ney, “Towards automatic sign language annotation for the elan tool,” Procs. of Int. Conf. LREC Wkshp: Representation and Processing of Sign Languages, Marrakech, Morocco, 2008.

[16] K. Aitpayev, J. Gaber, “Collision Avatar (CA): Adding collision objects for human body in augmented reality using Kinect,” Application of Information and Communication Technologies (AICT), 2012 6th International Conference on. – IEEE, 2012, pp. 1-4.

[17] X. Zhu, D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. – IEEE, 2012, pp. 2879-2886.

[18] J. Alon, Q. Yuan, “A unified framework for gesture recognition and spatiotemporal gesture segmentation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, pp. 1685-1699, September 2009.

[19] J. Kovac, P. Peer, F. Solina, “Human skin colour clustering for face detection,” EUROCON. Computer as a Tool. The IEEE Region 8, vol. 2, pp. 144-148, September 2003.

[20] Y. Boykov, M-P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images,” Internation Conference on Computer Vision, IEEE, vol. 1, pp. 105-112, July 2001.

[21] H. Wang, A. Kl¨aser, C. Schmid, L. Cheng-Lin. “Action recognition by dense trajectories,” CVPR 2011 - IEEE Conference on Computer Vision & Pattern Recognition, pp. 3169-3176, Jun 2011.

[22] P. Viola, M. Jones, “Robust real-time object detection,” Second International Workshop on Statistical and Computational Theories Of Vision Modelling, Learning, Computing, and Sampling, July 2001.

[23] S. Marcel, O. Bernier, J-E. Viallet, and D. Collobert, “Hand gesture recognition using Input/Ouput Hidden Markov Models,” In Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (AFGR), 2000. URL: http://www.idiap.ch/resource/gestures/.

[24] Website: “http://www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html”, visited – 21.07.2016.

[25] Website: “http://www.polsl.pl”, visited - 21.07.2016.

[26] Website: “http://www.massey.ac.nz/~albarcza/gesture_dataset 2012.html”, visited - 21.07.2016.

Education of Goniometric Functions with MATLAB software

Lucia Rumanová
Department of Mathematics
FNS CPU in Nitra
Nitra, Slovakia
lrumanova@ukf.sk

Viliam Ďuriš
Department of Mathematics
FNS CPU in Nitra
Nitra, Slovakia
vduris@ukf.sk

***Abstract*—Various research, experience and insights of teachers has shown that appropriation of information and communications technology (ICT) lead to improved learning of students and they are also suitable teaching methods. With the rapid advances in technology and the prevalence of various ICT in schools, schools must have thought-out work for good integrating ICT and visualization into the teaching of mathematics. In this article we show the possibility of using certain a mathematical system MATLAB in solving math problems.**

***Index Terms*—Education, graphs, simulation, examples, MATLAB.**

I. INTRODUCTION

Educational process is a system of three main components, which are in constant mutual interaction. These components are namely the teacher, the curriculum, and the learner. Nowadays it is a customary convention that the curriculum is prepared exclusively by the teacher who, using various teaching methods and organizational forms of instruction, exposes the curricular content to the learners. Afterwards, learners try to learn the curricular content and they are expected to be able to present their newly acquired knowledge in written and/or oral form. Finally, the teacher assesses the learners, especially to what extent they are able to reproduce the curriculum [1].

Education of math is a practical discipline and the mainly goal of this education is to promote better learning of mathematics by students [2].

National Program of Education is official document for all Slovak schools in which are included main goals and requirements of the math education. This program for Upper Secondary Education in Slovakia named ISCED 3A specifies standards for lessons of mathematics in Slovak schools. In this document is also states: (1) throughout studies to try include the problematic tasks for pupils, (2) during the study of mathematics contributes to the development of key competencies for solving of problems, it means to apply appropriate methods to solving these problems, using different and innovative practices for solving problems and to correctly formulate arguments and proofs pupils´ results [3].

In last several years is evidence that various appropriation of ICT help to improve the efficiency and effectiveness of education at different school levels, in formal or non-formal settings, too [4].

Using of ICT is essential in mathematics teaching and also learning. Teaching process of mathematics we can make more effective with using of ICT. These appropriations can enhance the capabilities of students and also better understanding the basic mathematical concepts [5].

In the paper we show the possibility of using certain a mathematical system MATLAB in solving math problems. We focused especially on teaching of goniometric functions with MATLAB, because for numerical modelling some properties of functions is this mathematical system very good tool and also for teaching them.

II. EDUCATION OF GONIOMETRIC FUNCTIONS

According to ISCED 3A the mathematics content is divided into five thematic areas, while goniometric functions include in area named *Relations, functions, tables, diagrams*. Students learn to dependence and functional relations of goniometric functions, specifically:

- Generally information of functions, specifically term of function, definition scope and range, graph of function and its properties.
- Goniometric functions.
- Relationship between the goniometric functions.
- Determination of goniometric functions.
- Properties of goniometric functions.

Pupils have knowledge of the thematic units as triangles, namely their types as right sided, isosceles or equilateral triangle and surface shapes (quadrilaterals - square, rectangle and trapezoid). After these units pupils learn according to ISCED 3A these facts about goniometric functions:

- Definition in the right sided triangle.
- Calculation in the different types of this triangle.
- Calculation of the various lengths in the right sided triangle.
- Calculation of angles in this triangle.
- Using goniometric functions in the application tasks.

- Using Pythagoras' theorem in solving various geometric problems.
- Drawing goniometric graphs.
- Properties of goniometric function.

Teacher can use frontal teaching or problem-oriented learning, maybe conversation with the pupils, different activities with suitable tools (PC, calculator, interactive whiteboard, software …), also used problem solving worksheet or collection task from recommended textbooks of mathematic.

Educators present that the main problems for the successful understanding of this thematic unit are:

- Pupils incorrectly work with tables in mathematics.
- Ignorance curriculum for given problematic (plane figures, Pythagoras theorem …).
- Pupils obviously know apply their knowledge in right sided triangle).
- Pupils do not work with calculators properly.
- Pupils have limited skills in drawing graph or solving geometric problems (many of these shortcomings come from primary schools).

So in the next part of article we present teaching selected examples of how pupils can present specific knowledge of the issues with using software MATLAB. We have selected a few specific tasks.

III. Teaching Examples with Using MATLAB

For working with numbers use the basic elementary mathematical functions, we provide an overview of goniometric functions in the table I.

TABLE I. Goniometric functions

Goniometric functions	
Function	**Meaning of function**
`sin`	Sine function
`sinh`	Hyperbolic sine function
`asin`	Inverse sine function
`asinh`	Inverse hyperbolic sine function
`cos`	Cosine function
`cosh`	Hyperbolic cosine function
`acos`	Inverse cosine function
`acosh`	Inverse hyperbolic cosine function
`tan`	Tangent function
`tanh`	Hyperbolic tangent function
`atan`	Inverse tangent function
`atanh`	Inverse hyperbolic tangent function
`sec`	Secant function
`sech`	Hyperbolic secant function
`asec`	Inverse secant function
`csc`	Cosecant function
`csch`	Hyperbolic cosecant function
`acsc`	Inverse cosecant function
`acsch`	Inverse hyperbolic cosecant function
`cot`	Cotangent function
`coth`	Hyperbolic cotangent function
`acot`	Inverse cotangent function
`acoth`	Inverse hyperbolic cotangent function

We deal with these functions as a mathematical function, which means that the function has determined the definition scope and range. For goniometric functions argument is entered and displays the result in radians. The arguments are always given in brackets. All functions and their descriptions can be obtained directly from *Prompt*, for example with using command `doc elfun`.

A. Example

Calculate: sin 60°, $e^{\ln 4}$, cos 45° - sin 45°.

```
>> sin (60/180*pi)
ans =
      0.8660
>> exp(log(4))
ans =
      4
>> z = 45/180*pi;
>> cos (z)-sin(z)
ans =
      1.1102e-016
>> tan (30/180*pi)/(tan(pi/4)+tan(pi/3))
ans =
      0.2113
```

Result of the difference cos 45° - sin 45° is the number 0, which was shown as a `1.1102e-016`. It´s number no less than accuracy `eps` in MATLAB and it reflects the accuracy with which the calculation was made.

B. Example

Create a graph of function y = sin (t), $t \in \langle 0, 2\pi \rangle$.

First we define in the *Prompt* a row vector axis of the independent variable. Then we define the second row vector for the dependent variable.

```
>> t = 0:0.1:2*pi;
>> y = sin(t);
```

With basic commands `plot` we are created a two-dimensional connected graph which causes an automatic opening of *Figure window* with graph. *Figure window* is automatically numbered, but this numbering can be changed with command figure (`<figure window>`. This command is before the command of graph.

```
>> plot(t, y);
```

In the command `plot` the parameters can change color of graph, the type of line or can place different symbols on the graph. In MATLAB is appropriate to set a range of graph axes, also enable grid of auxiliary lines, to create a headline and describe the graph axes.

```
>> axis([0 2*pi -1 1]); % Sets the axes
>> grid on %Grid view
>> title('Graph of function sine'); %Title
of graph
>>     xlabel('t',     'Fontsize',     16);
% Description of the x-axis
>>     ylabel('y',     'Fontsize',     16);
% Description of the y-axis
```

Very useful characteristic is the description of the graph by chains. One of the key possibilities is to use the command `text` (or `gtext`) and so the text placed on the certain position.

```
>>      text(2.3,      0.8,      'Function
{\ity}=sin({\itt})');
```

In the same way we add function $y = \cos t$ (red colour with dots). If we want to draw more curves into one graph so we use the command `hold on`. This command will ensure that the original curve (function) not lost.

```
>> hold on
>> y = cos(t);
>> plot(t, y, '-r.');
>>text(1.6, 0.1, 'Function{\ity}=cos({\itt
})');
```

The following command adds a legend yet.

```
>> legend ('sin', 'cos', 2);
```

The last parameter of the legend determines the position of its location, for example 2 mean that legend is arranged at top left in the picture.

The final graph is shown in Fig 1.

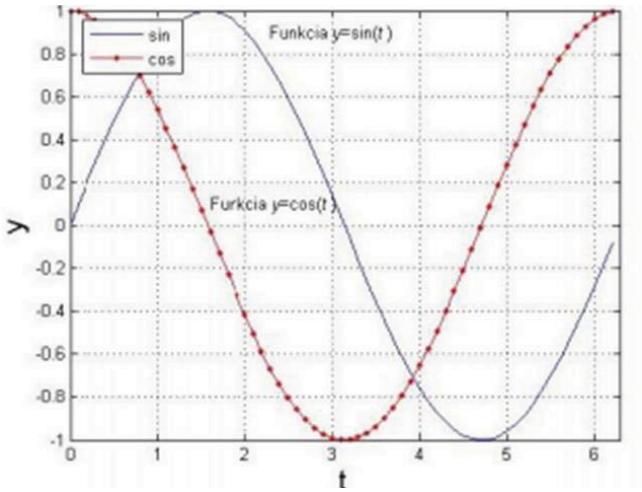


Fig. 1. Graphs of goniometric functions y = sin t and y = cos t

Now we construct a function that draws the time course of harmonic function sine with variable average value. Input parameters of the function are the number of periods and translation.

```
Function Graph_fun_sine(No_periods, Trans)
t = 0:2*No_periods*pi/100:2*No_periods*pi;
y = Trans + sin(t);
plot(t, y)
if Trans >= 0 %for better representation
 ymin = -1;
 ymax = Inf;
else
 ymin = -Inf;
 ymax = 1;
end
axis([0 2*No_periods*pi ymin ymax])
```

Call function from Command window while 4 is the number of periods and 0.5 is translation (Fig. 2).

```
>> Graph_function_sin(4, 0.5)
```

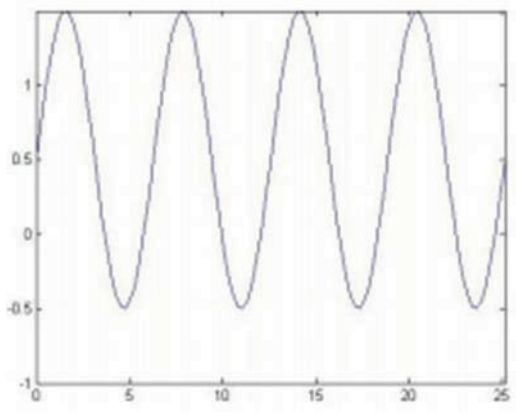


Fig. 2. Graphs of goniometric function y = sin t + 0.5, t $\in\langle 0, 8\pi\rangle$

IV. CONCLUSIONS

Various appropriation of ICT is occurring in almost of educational levels. In Slovak classrooms at different levels of education, there are in the context of teaching and learning of mathematics studies on the use of ICT.

From our experience we can say that use of ICT in education is good way as to teach or learn mathematics and use it in relation to other subjects in the school's environment. In the classroom the teachers themselves have to see what is appropriate for their pupils. Appropriations of ICT are interesting of technology tool for pupils´ work. So they have opportunity for active learning and development of knowledge or skills. In educational process the teachers and various educational methods create very important part. The educational outcomes depend on co-operation between student and teacher or student and student, also other appropriate methods or forms which are chosen by their teachers.

We focused on the implementation of software MATLAB to selected theme of mathematics. We are aware that the lessons of mathematics with using MATLAB can be made much more interesting or clearly for pupils and teachers too. Also can to present new approaches and so to develop better understanding of the needs mathematical concepts for pupils.

ACKNOWLEDGMENT

The paper is supported by research project no. VII/5/2015 named “Visualization of solution to various mathematical problems”.

REFERENCES

[1] I. Jakab, “Modernization of education using GIS”, in Staircase to even more interesting mathematics teaching. Nitra: Constantine the Philosopher University in Nitra, Slovakia, 2016, pp. 165–182.

[2] L. Rumanová, E. Smiešková, “Fostering creativity in geometry through the lesson problem with using communication in classroom”, in APLIMAT 2016: Proceedings from 15. International Scientific Conference from Applied Mathemathics, Bratislava: STU, 2016, pp. 970 – 977.

[3] The National Institute for Education, “The National Educational Programme – Mathematics ISCED 3A“, [online]. Retrieved from http://www.statpedu.sk/files/documents/svp/gymnazia/vzdelavacie_oblasti/matematika_isced3a.pdf

[4] V. L. Tinio, ICT in education: e-Primers for the information economy, society and policy. New York, 2003.

[5] G. Pavlovičová, L. Rumanová, V. Švecová, “Creativity in geometry on the playground“, in Staircase to even more interesting mathematics teaching. Nitra: Constantine the Philosopher University in Nitra, Slovakia, 2016, pp. 113–129.

Geometry Software Cabri 3D in Teaching Stereometry

Dusan Vallo
Department of Mathematics
Faculty of Natural Sciences, CPU in Nitra
Nitra, Slovakia
dvallo@ukf.sk

Julia Zahorska
Department of Mathematics
Faculty of Natural Sciences
Nitra, Slovakia
jzahorska@ukf.sk

***Abstract*: In this paper we focused on some advantages of software Cabri 3D based on idea of comparisons students incorrect results of the stereometric tasks and visualization via software environment. In the theory of mathematics education which is concerned with development of spatial and geometric imagination, some experts and researchers agree in opinion that these imaginations are for a long-time underdeveloped in high school students in Slovakia. Based on the theory of didactic constructivism we present a few research excerpts from brief probes and studies which inquiries are face to face with possibilities of using geometric software to eliminate the problems' results.**

***Index Terms:* geometry, teaching, dynamic geometry software, Cabri 3D**

I. Introduction

The National Curriculum for Upper Secondary Education in Slovakia (ISCED 3A) specifies the mathematical standards in Slovak high schools. One of the major topic in ISCED 3A is stereometry-the geometry of the 3-dimensional Euclidean space. The stereometric curriculum is focused to the teaching of solids and to solution of positional and metric tasks of linear figures. This inclination has its origin in historical context of the last century. The main role of education in geometry and mathematics was goal oriented to a development of geometric imagination and spatial imagination of the students as one of the most important ability for their future profession like engineers, architects, structural engineers, etc.

Over the years, it has been develops a detailed methodology for teaching geometry, which was built on the systematic pedagogical research native and foreign experts in didactic of mathematics and mathematics education. It turned out that the ability of students to manage stereometric curriculum are conditional upon the level of psychological predispositions, such as spatial imagination and geometric imagination. Their development can be greatly influenced by a work with models. Static models of solids were available traditionally in the past. Currently, courses in stereometry are supported by introducing ICT in schools, especially by using interactive geometric software e.g. Cabri 3D. This kind of 3D software supports the constructive thinking of students and plays an important role in the educational process. We illustrate some applications of 3D software in solutions of specific stereometric tasks in this paper.

II. Problem overview

A lot of Slovak and Czech researchers oriented to the theory of mathematics education diverge in exact definition of the spatial imagination. In agreement of the most of them, the spatial imagination can be characterized as an ability how to perceive space, to imagine geometric planar or 3-dimensional objects in different positions; to characterize shapes, their properties and relationships among their elements [1, 2, 3, 4, 5, 6].

Geometric imagination is considered as a higher form of the spatial imagination. It is stated that geometric imagination is the specific ability to work with geometric objects in human mind at situations, which are based on the own vision of figures and its imagination; to use the knowledge necessary to determine relationships; to deduce geometric shapes of real objects; to represent shapes graphically in different positions and finally, to solve geometric problems of real life in plane and space [7].

The definition is not exhaustive. The researchers are not in the definition of united, but it seems to agree in opinion that the both levels of imagination are for a long-time underdeveloped in high school students in Slovakia due to various reasons. This general conclusion is confirmed by several studies [6, 8, 9, 10, 11, 12, 13, 14].

It turns out that one of the main reasons of the undeveloped geometric and spatial imagination is formalism in geometry knowledge and mathematics education in generally. In fact formalized knowledge in mathematics well stated A. J. Chintchin [15].

It states that knowledge of the pupils are memorized only in formal and symbolic representations of mathematical knowledge as well as knowledge itself remains unknown, not associated with any imagination or projection of the figures [16].

In the theory of constructivism the learning is presented like knowledge construction based on the assumption that learners actively create and restructure knowledge in highly individual ways, through experiences. Specially, theory of didactic constructivism is based on ideas of constructivism which are implemented in math education. In this theory to diagnose of the formalism are used conventional diagnostic tools such as tests and exams. It is recommended to test the students' knowledge

via non-standard tasks which difficulty is at an appropriate level to the student's knowledge structure [17].

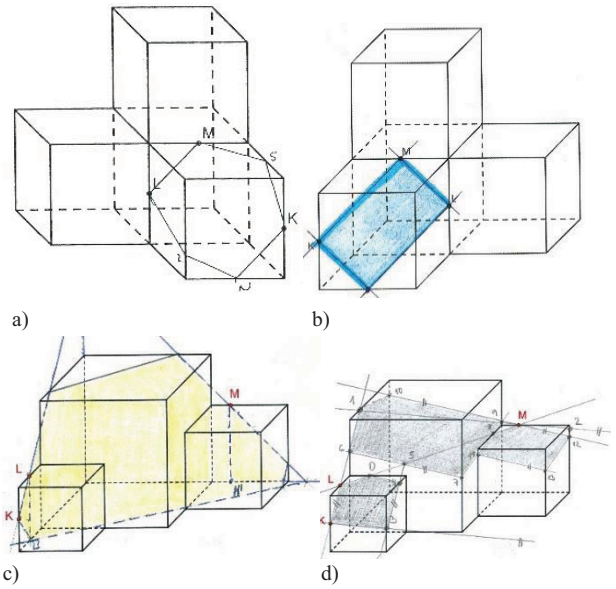


Fig. 1. A few samples of mistaken students'results of solutions of extraordinary stereometric tasks.

Based on the ideas of the didactic constructivism described above, in years 2013-2016 we have performed several research probes, in which we have demanded the high school graduates to solve extraordinary positional stereometric tasks [12].

From outcomes we select a few absolutely in-correct samples for illustration the statements about un-sufficient level of geometry knowledge some students. In the tasks we asked the students to cut the cube buildings by plane given with three points *K, L, M* (Fig. 1).

We remark that in generally 65% of all respondents did not solved the tasks correctly. In this count we cover the cases where the students had evident a poor geometrical thinking or low spatial imagination what they demonstrated in their results shown in Fig. 1a- d, Fig. 2a or Fig. 3a.

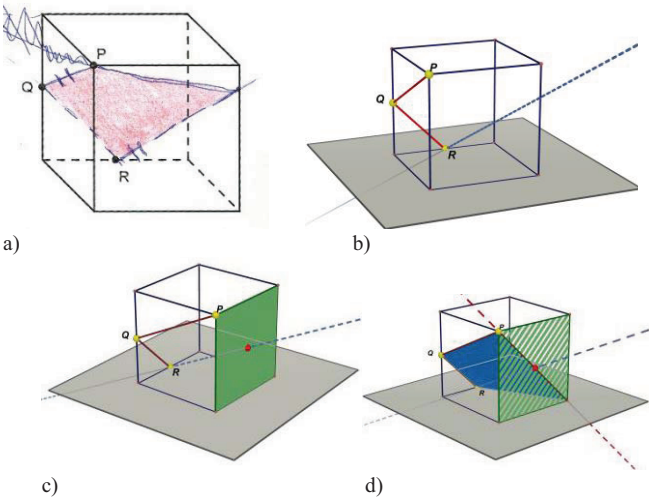


Fig. 2. Student's result with an inadmissible sketch of parallel lines passed through the points *P, R* and conter-example with correct construction in software Cabri 3D.

These respondents also have stated in running commentary that this kind of tasks were difficult and they have not considered the Geometry for their favorite subject, too.

III. Stereometry via dynamical geometry software Cabri 3D

Lack of adequate geometric ideas, projection of the stereometric solids or linear figures is a reason why the teacher should use some 3D dynamical geometry software in teaching of stereometric curriculum [18].

In this section we present some advantages of software Cabri 3D based on idea of comparisons students incorrect results of the task and visualization in Cabri 3D.

It is a general pedagogical observation that the students whose have a problem to solve stereometric tasks, used to approach to problem via algorithmic way or mostly by analogy. They do not use logic, but they aim their attention to some specific procedure or familiar words in an assignment of task. This phrase activates some method of construction.

In the theory of didactical constructivism is this approach called "*anti-signal*" because may mislead the student how to solve mathematical problem. The anti-signal is a suitable tool to diagnose the formalism in mathematical knowledge.

In Fig. 2 is demonstration of the procedure when the student used anti-signal "construct parallel line" through points *P, R*. Student obtained any intersection point the parallel line with the edge in the front face of the cube. In the usage of the software Cabri 3D this situation must not arise due to the skew lines. The student has a possibility to change a view or add the front face of the cube for the better visualization, too.

In Fig. 3a one can see that the student has no idea about the section of the cube building. There is any incorrect "copying of the edges" of the solids at the bottom, side and back planes. The software can be helpful in elimination of problem with graphic visualization.

In the cases when the student has knowledge and skills to complete the solution without any problems, can be the software Cabri 3D useful, too. The direct construction of the lines of the cutting plane, polygons, construction of the intersection points, etc. presents a strong motivational factor in the geometry teaching [20, 21, 22].

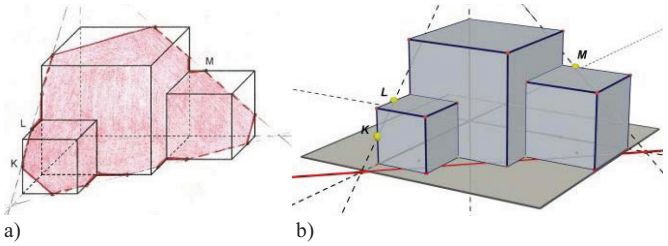


Fig. 3. Student's result with a wrong sketch of planar section which is copying the edges and conter-example with correct construction in software Cabri 3D.

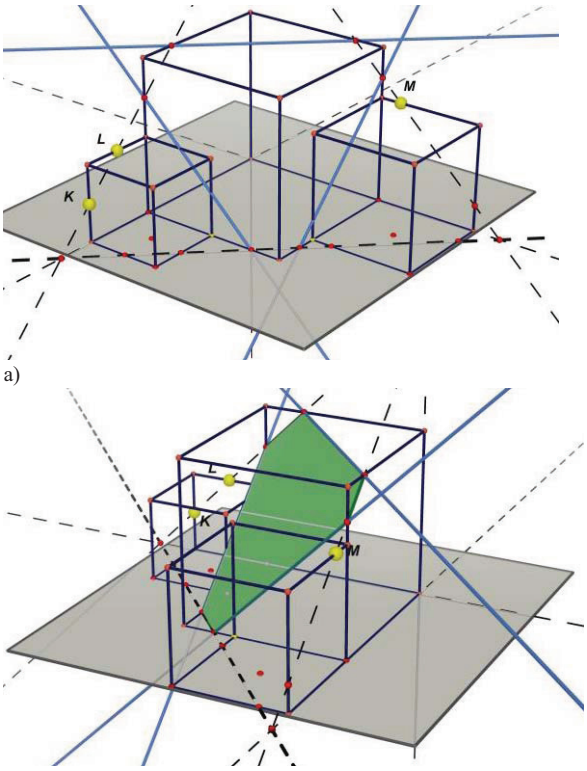


Fig. 4. Partial construction of planar section via software Cabri 3D. Visualization of the solution from another perspective is presented, too.

The visualization of the model in 3D space and the possibility to change the position of model in many directions can significantly increase of the level of the spatial imagination and help students to be successful in solutions of stereometric task.

By the usage of special tool „*Cut solid*", which has Cabri 3D implemented, one can solve the problem automatically. A teacher can integrate this tool in teaching in two ways.

First, when the student is familiar with the solution. It can be used like a feedback.

Second, it is also suitable as an instruction when the student has no idea how to construct the solution or how the result looks like in generally (Fig. 5).

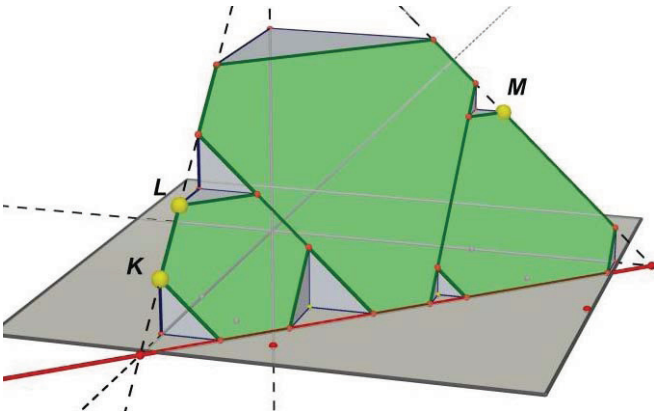


Fig. 5. Final visualization of the solution in Cabri 3D

IV. Discusion

In the traditional approach to teaching stereometry teachers use some representation of the 3D objects in the plane of the blackboard. It is requiring that students will solve stereometric task like planar geometric problem in assumption that they are able to imagine the situation in the space. It is important to remark that the significant graphical representation of the 3D objects in the plane need to imply from mental images about the solids [10, 18].

The probe realised in a continuance of a few years indicates that the majority of the graduates of the high schools have not sufficient level of the spatial imagination, the errors in results of the tasks are cardinal. Some pedagogical researches validate that the spatial imagination is developed in pupils 5 - 6 year old and in students 16 – 18 year old.

From this reason we believe that this mental ability of student can be increased using 3D interactive geometry software.

Of course, the progress of society call for new methods of teaching, newiest curriculum and it also enforces an implementation of the ICT in teaching in generally. Technology of visualization 3D space brings to students and their teachers great opportunity to individualize learning, to eliminate formalism in maths teaching, to present examples, to choose their own method and tools for solving stereometry problem.

The dynamical geometry software such Cabri 3D are softly implemented in mathematics teachig, any technical obstruction like hardware, licences, preparedness of the teachers are step-by-step overcome in Slovak schools and will be investigate like current didactic research problems.

Acknowledgment

This paper was supported by project KEGA No. 016UKF-4/2016 MINE Slovak Republic titled The Implementation of Constructivist Approaches to Mathematics Teaching with a Focus on Active Acquisition of Knowledge by Pupils in the Context of Bilingual Education

References

[1] A. Šarounová, "Development of geometric imagination at school" [Rozvíjení geometrické představivosti ve škole] (Czech), in Matematika a fyzika ve škole, Praha: JČMF, vol. 18 (3), 1988, pp.346 – 350

[2] M. Hejný, Theory of Teaching Mathematics 2 [Teória vyučovania matematiky 2] (Slovak), Bratislava: SPN Bratislava,1990, pp. 560

[3] O. Šedivý, "Illustration for teaching the geometry of space," [Názornosť vo vyučovaní geometrie priestoru] (Slovak) Názornosť vo vyučovaní matematiky," Nitra: FPV UKF v Nitre: Prírodovedec 214, 2006, pp. 3-14

[4] J. Molnár, Developing Spatial Conception (not only) in Stereometry [Rozvíjení prostorové představivosti (nejen) ve stereometrii] (Czech), Olomouc: Univerzita Palackého, 2009, pp. 183

[5] J. Perný, J. Hankova, T. Nováková and T. Votrubcova, "Future teachers and the development of geometric imagination of pupils," [Budoucí učitelé a rozvoj geometrické predstavivosti

žáku] (Czech), in. Acta Mathematica 13A Nitra: FPV UKF v Nitre: Prírodovedec 423, 2010, pp. 177-182

[6] V. Tomková, Technical Non-verbal Communication [Technická neverbálna komunikácia] (Slovak), Nitra: PF UKF v Nitre, 2013, pp. 223

[7] M. Hejný, J. Novotná and N. Stehliková, Twenty-five Chapters of Mathematics Education [Dvadset pet kapitol z didaktiky matematiky] (Czech), Praha: PF UK v Prahe, 2004, pp. 213-236

[8] A. Dubec, Methodology of Construction Geometry [Metodika deskriptívnej geometrie] (Slovak), Bratislava: SPN Bratislava, 1955, pp. 136 - 145

[9] F. Kuřina, The Art to See in Mathematics [Umění vidět v matematice] (Czech), Praha: SPN Praha, 1989, pp. 215-242

[10] K. Vidermanová, "Teaching Stereometry and development of spatial imagination through computer software," [Výučba stereometrie a rozvoj priestorovej predstavivosti pomocou počítačových programov] (Slovak) in. Informačné a komunikačné prostriedky vo vzdelávaní v matematike, Nitra: FPV UKF v Nitre, Prírodovedec 199, 2005, pp. 156-163

[11] M. Pomp and Z. Václavíková, "Factors with impact to geometric imagination of pupils," [Faktory ovplyvňující geometrickou predstavivost žáku] (Czech), in. Acta Mathematica 13A, Nitra: FPV UKF v Nitre, Prírodovedec 423, 2010, pp. 197-202

[12] D. Vallo, L. Rumanova and V. Duris, "Some Spatial Competences and Formalism in Solutions of Stereometrical Tasks," in. Procedia-Social and Behavioral Sciences 197, 2015, pp. 2320-2324.

[13] D. Vallo, V. Duris and J. Zahorska, "Specific method how to solve selected stereometry tasks in educational process," in. Procedia-Social and Behavioral Sciences 116, 2014, pp. 2957-2961.

[14] D. Vallo, L. Rumanova and V. Duris, "Spatial Imagination Development through Planar Section of Cube Buildings in Educational Process," in. Procedia-Social and Behavioral Sciences 191, 2015, pp. 2146-2151.

[15] A. A. Stoljar, Pedagogika matematiky [Педагогика математики] (Russian), Minsk: Vyšejšaja škola, 1974, pp. 414

[16] P. Květoň, Chapters in Mathematics Education II [Kapitoly z didaktiky matematiky II] (Czech), Ostrava, PF v Ostravě, 1986, pp. 13 - 16

[17] M. Hejný and F. Kuřina, Child, school and mathematics: constructivist approach to teaching, [Dítě, škola a matematika: konstruktivistické přístupy k vyučování] (Czech), Praha: Portal, 2001, pp. 240

[18] O. Chávez, R. Reys and D. Jones, "Spatial visualization: What happens when you turn it?" in. Mathematics Teaching in the Middle School,, Vol. 11, No. 4, 2005, pp. 190-196.

[19] K. Jones, "Learning Geometrical Concepts using Dynamic Geometry Software," in. Irwin (Ed), Mathematics Education Research: A catalyst for change. Auckland, New Zealand: University of Auckland, 2001, pp. 50 - 58

[20] N. Presmeg, Visualization in high school mathematics. For the Learning of Mathematics, 6(3), 1986, pp. 42-46

[21] J. M. Laborde, (Ed). Intelligent learning environments: the case of geometry, Berlin: Springer-Verlag, 1996

[22] J. M. Laborde, "Dynamic geometry as a window on mathematical learning:empirical research on the use of Cabri-Geometry", in. Procedings of 2nd mediterranean conference on mathematics education, Nicosia, 2000, pp. 161 – 173

Mathematical and Computer Modeling of the Process of Groundwater Pollution

Vladimir Krivykh, Saule Rakhmetullina, Zhanar Beldeubayeva
Department of Information Technology and Power Engineering
EKSTU after D.Serikbaev
Ust-Kamenogorsk, Kazakhstan
kanewardreinth@mail.ru; rakhmetullinas@mail.ru; zh.beldeubaeva@mail.ru

***Abstract*—This article is devoted to questions of mathematical modeling of the process of pollution of groundwater.**

The article describes the results of the work on the practical implementation of modeling algorithms. Work carried out by experts of the East Kazakhstan State Technical University. D.Serikbaev as part of a research project.

The main objective of the project is the development of information-analytical system of monitoring of the condition of groundwater contamination.

***Index Terms*—groundwater pollution, transfer of pollutants, modeling, mathemateical model, information-analytical system.**

I. Introduction

The specialists of the East Kazakhstan State Technical University D.Serikbaev (East Kazakhstan, Ust-Kamenogorsk) now, as part of a research project, work on the development of information-analytical system of groundwater contamination Monitoring [1]. Defining and testing of effective methods of mathematical modeling, which allow to transcribe contamination of groundwater authentically, have crucial meaning in the concept of development of IAS (information-analytical system) [2].

II. Description of the Complex of Mathematical Models

To achieve the target there was used such mathematical models as:

- Mathematical model of groundwater geofiltration described by partial derivatives;
- Model with a three-layer filtration region [3];
- Mathematical model of pollutant transport in groundwater in stationary mode [4];
- Mathematical model of groundwater contamination using the transport equation in porous media [5];
- Mathematical model describing the spread of chemical and radioactive contamination [6].

For practical implementation of the mathematical model of the spread of contamination in groundwater was used three-dimensional numerical method which is called MIF [7], created by RFNC-VNIITF. MIF methods is designed to solve the three-dimensional migration equations and filtering on arbitrary hexagonal grids. This technique has a set of conservative difference schemes, which allows to get monotonous solutions for various classes of problems.

In the article [8] on a row of model problems the MIF technique is compared with technique, which is widely used in the US to solve the three-dimensional migration equation. Soltuing the three-dimensional problems, results, which are obtained by different methods are close to each other, but the MIF technique gives more monotonous distribution of radioactivity in the area. Based on the MIF method has been created numerical method, which calculates contamination concentration spread of geofiltrational underground stream using the finite difference method, method of alternating directions and sweep method.

For the calculation of the mathematical model requires the following input data:

k – filtration coefficient
Qist – volumetric power sources within the body of water
Qst – volumetric power drains
ρ_{ist} – density of the solution in the sources
h- grid spacing
τ - the time step
η- elastic reservoir capacity factor
z – the average depth of the formation of the earth's surface
μ - water-retaining layer factor
λi - radioactive decay constant (a substance applies)
βi - dimensionless coefficient of interphase distribution
n- effective porosity
μ - water-retaining coefficient of layer
ρ_{wist} – density of water at the source
ρ – the density of pure water
δL - longitudinal coefficient of dispersion
δT - transverse coefficient of dispersion

Keeping the above values of parameters provided in the IAS database developed within the framework of a research project.

Fragment database with data necessary for the calculation of the mathematical model shown in Figure 1

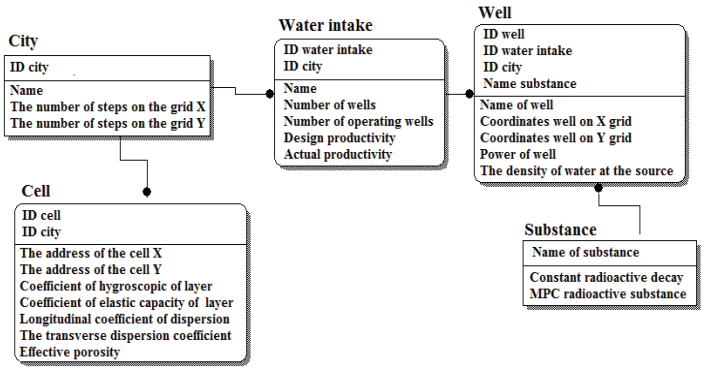


Fig.1. Information and logical database model

The table "City" contains basic information on the mentioned city. In this table it is determined by how many cells the city map will be divided.

The table "Cell" stores information about the formation and volumetric activity of radioactive substances in the rock.

The table "Water intake" keeps the information on water intake, its productivity and efficiency.

The table "Well" identifies the source, which will be used in calculation of spread of radioactive substances. This table also contains important information about the power of well and the density of water in the well.

The table "Radioactive substance" contains a constant of radioactive decay and the MPC of radioactive substances, which is necessary in comparisons with results.

In the process of modeling, grid with given dimension is superimposed on the local map (Ust-Kamenogorsk), where each grid cell has certain values for the required data. In calculations grid of 10 by 10 is superimposed on the map of city of Ust-Kamenogorsk.

The model is implemented using the finite difference method, method of variable directions. The system of linear equations is solved by the sweep method. [10]

For the test calculation there was used a mathematical model which describes the hydrodynamics geofiltration underground stream that includes equation of motion and the equation of continuity, also the uniqueness of conditions consisting of initial and boundary conditions.

Firstly, the calculation of hydrodynamic pressure is performed, neglecting the spatial variability of the properties of aqueous solutions.

Coefficient of filtration is calculated according to the formulas of linear interpolation. Interpolation is carried out along and by depth of profile section.

The next step is the calculation of distribution of pressures at the filtration flow.

The values obtained for the hydrodynamic pressure on the X and Y coordinates are used in the equation of filtrtion flow.

The next step is the calculation of filtration flow velocity.

The calculation is performed for a constant density flows with zero power drains.

The last step - calculating the spread of the radioactive components. The spread of contamination initially calculated tensor hydrodispersion. First, the transposed matrix is calculated and multiplied with the original. Thereafter, the resulting matrix is divided into module of velocity vector of the matrix. The cross ratio is multiplied by the unit vector.

After that, the obtained values of hydrodispersion are substituted into the formula for calculating the concentration of pollution.

Concentration measured in mg / l.

General modeling algorithm is presented in Figure 2.

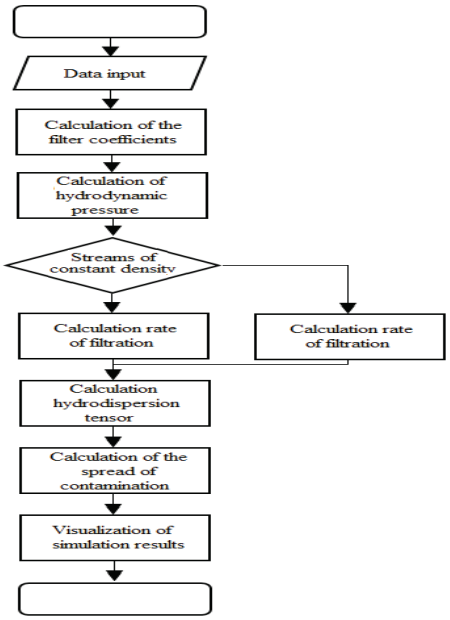


Fig.2. General algorithm of model

To test the algorithm calculations were performed for a point source of radioactive material located inside the plant.

In the calculations it was assumed that the concentration of the substance is equal to 20 mg / l, which is 200 times higher than normal.

The values are calculated as 100 meters, and only the last step value recorded in the table. The calculation is performed for a constant flux density with zero power of effluent.

Figure 3 shows the visualization of the model of contamination spread from industrial site UMP, which is superimposed on a map of the city of Ust-Kamenogorsk.

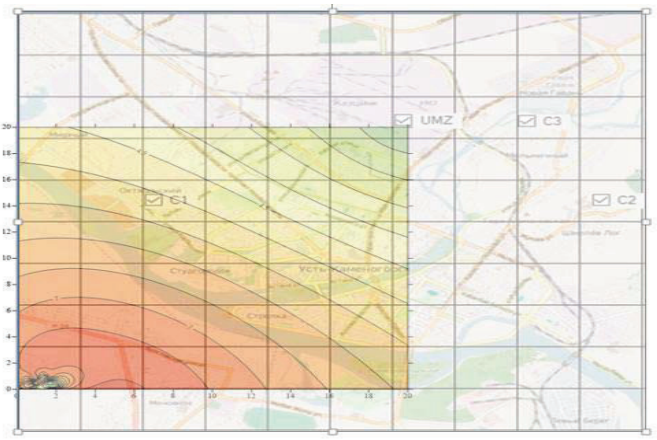


Fig.3. Visualization of uranium contamination from a point source

Also there were reviewed pollution modeling of wells, in a contact with radioactive substances in the form of uranium.

The calculation was made for three intakes in Ust-Kamenogorsk city.

Data for visualizing is represented by three columns with the values of X and Y for concentration, and Z for time. The time step is equal to days. On the scale on the right there can be seen the change in the spread of contamination within 10 days.

Approbation of numerical algorithms and algorithms of visualization the groundwater contamination process has been performed for a number of wells, located in the city of Ust-Kamenogorsk.

The following figure (Figure 4) you can see the visualization of the simulation results obtained on the basis of the well named October.

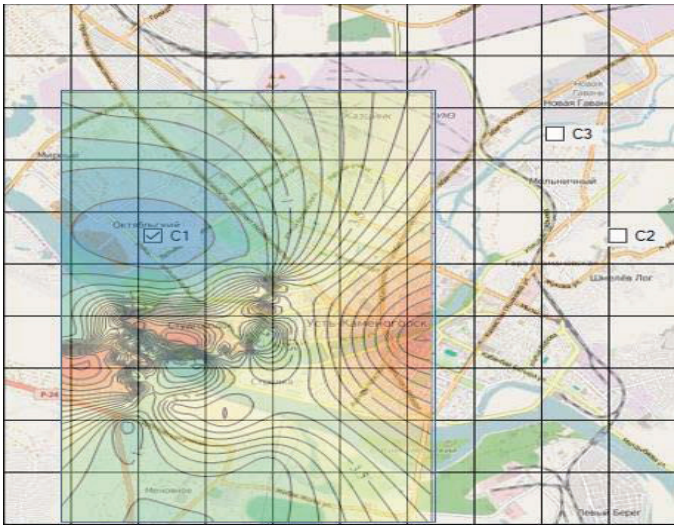


Fig.4 Imposition of the model of pollution from the well of the October.

III. RESULTS

The works were carried out in accordance with the schedule, compliance with the principles of the system approach in modeling and constructing mathematical models [11].

The main result is:

- database was designed and developed, which is necessary for the implementation of the mathematical modeling of process of groundwater contamination with radioactive substance;
- numerical scheme has been developed and was presented the algorithm for calculating the hydrodynamic pressure and radioactive contamination;
- was simulated a process of spread of radioactive substance from a point source;
- was modeled process of uranium spread by ingestion into water wells.

CONCLUSION

Mathematical modeling of groundwater pollution processes is being an important task in the development of information-analytical system of groundwater monitoring. It is important to note that the developed modeling algorithms will significantly enhance the analytical subsystem IAS, receive qualitative results of evaluation and prediction of pollution level of the groundwater.

REFERENCES

[1] Zh.Beldeubayeva S.Rakhmetullina, E.Turganbayev, V. Krivykh Information system of efficient data management of groundwater monitoring the Republic of Kazakhstan // Proceedings of the 9th International Conference on Application of Information and Communication Technologies AICT 2015 Russia, October, Rostov-on-Don, Russia, pp. 195-201, 2015

[2] Zh.Beldeubayeva S.Rakhmetullina, E.Turganbayev, V. Krivykh The concept of development of groundwater quality monitoring system // Proceedings of the conference «Actual problems of Computational and Applied Mathematics» 2015, October, Novosibirsk, Russia, pp. 112-117, 2015

[3] N.I. Panasiuk, A.M. Alferov, N.B. Starikov, I.A. Litvin, E.P. Lushnja, The results of a detailed simulation of influence of pile foundation in the hydrogeological conditions in the area of construction of the New Safe Confinement // Проблеми безпеки атомних електростанцій і Чорнобиля, vol. 16, pp. 124–129, 2011.

[4] M.U. Murzakatov, N, T, Azhygulova, Mathematical model of pollutant transport in groundwater under steady-state conditions // Vestnik of the OSU, Orenburg, Russia, vol., pp.119-122, 2005

[5] A.A. Sukhinov, A, I, Sukhinov Uniqueness of the solution of convection-diffusion-aggregation problem suspensions in an aqueous medium // Contemporary problems of continuum mechanics, Nauka, Moscow, Russia, p. 257, 2008

[6] P.I. Chalov, I.A. Vasiliev, V.V. Komosko, I.A. Kondakov, I.I. Pestov, V.N. Seleznev, V.A. Simonenko, L.V. Sokolov, O.I. Starodumov, V.V. Chernov, A.A. Shestakov Numerical simulation of radiation and other dangers of the Kara-Balta tailings hydrometallurgical plant for processing of uranium-bearing ores // «Academkniga», Moscow, Russia, p.115, 2009

[7] L.V. Sokolov, K.I. Simonova, A.A. Shestakov, O.S. Shirokovskaya Methodical research for the one-dimensional numerical calculations of the spread of contamination in porous media // Predpint, RFNC-VNIITF, Snezhinsk, Russia, vol. 92, 1996.

[8] O.V. Golubkina, L.V.Sokolov, G.A. Filimoncev, A.A. Shestakov Numerical comparison of techniques and MIF MT3D method for solving the three-dimensional migration equation //. Predpint, RFNC-VNIITF, Snezhinsk, Russia, vol. 118, 1997.

[9] C.Zheng, S.S. Papadopulos A Modular Three-Dimensoional Transport Model for Simulation of Advection, Dispersion and Chemical Reaction of Contaminants in Groundwater Systems, Inc. Rockville, Maryland, 1990.

[10] V.S. Krivukh, S.Zh. Rakhmetullina, Computer modeling of groundwater pollution, // Proceedings of the II International scientific-technical conference of students, graduate students and young scientists «Creativity of youth for the innovative development of Kazakhstan», EKSTU, Ust-Kamenogorsk, Kazakhstan, pp. 118-120, 2016

[11] D.N. Khorofas Systems and modeling // D. N. Khorofas; Translation from English under the editorship I.N. Kovalenko, Mir, Moscow, Russia, pp.418, 1967

Electronic Science In Azerbaijan: Current Status, Problems And Perspectives

Rasim M.Alguliyev[1], Rashid Q.Alakbarov[2], Tahmasib Kh.Fataliyev[3]
ANAS Institute of Information Technology
Baku, Azerbaijan
[1]director@iit.ab.az, [2]rashid@iit.ab.az, [3]depart3@iit.ab.az

Abstract*—The article is devoted to topical problems of e-science project carried out in the Republic of Azerbaijan. In the context of information society the essence of e-science, its goals and objectives are clarified, international experience and challenges, scientific-theoretical and practical problems are analyzed. Conceptual problems of e-science are considered, its modern state and perspective trends for the republic are *presented.

***Index Terms*—information society, e-science, AzScienceNet, internet services, cloud technologies.**

I. Introduction

The forming of Information Society (IS) in the Republic of Azerbaijan (RA) is considered as a priority target and its legal-normative basis, the coverage and problems it solves are continuously expanding and developing.

General concept of IS was adopted at World Summit on Information Society (WSIS) held in 2003 (Geneva) and 2005 (Tunisia), its main principles and goals were specified. Herein, the use of information-communication technologies (ICT) for improvement of people's living standards and reduction of digital divide was adopted as the main target and the motive force of development. Application and development of ICT in different fields of activity in the country during the past years have become an integral part of an everyday life in acceleration of the socio-economic growth and sustainable development, as well as increase of transparency and accountability.

It is known that in the C7 activity direction of WSIS Action Plan (AP) application of ICT in all spheres of life is considered and this covers seven diverse fields such as e-government, e-business, e-education, e-health, e-environment, e-agriculture and e-science. The principles proposed in C7 and State Programs carried out in the field of IS formation in Republic of Azerbaijan (RA) has set forth specific duties before appropriate bodies of Azerbaijan National Academy of Science (ANAS), IS ideas are being successfully implemented in science alongside the mentioned fields of activity (state governances, education, healthcare, etc.). The application of information technologies in scientific activity and scientific governance is realized within the framework of the e-science project as one of the priority fields of reforms carried out in ANAS [1]. The aim of the project is to provide the joint activity of scientific enterprises, organizations and staff, also individuals active in research that have information-communication infrastructure and access to information and computing resources via high-speed Internet network in virtual space.

The aim of the article is to investigate the topical problems of e-science project carried out in the RA.

II. Emergence And Development Stages Of E-Science

There are different international approaches, standards and state programs regarding e-science. There are multiple adopted projects titled as "e-science" in USA, England, Germany, Japan, India, Australia, CIS, and other world countries, and the research continues intensively. Those projects carrying out the implementation of application problems of ICT in different fields of scientific-research activity can be divided into three groups:

- Complex automation of scientific-research activities based on modern ICT (management systems of projects, grant, publications, etc.) (this is called CRIS-Current Research Information Systems in international scientific environment);
- Establishment of solid online scientific infrastructure for research (research e-infrastructure is achieved based on formation of solid online research environment with integration of CRIS of different scientific enterprises);
- Socio-economic benefit from technical innovation data for scientists and scientific organizations (this is interpreted as a social development of e-science).

"E-science" notion was first introduced by John Taylor, the general director of academic boards of United Kingdom in 1999 and includes new conduction methods of collective experimental research including computer modeling and the organization of virtual experimental environments [2]. "E-science" program announced in United Kingdom in 2000 has played a significant role in expansion of this novelty.

According to the professor of the University of Edinburgh Malcolm Atkinson, the goal of e-science is to create better research opportunities on all subjects by supporting the development of scientific collaboration with rich information resources obtained and analyzed as a result of leading distributed computations [3].

The research indicates that the history of "e-science" development has started a long time ago and is closely related to the emergence of network technologies. As an example, ARPANET project of Perspective Research Projects Agency of USA in the middle of 20th century can be shown, the main goal of which was to link the research enterprises among themselves [4]. ARPANET is the prototype of Internet and TCP/IP which is the current main protocol of data transmission in Internet was first used in that project. Moreover, the JANET (Joint Academic Network) including scientific and academic fields in Europe and grid infrastructure of CERN (Council for European Nuclear Research), GEANT (Gigabit European Advanced Network Technology) Association and CEENet (Central and Eastern European Networking Association) network associations have greatly contributed to the development of "e-science".

The establishment of e-science implies the forming of online research e-infrastructure at a national level. This, in turn, forms the solid global information space of information systems of separate scientific organizations. The idea of establishment of national e-infrastructure is carried out within the framework of state programs in several countries. Cyber infrastructure of USA, European research infrastructure, Japan scientific grid, national e-infrastructure projects of Great Britain, Australia, Canada, Russia and Moldova can be shown as an example in international environment.

E-science was established in the 80's of past century in Azerbaijan and the chronological order of some carried out works are as follows:

- In the 80's of past century "ACADEMSET" (The network of Computing Center of Scientific Organizations of Academy of Sciences of USSR), "RASUNT" (Republic Automatized Control System of Management for Science and Technology) and "ASOIAN: (Automatized System for Information Processing of Academy of Sciences of Azerbaijan) projects have been carried out;
- In 1991 the first access to Internet was realized in Azerbaijan;
- In 1995 www.ab.az (www.science.az) was the first website to be put into operation in Azerbaijan;
- In the middle of 90's with the support of Turkish Republic a network infrastructure covering scientific enterprises located in the ANAS campus was established and the access to TURKSAT satellite was provided;
- Network of the Academy of Sciences was provided with several work stations and equipment by British Petroleum enterprise;
- The network was connected to the Telecommunication satellite within the framework of "Virtual Silk Way" project of NATO in 2003.

The concept of national "e-science" is based on the solution of two fundamental problems such as re-establishment of existing scientific environment in accordance with IS requirements and the application of ICT in that environment [5]. For this purpose, consecutive solution of complex problems such as studying the world experience, monitoring of ICT application status in scientific activity, "e-science" management and the investigation of problems of information security provision, development and practical realization of scientific-theoretical principles are considered. Following results will be achieved when those problems are solved.

First, national e-science" program will be implemented on world standards level due to studying world experience, the regulation of informatization of science based on monitoring results and improvement of legal-normative base.

Second, communication-network infrastructure of e-science is formed. This is carried out based on the maintenance of material and technical base, establishment of local networks of scientific organizations and provision of high-speed Internet access, establishment of solid network linking the scientific enterprises of Republic of Azerbaijan, development of security strategy and realization of integration to international scientific networks.

Third, information resources with different appointments are created.

Fourth, with the application of ICT in scientific activity, the problems of organization of work place of researcher and problems covering the activities of scientific collectives, the establishment of new research relations based on e-environment on different fields of science, the organization of mutual relations with international scientific organizations, forming of scientific information spaces, establishment of computing environments based on supercomputer and grid technologies for solving problems requiring major computing and information resources, the problems of commercialization of science are solved.

Fifth, the education of scientific personnel in ICT field and staff training for organization of maintenance of technical-program tools are carried out.

III. E-Science And State Policy

The successful reforms carried out in e-government building under the leadership of the President of the country – National Strategy on ICT for the development of Republic of Azerbaijan (2003-2012 years), "Electron Azerbaijan" State Program, Action Plan on "Electronic government" formation in the Republic of Azerbaijan, National Strategy on scientific development in 2009-2015 years in Republic of Azerbaijan, "Azerbaijan 2020: future outlook" Development Strategy and National Strategy for 2014-2020 years on development of IS in Republic of Azerbaijan preserve their own place in formation of e-science.

The National Strategy on science development in 2009-2015 years in the Republic of Azerbaijan consists of 20 clauses covering problems such as the main goals and objectives standing in front of the science in Azerbaijan, specification of innovative and priority trends, modernization of scientific infrastructure, international scientific collaboration, highly qualified staff training, integration of science and education and others. The 19th clause of the Strategy and one subclause are devoted to information provision problems of the science and development of "e-science" model respectively. The strategy brings up the problems of formation of a new type of economy based on knowledge, modernization of management system in science field and scientific infrastructure and the training of highly qualified staff. One of the goals considered in the Strategy is the conduction of reforms in scientific activity, wide application of ICT, and development of e-science formation process. In the strategy reflecting the expansion of research in fundamental sciences important problems such as the improvement of management system in fields of science and technology, establishment of normative legal framework, provision of scientific information, strengthening of integration to international scientific space, the increase of efficiency of research and innovation policy, technological modernization of country are considered, which are the problems directly related to formation of e-science.

The development of IT infrastructure, emergence of e-science and knowledge economy which are among main requirements of the forming IS in Azerbaijan, have posed important duties in front of ANAS alongside with other relevant institutions. The provision of active participation of science in development of the society necessitates establishment of improved organizational structure of science itself and brings up the transformation of all traditional models in fields of science and education.

Establishment of IS, efficient use of opportunities enabled by IS for the development of citizens, society and state, comprehensive application of ICT in state governance, also development of IS as an economic sector stimulating socio-economic and cultural fields are considered in "Azerbaijan 2020: future outlook" Development Concept and National Strategy on IS development in Republic of Azerbaijan for 2014-2020 years. E-science holds a special place in priority directions of the Strategy such as development of "e-government" and national content. This includes increasing the efficiency of research and expansion of ICT in infrastructure development; the application of e-document circulation and provision of application of necessary systems (archiving, analyses, reports, etc.); for provision of electronic management of documents, the creation and development of internet resources (digital archives) on Azerbaijan history, patriotism topics, literature and cultural heritage; application of ICT in libraries, archives and museums, expansion of e-library network and etc. can be attributed herein.

It must be noted that, due to the lack of solid concept on formation of e-science the organizations located in different geographical locations of the republic and engaged in scientific activities encounter several technical, economic and other problems during the conduction of works in this direction. However, these problems are easily solved as a result of realization of "e-science" as a part of the forming e-government within the content of "Electronic Azerbaijan" State Program.

If we review the integral parts of e-science (Fig. 1) the area of coverage of scientific-theoretical and practical problems of its formation can be seen.

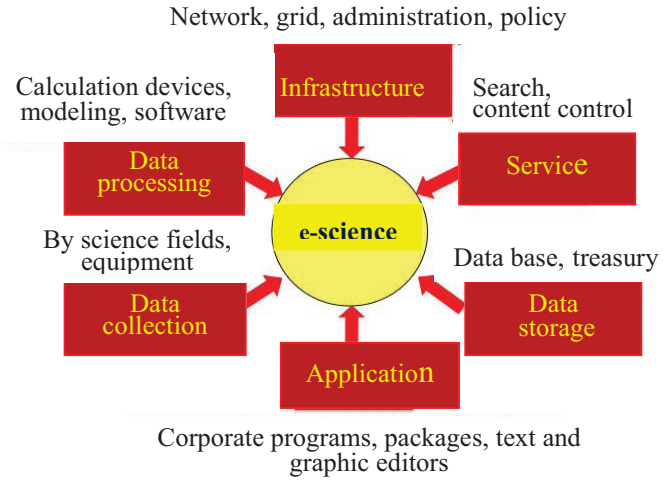


Fig. 1. Integral parts of e-science

The solution of problems such as formation of infrastructure, data collection, storage, processing, search, analysis, transmission, introduction and etc. determines the scientific-theoretical and practical problems of e-science.

IV. Status Of National E-Science Concept Formation

Researches and practical results are of major importance in e-science field in the Republic of Azerbaijan. E-science has a territorially distributed infrastructure and its duty is to cover the scientific institutions across the Republic entirely. It is known that the institutes and organizations of ANAS, higher education institutions and other academic institutions constituting the majority of scientific institutions are located in Baku city, Nakhchivan and Ganja Branches, Sheki and Lenkaran Regional Scientific Centers of ANAS are located in geographically distributed areas of the republic. E-science project carried out in the Republic has such a complex infrastructure and is formed based on AzScienceNet network infrastructure. Following can be mentioned as a result of analysis of the current status of implemented works within the framework of e-science.

A. The current status of e-science network infrastructure

Monitoring of application status of ICT in scientific activity must be primarily conducted for development of national e-science concept, priorities and trends of science

informatization must be specified based on obtained results. For this purpose, ANAS Institute of Information Technology (IIT) has conducted a monitoring in order to investigate the application status of ICT in approximately 150 scientific organizations active in the Republic of Azerbaijan including ANAS institutions in 2009 [6,7]. Second such monitoring was conducted exclusively for ANAS organizations in 2013.

- AzScienceNet which is a network platform of e-science covers all scientific enterprises of ANAS and scientists experience no problems while using the Internet services.
- Around 4300 computers are connected to AzScienceNet.
- Internet access speed of AzScienceNet is 410 Mbit/sec.
- The Data Center of AzScienceNet with a large memory and computing resources (300 terabyte, 15 Tflops) was created and its technical characteristics are continuously improving.
- This active network and computing e-infrastructure creates a high-speed connection among scientific institutions, provides multiple services to users, and at the same time, enables integration to international systems.
- Within the framework of AzCloud service, virtual computing machines are provided for solution of complex scientific problems requiring major computing resources in user-distributed environment.
- AzScienceStorage service provides memory resources for storage of information, security provision of which is considered important for ANAS institutions and organizations.
- Monitoring and Security center of AzScienceNet is operating.
- AzScienceCERT service was launched for providing operative response to information security incidents in AzScienceNet, data collection in this regard, conduction of scientific-analytic research and realization of mutual relations with international institutions. This service has been registered in Trusted Introducer system operating within the framework of TERENA since May 31st, 2011.

B. The status of e-science informatization

One of the main priority directions of e-science is informatization of science and formation of information supply. Multi-purpose and multi-goal works are conducted in this course:

- Information Resources Center of ANAS Central Library has been established based on the order of the President of Republic of Azerbaijan Mr. Ilham Aliyev.
- The operation of ANAS Presidium is rapidly informatizing.
- The rating of www.science.az website of ANAS created as the first web-site of the country in 1995 has been rapidly increasing in the society recently, this, without a doubt, is a webometric indicator of ANAS.
- Multi-purpose information resources are created at scientific enterprises of ANAS, and at the same time, each of them has an active website.
- A monitoring is conducted for adaptation of country-wide scientific journals to international requirements and their electronic versions are located at DATA center of AzScienceNet network.

C. Scientific problems and research status of e-science

- The scientific-theoretical foundations of formation of e-science are researched and suggestions are developed. Investigation of creation of national scientific reference index [8, 9], scientometrics [10], the assessment of scientific activities, the application of grid and computing clouds [11], the organization and management of scientific activity with application of ICT and etc. problems can be listed as examples.
- Methods have been developed for assessment of activities of scientific enterprises and scientists, their information culture by means of web-analytic technologies.
- The protection of copyrights, intellectual property and technologies against plagiarism are being developed in online environment.
- As a trend in e-science, citizen science formation problems are investigated.
- The solution of problems of application of Big Data technologies, data storage and processing problems that are rapidly increased as a result of scientific activities is investigated.
- "1st Republic-wide scientific-practical conference on electronic science problems" organized by the Ministry of Communications and High Technologies, ANAS and IIT was held, the results of the conference were very significant and contributed to the development of e-government building [12].

D. State of the art of e-science integration to international institutions and science markets

- Since mid-1994, ANAS has been the member of TERENA international institution carrying out the mission of formation and development of computer networks of scientific and higher education institutions of Europe.
- Representation of science and education societies of Azerbaijan in GEANT Association which was founded as a result of merging of TERENA and DANTE (Delivery of Advanced Network Technology to Europe) has been considered in October 2014.

E. The status of e-science staff training

- Staff training is one of the important duties of e-science. Successive and purposeful works in this field are conducted at Training-Innovation Center of ANAS IIT.
- In 2003-2015 years in Training-Innovation Center of IIT "The principles of Science Informatics" subject has been taught to more than 15000 doctoral and PhD students of scientific and academic institutions of the country.
- Distant Training Centers were founded in Nakhchivan State University and Ganja Department of ANAS. "Scientific Informatics" subject is being taught to doctoral students and dissertationists of scientific and higher education institutions located in those regions by the employees of ANAS IIT.

V. Development Perspectives Of E-Science

- The integration process of AzScienceNet with international scientific and academic networks must be carried out continuously;
- Scientific-organizational and management activities of scientific institutions of the Republic must be completely informatized;
- Researches on scientific-theoretical principles of e-science and their practical realization must be carried out continuously;
- The works in direction of formation of information provision of science must be continued;
- The capabilities of e-science in the direction of knowledge economy formation in the country must be implemented;
- The works in direction of achievement of development of all science fields in accordance with modern world standards and their rapid integration into world science with application of e-science must be carried out.
- The storage and processing problems of scientific data must be continued with application of Big Data technologies.
- Citizen science as one of the new trends of e-science development in the country must be improved, etc.

VI. Conclusion

The rapid development of modern ICT and its broad capabilities impact the different fields of human life and activities. One of these fields is the scientific activity. For specification of structure of scientific institutions in Republic of Azerbaijan, the conduction and specification of investigations at the world standards level, the increase of scientific staff potential of the country and strengthening of their social security, the conduction of national strategy on development of science in general and the provision of integration of Azerbaijan science to international scientific environment, the realization of "e-science" problems have been necessitated. Conducted research, world experience and the analysis of the current status confirm this once again. Thus, successive works must be carried out at each institution engaged in scientific activity in the republic for the solution of those problems with consideration of realities and requirements of 21st century and international experience. At first instance, their scientific-organizational and management activity must be brought into online environment; unique national e-infrastructure providing their mutual integration must be established.

Acknowledgements

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – **Grant № EİF-2014-9(24)-KETPL-14/02/1**

References

[1] R. M. Alguliyev, T. Kh. Fataliyev. Electronic science: goals, objectives and development perspectives // 1st republic scientific-practical conference on electronic science problems, November 15-16, 2012, Baku, Azerbaijan, pp 11-12.

[2] T. Hey, A. E. Trefethen. The UK e-Science core program and the grid. Future Generation Computer Systems, October 2002, volume 18, no. 8, pp. 1017–1031.

[3] M. Atkinson. What is e-Science? //www2006.org/speakers/atkinson/atkinson.pdf

[4] ARPANET. https://en.wikipedia.org/wiki/ARPANET

[5] R. M. Alguliyev, T. Kh. Fataliyev. Toward the e-science: state-of-the-art and perspectives//Proceedings of the International conference-ICKIMICS, 2011, June 28-29, 2011, Tashkent, Uzbekistan, pp.229-232.

[6] R. M. Alguliyev, T. Kh. Fataliyev. Informational-analytic system of monitoring of ICT application in scientific activity // Proceedings of 20th International Conference "Application of new technologies in education", June 26-27, 2009, Troitsk, pp. 65-66.

[7] T. Kh. Fataliyev. Monitoring of application of information technologies in scientific activity. Baku, "Information Technology" publishing house, 2010, 68 pp.

[8] R. M. Alguliyev, R. M. Aliguliyev, T. Kh., Fataliyev, R.Sh. Hasanova. On Necessity of the Azerbaijan Citation Index (AzCI)// Proceedings of the 4th International Conference on "Problems of Cybernetics and Informatics", 12-14 September, 2012, Baku, vol. I, pp.82-85.

[9] R. M. Alguliyev, R. M. Aliguliyev, T. Kh., Fataliyev, R.Sh. Hasanova. Conceptual Challenges in Developing Azerbaijan Citation Index//Proceedings of the 4th International Conference on "Problems of Cybernetics and Informatics", 12-14 September, 2012, Baku, vol. I, pp.50-53.

[10] R. M. Alguliyev, R. M. Aliguliyev, T. Kh., Fataliyev, R.Sh. Hasanova. Scientometrics: current status and opportunities. Baku, "Information Technology" publishing house, 2013, 96 pp.

[11] R. Alguliyev, R. Alakbarov. Cloud Computing: Modern State, Problems and Prospects. Telecommunications and Radio Engineering, 2013, vol. 72, no.3, pp.255-266.

[12] The Proceedings of the 1st Republic scientific-practical conference on electronic science problems. Baku, "Information Technology" publishing house, 15-16 November, Baku, 2012, 138 pp.

The Possibilities of Using Experiments in Education of Special Technical Subjects

Gabriel Bánesz, Danka Lukáčová, Jana Depešová, Lukáš Vaněk
Department of Technology and Information Technologies
University of Constantine the Philosopher
Nitra, Slovakia
gbanesz@ukf.sk; dlukacova@ukf.sk; jdepesova@ukf.sk; lukas.vanek@ukf.sk

***Abstract*— The use of experiments in the education students of technical disciplines gain the necessary knowledge and skills for controlling technical equipment. Such experiments can be carried out in education directly or through remote means of so-called remote real experiments. The report presents two possible ways of using electro panel in education through remote and direct exercises on the subject called Parts of machines and machines. The students have an available e -learning course for preparation for the direct work with the electro-pneumatic panels. Experiences and results from the use of this course are summarized in this report. As a research method to verify the knowledge from the usage of the course an electronic test was used, which results are prepared by descriptive statistic.**

***Index Terms* — remote real experiment, educational requirements, use of remote real experiments**

I. INTRODUCTION

Experiments in education have their special position, because they allow to obtain new knowledge about the observed or researched phenomenon, validate formulated assumptions or hypotheses that form the basis of any further scientific work. Another major benefit is the fact that experiments enhance, accentuate and extend skills and knowledge of the students.

Some types of experiments, however, are demanding in terms of technical equipment, which is reflected in the price. It is ideally need to have such technical equipment that all students can work simultaneously when assigning experiments in education [1]. In practice, the inclusion of experiments in teaching is not as frequent, because it lacks the necessary technical equipment.

In recent years some university workplaces open new experiments that can be controlled in real-time and distant form. These experiments are called the remote real experiments (VRE). After introduction of these experiments into practice the following advantages and disadvantages had been proved:

Advantages:

- Real time availability
- The ability to do experiments on real machines
- Learn the method of try-mistake
- The opportunity of the experimenter to work with real results
- Solve the problem of expensive apparatuses

Disadvantages:

- experimenter does not obtain the necessary skills to work with VRE
- are only available with high-speed Internet
- Lack of involvement of more senses into the experiment
- Absence of teacher who controls the learning process itself. [2].

II. REMOTE-CONTROLLED REAL EXPERIMENT WITH THE USE OF ELECTRO-PNEUMATIC PANEL

Remote controlled real experiments in the school practice whose main objective was to develop and verify in practice the real remote experiments using the elements of industrial automatization with programmable PLC, have been used on the Department of Technology and Information Technology for several years. Therefore, another aim of the department is to develop and verify a simulator for plugging in and controlling a pneumatic system in practice. This will be used to practice plugging of simple and more complex pneumatic circuits. The system itself is planned so that it can be used by direct contact with students on lessons, but also in a distant form assembled as remote real experiment. Both ways of using this system will allow the direct handling of the device by which students acquire the necessary skills, but also will be able to deal with a distant form of homework, or practice the acquired knowledge. The acquisition of basic knowledge to work with real PLC systems is to be ensured when working with a simulator for pneumatic systems and their control through a computer program.

The proposal of the simulator for pneumatic systems will include two aspects:

1) Technical assembly

2) The establishment of teaching content for handling the simulator directly as well as through remote access via the Internet.

In terms of technical implementation will firstly be the electro-pneumatic systems assembled for direct action, as well as for remote form of control. A software will be used on the remote control, by means of which through a graphics program a pneumatic design can be formed and it will direct a pneumatic connection through the processor on the pneumatic panel.

From the perspective of didactics an e -learning course will be created to gain the necessary knowledge and skills needed to work with the simulator in direct and remote way. In addition, worksheets will be created by which students learn to create the main connections of pneumatic systems and also programming PLC systems.

III. Electro-pneumatic panel Uni Train, its description and usage

Electro-pneumatics from LUCAS NÜLLE is used for training connection of simple and complex pneumatic circuit. The utility allows you to engage individual pneumatic engagement with electrically controlled pneumatic systems.

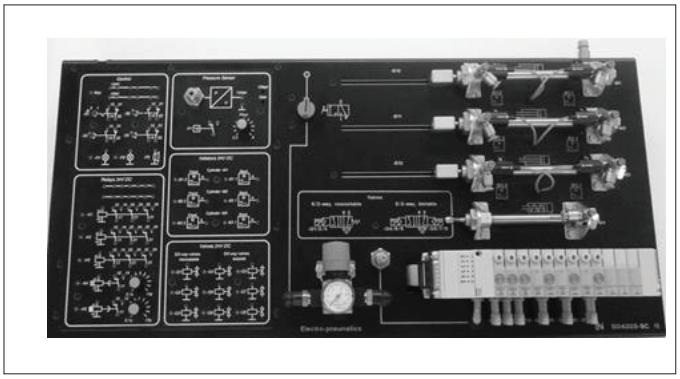


Fig. 1. Electro-pneumatic panel Uni Train [3]

The entire panel Uni Train Electro-pneumatics is divided into two parts. The first part is an electrical system for controlling the pneumatic part (left in Figure 1) and the other part consists of the elements of the pneumatic system (Figure 1 right).

The electrical system is situated on the right side of the base panel and contains the following elements:

Supply Part - Control: Power supply voltage DC 24 V / 2 A; 1 x toggle switch S1; 3 x pushbutton switches S2, S3, S4; 2 x signal diodes -P1, -P2; 1 x speaker -P3. Part of electrical relays - Relays 24V DC: 3 x electrical relay DC 24 V (K1, -K2, K3) each controlling three on-off switches; 1 x electrical relay (-K4) with a time-delay setting of the two-switch in the time interval from 0.1 seconds to 15 seconds; 1 x electrical relay (-K5) with time by closing the switch of the two-time at an interval of 0.1 s to 15 s. Part of the pressure sensor - Pressure sensor: Pressure sensor 1 / bar output from 0 V to 10 V; pressure switch from 0.5 bar to 8.0 bar. Part of the indicators extreme positions - Initiators 24 V DC 3 x indicators extreme positions actuators - cylinders (-M1, -M2, -M3) 24 V DC output and signal lights. Part of the electro-pneumatic valve - Valves 24 V DC 3 x relays for controlling 5/2 solenoid valves Q1, Q2, -Q3; 3 x 2 relays for control of bistable 5/2 valves -Q4 / -Q5, -Q6 / -Q7, -Q8 / -Q9 (2).

Second, the left part of the panel, consists of pneumatic valves and elements. Input - supply of compressed air from the compressor via quick-plug; 1 x inlet valve 3/2; 1 x control pressure 0 to 10 bar. (Default is to be set to the air pressure of 300 kPa (3 bar), unless otherwise indicated); 3 x double-acting cylinder with a piston and 80mm piston rod sensors and extreme positions; 1 x single-acting cylinder with a piston 50 mm piston rod; 3 x 5/2 single solenoid valves; 3 x 5/2 double solenoid valves. Single solenoid and double solenoid valves are integrated into one valve block. The valve block has an air inlet for all valves. If any of the valves is not in use, the outputs 2 and 4 have to be stopped. For air flow control valves are used for throttling the airflow limitation in both directions at a constant volume.

The equipment of the electro-pneumatic panel is provided a software through which it is possible to acquire the necessary knowledge and skills in assembling the pneumatic circuits. The big advantage is even the fact that within the software it is possible to draw diagrams of pneumatic connections, the accuracy of which can then be verified on the panel without having the electrical components physically engaged. The panel works with a separate processor that allows this approach. In this aspect, we see the use of this device within the remote access [4].

The mentioned electro-pneumatic panel is used in the Department of Technology and Information Technology PF UKF only directly in education without the remote access. E-learning portal course is made to help students to prepare for practical instruction for the mentioned device with a focus on integrating pneumatic systems [5]. Theme, where students can get basic information about the theory, methods of measurement and the treatment record of the measurements is structured as follows:

- The theory for practice with the panel
- Study materials
- Discussion to the exact theme
- Test

The theory includes learning materials that will bring students into the theory of pneumatics, with the emphasis on the practical work with the panel for electrical connections and especially for pneumatic connections. In the study literature part, the student has an available supplementary literature that can be used to supplement other information from hardcopy literature.

Forum allows discussion between registered participants of the course, where they can exchange information on this topic or send e-mail to a specific person or directly contact the teacher.

The last part of each topic is a test. If students pass the test, they receive permission to perform specific measurements. For successful completion it shall be deemed to control it to 51%. One aim of our review is to evaluate the results of the tests of students for the first year of application of the course and teaching aid.

IV. RESEARCH METHODS

The object of our research was the knowledge of students of the first class of OHS necessary to complete practical measurements of the thematic circle of pneumatics falling into the e-learning course for the subject Parts of machinery and Machines. An electronic test in the e-learning course was made as a research tool for the detection of knowledge, on the web of UKF in LMS Moodle system [6] [7].

The knowledge test measured the level of proficiency of the necessary knowledge of students on practical aspects of measurement of pneumatic systems. This means that it was the cognitive, output, NR-distinctive, non-standardized knowledge test. Its content has been incorporated in the curriculum theoretical basis of the subject. The number of tasks in the test was ten. The evaluation of the test was automatically done by Moodle, each test item was evaluated by one point for each correct answer. For each incorrect, incomplete or missed response the system gave zero points. If an item in the test was included ASSOCIATION response, the system for some mismatched answers counted the proportion out of one point. For the correct pass of the test the student can score 10 points. Time for drawing the test was set to 12 minutes.

The test focused on the theoretical foundations of pneumatics and pneumatic systems, while the actual measurement is performed directly in the exercises. The test contains one entry with the correct answers and assigning nine items with selection of the correct answer from four options. The focus of the questions was: 1. Properties of gases by Gay - Lussac and Boyle - Mariotte law, 2. Percentage by air, 3. Unit of gas pressure, 4. 2/3 electro-pneumatic valve scheme, Electro-pneumatic scheme monostable 5/2 valve, 5. Electro-pneumatic scheme bistable 5/2 valve, 6. The schematic indication of the compressed air supply, 7. control capability of compressed air in pneumatic systems, 8. input identifier for pneumatic valves, 9. Use 3/2 valve for controlling single acting linear pneumatic motor, 10. Use 5/2 double solenoid valve for controlling double-acting linear pneumatic motor.

V. THE PROGRESS AND RESULTS OF THE TESTS

The research was done on the subject of Parts of machinery and machines during the summer term in the academic year 2015/2016. Students had the opportunity within the e-learning course in one week to prepare for the theoretical foundations of the exercise. The real testing was done before the actual measurements at the department. 49 students worked with the course and who participated in the test.

VI. THE RESULTS OF THE TESTS FOR CONNECTING PNEUMATIC SYSTEMS

Descriptive statistics were used to compare the performance of students on the tests, which are presented in Table 1.

TABLE I. DESCRIPTIVE STATISTICS

Descriptive statistics	Table
Mean	*5,2*
Median	5
Mode	7
Standard Deviation	2,2
Sample Variance	5,2
Range	9
Minimum	1
Maximum	10
Count	49

The number of points acquired by students are summarized in Figure 1. It is obvious from the description of statistics from the test, the average number of points for all respondents is 5,2 with a relatively wide spread up to 5,2 points. This is also evident from Chart 1. Also, the same chart shows that eight students gained 4 points. The evaluation results of the test using Graph 1 shows that from 49 students, 24 students obtained more points than the diameter of the sample was achieved that is 48,9 %., only one student of these has acquired the maximum 10 points and two students gained 9 points. [8]

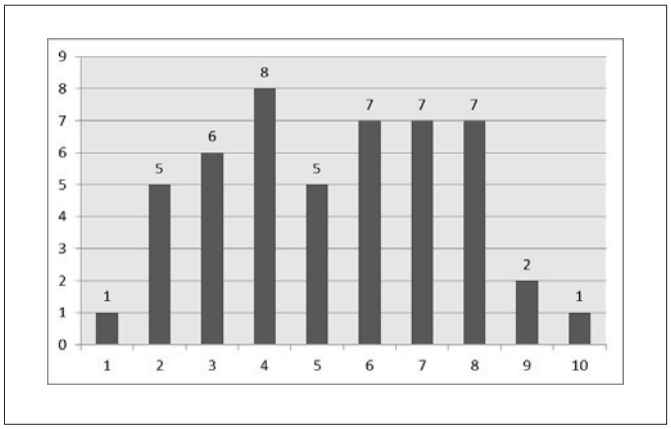


Fig. 2. Graph of the results

Similarly, a relatively strong group consist students who received less than the average (5,2 points). This is 40 % of them, from which one student get from a possible total of 10 points only one point.

As it is clear from the obtained data, just under half of the students received the necessary 51 % of the total number of points needed to work with an electro-pneumatic panel on the exercises. The second group of students had to step up their preparations for the learning, so they receive consultations with a recommendation for further study of appropriate literature.

VII. CONCLUSION

The presented results speak about the preparation of students for the immediate action with an electro-pneumatic panel. As we mentioned in the beginning of our review, we plan this utility for the future to make it available in training as

real remote experiment through the simulator. In this form, the device will be available for students above contact hours. We believe that the performance of students will improve when using such a form of practice. [9]

ACKNOWLEDGMENT

Publication of the paper was funded from the project KEGA 021UKF-4/2014 Remote Real Experiments in School Practice.

REFERENCES

[1] Ferrero, A. - Salicone, S. - Bonora, C. - Parmigiani, M. (2003), ReMLab: A Java - Based Remote, Didactic Measurement Laboratory. [w:] IEEE Transactions on Instrumentation and Measurement. roč. 52, č. 3, s. 710-715. ISSN 0018-9456

[2] Kozík, T. – Šimon, M. – Kuna, P.: Remote Experiment in Terms of View of Didactics of Education. I.: Dydaktika Informatyki, 9 (2014). Wydawnictvo uniwersytetu Rzeszowskiego, 2014. s. 200-2011. ISSN 2083-3156

[3] Lucas – Nule, 2016. On line. (24-05-2016) Available at: http://www.lucas-nuelle.com/316/apg/3751/Pneumatics-with-UniTrain.htm

[4] M. F. Paulsen, Online Education and Learning Management Systems – Global Elearning in a Scandinavian Perspective. Oslo: NKI Forlaget, 2003

[5] G. Bánesz, "LMS Moodle in safety technicians education", in. "Modernization of technical subject teaching at higher education institutions", Hradec Králové: UHK, 2013

[6] Constantine the Philosopher University in Nitra, Education portal (2016). On line. Available at: https://edu.ukf.sk/

[7] D. Lukáčová, "Using of the Internet by tertiary students", in "Modernization of technical subject teaching at higher education institutions", Hradec Králové: UHK, 2008

[8] S. Neradová, J. Horálek, "Use of the statistics at evaluation in the e-learning systems", Využití statistiky pri evaluaci v elearningových systémech, in. Jorunal of Technology and Information Education. 1/2012, Volume 4, Issue 1

[9] Milan Klement, Jiří Dostál: Learning styles according to vark classification and their possible uses in tertiary education carried out in the form of e-learning. Journal of Technology and Information Education. 2/2014, Volume 6, Issue 2 ISSN 1803-537X p. 58-67

Architecture of Smart Learning Courses in Higher Education

Elena E. Merzon
Kazan Federal University,
Elabuga, Republic of Tatarstan, Russia
Elena.Merzon@kpfu.ru

Rinat R. Ibatullin
Kazan Federal University
Elabuga, Republic of Tatarstan, Russia
Rinat.Ibatullin@kpfu.ru

***Abstract*— The article discusses the organization architecture of smart learning courses in university students' training. The course architecture is displayed in the form of a diagram. The results of testing of the proposed architecture are presented and analyzed.**

***Index Terms*— architecture course, smart education, road mapping, student.**

I. Introduction

Modern information and technologically advanced society places new demands on the education system, which includes higher education. The organization of modern learning process should promote active and cooperative work of university students in the construction of new knowledge. The educational content should be practice-oriented, provide students with tools to training activities, and promote their self-development and self-improvement [1]. Therefore, it is necessary to introduce new forms and methods of educational process organization as the main task of universities is to ensure the high quality of education [2].

One of the promising directions of education development is smart education. The concept of smart education includes flexible learning in an interactive learning environment, suggesting availability of a large number of sources, maximum diversity of media, and the ability to adapt quickly and easily to the students' level and needs. Smart education helps to develop analytical competence, skills in solving complex problems, creativity that is an ability to develop new ideas and implement them, and skills in communication with teams, groups and individuals. A key figure of smart education is the Smart student who expects multi-format flexible and personalized learning in an interactive environment, multidisciplinary training programs, and an opportunity to study and work at any time and at any place in the world on the basis of free access to the educational content.

An educational smart event is considered to be the backbone component of smart education in higher learning, the smart event meaning the logical integrity of a sequence of educational-professional and professional activities, leading to a certain result such as formation of competencies, creation of intellectual products.

Smart training can be organized only when using e-learning technologies. Taking into account the new requirements for the students' competence it is necessary to change the approach to the design of training courses that should provide accessibility, openness, mobility, interactivity and flexibility applicable to the education of today. In these circumstances, it becomes urgent to develop the course architecture providing university students' smart training.

A course is understood as a complete meaningful unit of smart educational events, developed with view of the specific educational goals. The architecture of the course is a selection of different components and the course participants with a schematic indication of their interaction.

II. Research Objectives, Methodology and Stages

The aim of the research is in the design, specification and validation of the courses architecture, providing smart learning of university students.

The proposed course architecture is shown in Fig. 1

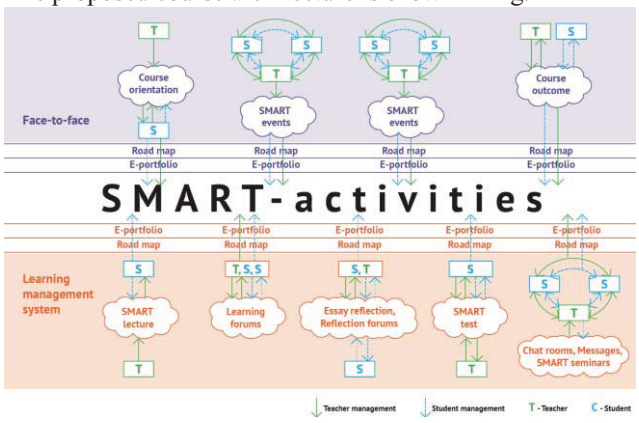


Fig. 1. Architecture of Smart learning courses in higher education

A distinctive feature of the proposed courses architecture is its focus on carrying out classes without the teacher's systematic, consistent, oral monologue presentation of the training material that is typical of the traditional lecture. Instead, introductory, goal setting and final sessions are held, bearing information-explanatory function, in which the teacher refers to the course problems and goal, the plan and the logic of learning paths, and explains the methods of work in the training course [3].

In smart education the concept of smart lecture takes on a new meaning. With the advent of smart technology in human life, the lecture becomes an interactive, personalized, multi-

format and easily accessible. Therefore, interactive lectures should become an integral part of training courses architecture providing a fundamental theoretical component to the educational process. The main part of lectures is based on the students' studying theoretical material through the element "Lecture" of Learning Management System, followed by a discussion of the studied issues at the round tables, seminars, conferences and other smart events. Smart lecture allows the teacher to place the content in an appealing and flexible way. The teacher can use interactive lecture circuit, which contains a variety of ways and options for learning. For example, the theoretical material can be divided into small portions, after each of which the students are encouraged to answer test questions or a question, involving a detailed response in the form of an essay, depending on the results of which the system or the student chooses the next portion of the training content. The teacher has a possibility to track the student's work on the item. For the purpose of the students' self-control of mastering the theoretical material, smart tests, such as multiple choice, or compliance, etc. are used.

As the need arises, the teacher in the mode 'face-to-face' conducts explanatory or corrective smart events, where the material causing difficulties for students and requiring the teacher's participation is studied. During these events the students discuss on an equal footing with the teacher the problem situations, illustrate their opinions and positions using a variety of visual materials. It should be noted that in contrast to the traditional training, in smart courses students become more active, learn to formulate questions, argue their position, and convince interlocutors. Such events are most effective if they are preceded by training in communication forums (chat rooms, wikis, etc.) of the e-learning course. Here the participants ask questions and discuss problems that they encountered in the implementation of practical tasks, as well as exchange views on educational issues. As a general rule, it is difficult for the teacher to listen and to assess each student's spoken language, and the use of educational forums allows overcoming this predicament, as well as evaluating the level of the students' competence in the theoretical part of the course and determining their scope and possession of the conceptual apparatus of the studied subject.

Modern educational process in university will be successful only if it is based on the principles of mandatory feedback [4, 5]. The undoubted advantage is the use of feedback in the process of students' reflection of their activity on reflection forums. Due to the use of this element students' reflection of their competencies is performed continuously. Namely, on the basis of performance in each block of educational activities the student, using a questionnaire for self-analysis and self-assessment, prepares a reflective report in an electronic form and sends it within the frame of the smart course. This allows the teacher not only to get an idea of the student's ability to adequately evaluate the level of completeness of the results of his work, but to analyze the results obtained in a timely manner, to predetermine the purpose of further work and to adjust the educational process to the particular student's needs.

TABLE I. FRAGMENT OF THE ROAD MAP FOR THE COURSE "IT IN EDUCATION" FOR STUDENTS OF PEDAGOGICAL DIRECTION

No	1	2
Event	Round table "Secondary education: studying a computer or studying with a computer?"	Individual work + mutual control + self-esteem
Student's activities	Participation in the collective discussion of the problem, comparison of different points of view, opinions and suggestions in the discussion.	Student's work in the electronic instructor-led course: using the recommendations and given (and found) sources independent (in a co-operation mode) study of the material, performing interactive tasks; participation in educational forums. Student's preparing for participation in the seminar "Portrait of an IT-competent teacher".
Reporting documentation	Essay-reflection, photo-report of presentation materials at the round table "Secondary education: studying a computer or studying with a computer?", publishing essays in electronic course for discussion and evaluation by the teacher and group-mates.	Running interactive activities and participating in the educational forum "Teachers' IT competence. Ways to effective change"."
Place of event	School, e-learning platform of Kazan Federal University	E-learning platform of Kazan Federal University
Date		
Mark		
Signature		

In order to ensure the course personification, the possibility of the student's reflection, correction and prediction of the student's ascent to the goals, one of the "smart" elements of the course, in our opinion, should be the student's road map, a fragment of which is presented in TABLE 1.

One of the main problems in the control of the student's progress on the learning path is the lack of a clear individual plan and the lack of his awareness of the interim results. Therefore, it is necessary to have a step by step scenario of training events with a clear indication of activities, reporting documentation, event dates, and etc. Such a scenario will be called the "Course road map". The phrase "road map" appeared from the English "road mapping", which means the process of mapping activities in any sphere. The student's road map is in fact a student's plan of development through the

events, taking into account all the known factors. In this case, in contrast, for example, with the course curriculum, the road map includes several development scenarios and is focused on a particular student. The basis of the road map is certain nodes or points at which it is necessary to make decisions important for the student in studying the course and depending on the current situation. The node corresponds to a definite stage of development and the expected time in which the student must reach this stage. It should be noted that changes can be introduced into the course road map. As a result, following the road map, the student together with the teacher is able to respond quickly to the changes in his progress.

Learning Management System creates and stores the results of each learner: the handed works, the teacher's evaluation and comments, and posts on the forum [6]. A mandatory student's road map extension is forming an e-portfolio of achievements. An important component developed by the author of the course architecture, is the formation of a single student e-portfolio for the course, including the results of classroom work. Thus, all the student's results of the course will be placed in the system.

The above-described components of the proposed architecture, and the indicated relationships between them, create the conditions for students' smart activity, their co-operation with the teacher providing successful development of professional activities. In these circumstances, the participants' roles change. Management of training course is carried out not only by the teacher but also by the students. The teacher acts as a consultant for the implementation of various educational-professional and professional actions arising in the course of training, and as an employee to address common problems in organization of cooperative work [7, 8]. A new smart content is formed by all participants while the students are going through the stages of the course.

III. Research Results and Discussion

The described course architecture was tested at Kazan Federal University. Kazan University keeps pace with the times, adhering to global trends in the development of science and education, and making a significant contribution to the development of the most advanced forms and methods of teaching. Among them are the innovative educational technologies, e-learning. The purpose of the e-learning introduction into the University education is to increase the efficiency of the learning process through the use of active learning methods and individualization of educational trajectories.

Smart learning testing took place in Elabuga Institute of Kazan Federal University from September 2014 to May 2016. The experiment involved 14 teachers and over 750 students.

As part of the experiment part of the classroom activities was transferred into the e-learning courses developed on the e-learning platform in Learning Management System. Part of teachers abandoned traditional lecturing almost completely.

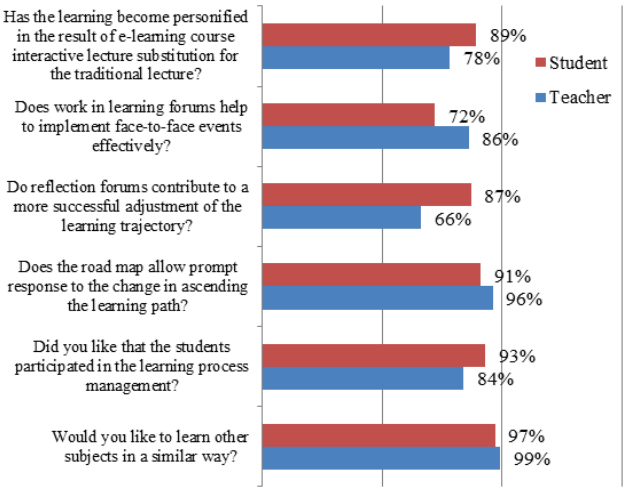


Fig. 2. Results of testing participants' questionnaire

In order to get an idea of the effectiveness of the author's course architecture teachers and students were surveyed. As can be seen from the survey results, the majority of the respondents approve of this form of training. The students note such positive aspects as accessibility and openness of the course content; the availability of means of interaction with all participants in the educational environment through the forums, chats, workshops, etc.; the teacher's evaluation of the student's performance transparency, timeliness and informative character; the ability to manage their time; participation in the organization of the learning process; roadmap allowing to see the learning path clearly.

All the testing participants, without exception, noted that the undoubted advantage is the ability of the system to evaluate the student's work with the theoretical material automatically, which is impossible with lectures-monologues. The teachers expressed the view that the students' responsibility for the learning outcomes increased; the teacher's awareness about the state of students' learning also increased as the teacher knows about difficulties and successes, which allows adjusting the learning process to the students' needs.

Some teachers faced the challenge of creating lectures that should meet the requirements of the new format. Therefore, they need to improve their competence constantly, in particular, by immersing themselves in the training course built on the basis of the proposed architecture. As a result, the teacher is able to try on the role of the student, experience all the features of this training and prevent possible errors in the construction of their copyright courses. Teachers studying in an environment based on this course architecture are able not only to learn, but also to exchange opinions, help each other to resolve the difficulties that they may encounter.

IV. Summary and Conclusions

The designed course architecture allows for a fresh look at the organization of learning through Learning Management System. This approach uses the system elements for the

organization of the students' and teachers' cooperative smart work. The results showed that the rejection of lecture-monologue in favor of smart event format is not only possible, but also effective. With this organization of learning students become more active, independent, learn to design their own learning path and create new knowledge in cooperation with all the participants of the educational environment. One of the significant results of the architecture application is the student's willingness to implement smart activities in an open, interactive and multi-format educational environment.

The undoubted advantage of the proposed architecture is the student's work on the road map, which in turn generates a student's ability to implement self-diagnosis, planning and forecasting of his activities. This solves the problem of learning personalization in students' academic groups.

Furthermore, the advantages of the proposed architecture is its flexibility, which is the ability to add, move, arbitrary arrange the course elements at the author's judgement. The tested course architecture is equally well applicable to both humanities and natural sciences. The results of testing can be used to build training courses, providing flexible learning in an interactive learning environment.

ACKNOWLEDGMENT

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

[1] E. M. Ljubimova and E. Z. Galimullina, "Development level of independent activity of undergraduates on the basis of Web-technologies", Life Science Journal, ISSN: 1097-8135, Scopus Coverage Years: from 2008 to 2013, 11 (SPEC. ISSUE 11), 110, pp. 485-488. http://www.lifesciencesite.com. 110. URL: http://www.lifesciencesite.com/lsj/life1111s/110_26040life1111s14_485_488.pdf.

[2] E. Z. Galimullina and E. M. Ljubimova, Training students of language on the use of information technologies, International Congress on Interdisciplinary Business and Social Science, Indonesia, 4-5 November 2013.

[3] E. M. Ljubimova, E. Z. Galimullina, and R. R. Ibatullin, The development of university students' self-sufficiency based on interactive technologies by their immersion in the professional activity, International Education Studies, vol. 8, No. 4, April 2015: pp.192-199. (ISSN 1913-9039). DOI: 10.5539/ies.v8n4p192. http://www.ccsenet.org/journal/index.php/ies/article/view/47011

[4] E. M. Lyubimova, E. Z. Galimullina, and R. R. Ibatullin, Practical orientation increase for future teachers training through the integration of interactive technologies, The Social Sciences, 2015, No. 10: pp.1836-1839. DOI: 10.3923/sscience.2015.1836.1839 URL: http://medwelljournals.com/abstract/?doi=sscience.2015.1836.1839.

[5] E. Z. Galimullina, Mechanisms of integration of interactive forms and methods in the educational process of higher education, Modern problems of science and education, 2014, No. 4. URL: http://www.science-education.ru/ru/article/view?id=13972.

[6] V. Y. Shurygin and L. A. Krasnova, Organization of students' independent work in studying physics based on distance learning technology in LMS MOODLE. The Education and science journal, 2015; No.8: pp.125-139. (In Russ.) DOI: 10.17853/1994-5639-2015-8-125-139.

[7] E. G. Galimullina and Y. M. Lyubimova, Model of network communication between establishment of higher education and school in terms of intensification of practical orientation of bachelor's training of pedagogical education, The Social Sciences, 2015, No. 10: pp.956-964. DOI: 10.3923/sscience.2015.956.964. URL: http://medwelljournals.com/abstract/?doi=sscience.2015.956.964.

[8] E. M. Lyubimova and I. A. Borisov, Networking school-high school as a means of students' immersion in professional activity, Modern problems of science and education, 2015 No. 1-1. URL: http://www.science-education.ru/ru/article/view?id=19426.

Evolution of Dynamic User Models for Adaptive Educational Hypermedia System (AEHS)

Nafisa Abdo
Department of Computer Science
University of West Kordofan
Elnihood, Sudan
nafisa.abdo@gmail.com

Noureldien A. Noureldien
Department of Computer Science
University of Science and Technology
Omdurman, Sudan
noureldien@hotmail.com

***Abstract*— *The Internet and the web change many human life aspects including education.* The web provides new means for delivering educational material, so the concepts of e-learning and distance learning have evolved.**

The web technologies are tailored to personalize learning process which results in what is known as Adaptive Educational Hypermedia Systems (AEHS). AEHS constitutes three basic modules, the content module, the adaptation module and the user module. The user module is the heart of the system, based on which the adaptation module personalize the content. A User models in AEHS is either static, i.e the information about the user characteristics is collected and updated before starting the adaptation or in habitual intervals, or dynamic where the user information is updated continuously as the user interact with the system.

In this paper we surveys systems that employ dynamic user models and research work that proposes a dynamic user mode, to determines the evolution of user characteristics that have been involved in building dynamic user models during the last decades.

***Index Terms*—** Adaptive Educational Hypermedia Systems (AEHS), dynamic user model, dynamic characteristics, knowledge, learning styles, goals/tasks, behavior.

I. Introduction

Adaptive Hypermedia systems are systems that use the hypermedia (hypertext, links,…..) to personalize the content material, presentation style, and navigation of a web application to individual users based on their characteristics and preferences [1] [2]. This personalization or adaptation is generally classified as either presentation support or navigation support. In presentation support certain techniques are used to adapt content material and the way of presentation of this content to support individual user needs. In navigation support, some techniques are used to control the links and therefore the navigation through the content material hypertext [3].

Adaptive Educational Hypermedia Systems (AEHS) provides a user tailored educational environment in contrast to the traditional educational environment in which the same educational material, presentation and navigation is provided to all users in a one size fits all model. This adaptive model provided by AEHS act in response to individual differences between system users in their background on domain knowledge, preferred learning style, attitudes and other individual learning characteristics and hence increases learning outcomes and user satisfaction [4].

An AEHS consists of three main components; the domain model (DM), the user model (UM) and the adaptation model (AM).The domain model is the knowledge domain area which the adaptive educational hypermedia system is used as a learning resource that is shaped in content and presentation to individual users. The user model is a collection of user characteristics that are used to adapt the content in the domain model. The adaptation model describes the parts of the hypermedia system that to be adapted and the circumstances under which this adaptation is to occur [5]. According to [3], mainly the presentation of the hypermedia and the navigation through the hyperspace can be adapted.

Early AEHS employ only what is known as static UM, where user characteristic are captured either prior the user use of the system using queries, forms, tests, …etc or during the system usage using the same methods in regular or predefined intervals. The adaption in such systems will sustains static until the next interval update. Later, the concept of dynamic user characteristics is evolved to describe some user characteristics that keep changing continuously during the user usage of the system. Utilizing such characteristics in building user models define the second type of user models, the dynamic user mode (DUM) [6].

Dynamic user models (DUM) facilitates the implementation of dynamic hypermedia applications that adapt themselves at each step to the user, the glory application example is adaptive educational hypermedia systems. The survey in this article focuses on dynamic user models for AEHS. The objectives of this survey are: to enumerate the user characteristics that have been employed in designing dynamic user models for AEHS, and to determine the evolution directions in the development of dynamic user models for AEHS.

The rest of this paper is organized as follows: Section 2 provides a survey to proposed dynamic user models for AEHS. Section 3 gives our observation on the evolution directions of dynamic user modeling. In section 4 conclusions are given.

II. Evolution of Dynamic User Models for AEHS

To implement a user model the system should collect information about each user and store and update this information. Information about individual user usually contains static information such as personal data, demographic and geographic information, and initial values for some user

characteristics that change dynamically such as knowledge about the domain subject, favorites, preferred learning style, the learning environment, …etc [7].

Data contained in the user model can be grouped into two categories; information that are related to the educational material or the domain content provided by the AEHS, such as user knowledge about topics and concepts in the domain, the user preferred style of learning, and unrelated domain information such as user demographic and geographic information, interest and aptitudes [8].

Researchers and AEHS developers have designed and developed many real world AEHS that use user models with variant number and type of user characteristics. It is obvious that early systems use one or two dynamic user characteristics in their user models. To track the later growth of dynamic user models in AEHS in terms of dynamic user characteristics in order to achieve the first goal of this survey, a literature review for user models proposed in academic research, and developed AEHS in the period 1998-2015 was conducted.

Table (1) summarizes the reviewed models. The forth column of the table lists the user characteristics used by the model, and the fifth column shows the dynamic ones.

TABLE 1: DYNAMIC CHARACTERISTICS FOR UM IN AEHS

Proposed User Models Systems	Authors	Years	User Characteristics Used by the Model	Dynamic Characteristics
Hybrid Dynamic User Model [9].	Karagiannis & Satratzemi	2014	• Knowledge • Behavior • Learning style • Goals • Experience.	• Knowledge • Behavior • Learning style • Goals
Triangular Learner Model (TLM) [10].	Loc, N. P	2014	• Knowledge • Learning style	• Knowledge • Learning style
PATS (Personalizing Affective Tutoring System) [11].	Soudani & Aghaee	2013	• Learning style • Emotion	• Learning styles
ULUL-ILM (Web – based AEHS) [12].	Bayasut et al.,	2013	• Learning style • User profile	• Learning style
LS-AEHS (Learning Style based Adaptive E-learning Hypermedia System) [13].	Samia & Abdelkrim	2012	• Learning style	• Learning style
WELSA (Web-based Educational system with Learning Style Adaptation) [14].	Popescu & Badica	2011	• Learning style • Preference	• Learning style
AEHS-LS Proposed Model [15].	Mustafa & Sharif	2011	• Knowledge • Learning styles • Learner profile	• Knowledge • Learning styles
Proposed Adaptive e-Learning System [16].	Esichaikul, et al.,	2011	• Knowledge level • Personal data	• Knowledge
Adjustable Three-layer Fuzzy Cognitive Map [17].	Georgiou, et al.,	2011	• Cognitive traits • Learning style	• Learning style
AIWBES (A new Approach for Building Student Model in an Adaptive and Intelligent Web-Based Educational System) [18].	Homsi, et al.,	2008	• Knowledge	• Knowledge
AEHS in Biomedical Engineering Project [19].	Almeida & de Azevedo	2008	• Preference • Knowledge	• Knowledge
New User Modeling Approach to Determine the Knowledge Status of User in AEHS [20].	Colak et al.,	2008	• Knowledge state	• Knowledge

User Models for Adaptive Hypermedia and Adaptive Educational Systems [21].	Brusilovsky, & Millan,	2007	• Knowledge • Interests • Goals • Background • Individual traits • Context of work	• Knowledge • Goals • Learning style
Design and Implementation of the User Modeling System [22].	Froschl	2007	• User profile	-
Proposed User Model with the Personality Type [23].	AL- Dujaily	2007	• Preferences • Knowledge • Personality type	• Knowledge
EDUCE (Intelligence Profile in AEHS) [24].	Kelly & Tangney	2006	• Knowledge • Preferences	• Knowledge
Jointzone (An Adaptive Web-based Learning Application) [25].	Maier, P et al.,	2005	• Knowledge	• Knowledge
Learning Styles and Pedagogy in Post-16 Learning [26].	Coffield et al.,	2004	• Learning style	• Learning style
INSPIRE [27].	Papanikolaou et al.,	2003	• Learning style • Knowledge • Interests • Goals	• Learning style • Knowledge • Goals
AHyCo [28].	Momar et al.,	2003	• Knowledge	• Knowledge
AEHS Based on Multiple Characteristics [29].	Surjono, H. & Maltby, J.	2003	• Knowledge • Learning styles • Experience • Background • Preferences	• Knowledge • Learning styles
Proposed User Model of an AHS [30].	Brusilovsky	2002	• Knowledge • Background • Goals • Experience • Preferences	• Knowledge • Goals
OPAL (Open Adaptive Learning Environment) follow SCORM standard [31].	Conlan et al.,	2002	• User profile • Preferences	• Knowledge • Goals
Engineering Approach for Adaptive Hypermedia Systems (UWE) [32].	De Koch	2001	• Knowledge • Preferences • User profile	• Knowledge
AHA! An open adaptive hypermedia architecture [33].	De Bra & Calvi	1998	• Knowledge • Goals	• Knowledge • Goals
ANATOMTUTOR [34].	Beaumont	1998	• Knowledge	• Knowledge
ISIS-TUTOR [35].	Brusilovsky & Pesin	1998	• Knowledge • Goals	• Knowledge • Goals

The literature survey shows that different characteristics have been proposed to be incorporated in designing user models for AEHS. These characteristics include: Knowledge, Background/ Experience, Preferences, Interests, Goals/Tasks, Personality Traits, Cognitive Traits, Learning Styles, Personal Data, Geographic Data, Demographic Data, Abilities/ Disabilities, Behavior, Emotion, Social/ Group, and Environment/ Work. User profile as a characteristic in table (1) normally contains demographic, geographic information and initial values for some dynamic characteristics. Table (2) shows a precise definition of the above user characteristics. Out of this list four characteristics are used in dynamic user models; knowledge, learning style, objectives (goals / tasks) and behavior.

TABLE 2: USER CHARACTERISTICS IN UM

Characteristic	**Meaning**
Knowledge	User Knowledge is the set of facts, concepts, and skills about a subject or a domain acquired by a user.
Background/ Experience	User Background\Experience is the previously user acquired knowledge about a subject or a domain.
Preferences	User Preferences are user characteristics that determine or affect user's likes and dislike.
Interests	User Interests are user concerns and things that get user attention and seem to be important to him.
Objectives/ Goals/Tasks	User Objectives or goals are attributes that reveal the user aim of using the AEHS application.
Personality Traits	User Personality Traits represent user's identity whether its introverted or extroverted.
Cognitive Traits	User Cognitive Traits represent user's individual characteristics such as affective, impulsive, …etc.
Learning Styles	User Learning Style is the preferred way of learning to the user, this style determines the user's preferred type of information, the way of presentation of learning content and other learning alternative attributes.
Personal Data	Personal Data, such as name, email, password… etc.
Geographic Data	Geographic Data such as, city, state, area code… etc
Demographic Data	Demographic Data are age, language… etc.
Abilities/ Disabilities	This is a static user attributes that determines whether the user is well off or having some biological, mental disability.
Behavior	The way in which a user acts in response to a particular situation or stimulus.
Emotion	Emotion, away to identify the affective state of users [11].
Social/ Group,	The group or society the user belongs to.
Environment/ Work	Is used to describe the surrounding conditions in which the user operates

III. OBSERVATIONS ON EVOLUTION DIRECTIONS OF USER MODELS

Based on the historical evolution in table (1), we deduce the following observations on evolution directions of dynamic user models for AEHS.

- The knowledge acquired by the user is a dominant factor in building user models. This is obvious since acquiring knowledge is a major learning goal.
- Learning style is emerging as an important dynamic user characteristic in last five years.
- Currently user behavior has been included in user model to enhance adaptation to fit with user behavior using special dedicated adaptation techniques.

IV. CONCLUSION

Theoretically it may seems to be logical that the increase of number of user characteristics involved in dynamicity of user model will increase the degree of adaptability of the educational system, and hence increase user satisfaction. The historical review of academic and research proposed dynamic user models do not reflect this expectation, but more dynamic user characteristics are expected to be employed in future user modeling

The lack of authoring tools for developing user model for AEHS, and techniques that capture dynamic change in different user characteristics affects the developing of sophisticated user models that are highly adaptive.

REFERENCES

[1] E. Martinez, Y. Sherry, and L. Xiaohui, "Survey of Data Mining Approaches to User Modeling for Adaptive Hypermedia," IEEE Transactions on System, MAN, AND Cybernetics: Applications and Reviews, VOL. 36, NO. 6, Nov 2006

[2] P. Brusilovsky, "Methods and techniques of adaptive hypermedia," User Modeling and User-Adapted Interaction, vol. 6, pp. 87-129, 1996.

[3] P. Brusilovsky, "Adaptive hypermedia," User Modeling and User Adapted Interaction, 11 (1/2), 87-110.

[4] E. Popescu, T. Philippe; B. Costin, "Towards a unified learning style model in adaptive educational systems," Proc. ICALT 2007, pp. 804-808, 2007.

[5] P. DeBra, A. Aerts, D. Smits, and N. Stash, "Aha! Version 2.0, More Adaptation Flexibility for Authors," Proceedings of the AACE ELearn'2002 conference, pp, 240-246. 2002.

[6] S. Baldiris, S. Graf, and R. Fabregat, "Dynamic User Modeling and Adaptation based on Learning Styles for Supporting Semi-Automatic Generation of IMS Learning Design," 11th

IEEE International Conference on Advanced Learning Technologies, pp 218-220. 2011.

[7] L. Nguyen, and D. Phung, "Learner Model in Adaptive Learning," World Academy of Science, Engineering and Technology 21 2008, pp 395-400

[8] P. Brusilovsky, "The Construction and Application of Student Models in Intelligent Tutoring Systems, " Journal of Computer and System Sciences International, vol.32, no. 1 p.p. 70- 89, 1994.

[9] I. Karagiannis, and M. Satratzemi, "Comparing LMS and AEHS: Challenges for Improvement with Exploitation of Data Mining," IEEE 14th International Conference on Advanced Learning Technologies, pp. 65-66, IEEE. 2014.

[10] N. Loc," A User Modeling System for Adaptive Learning, Standard Scientific Research and Essays," Vol, 2 (4), pp. 065-209.

[11] S. Soudani, and N. Aghaee, "A Model for Personalized Intelligent Tutoring Systems, International Journal of Advanced Computer Technology (IJACT), VOLUME 2, NUMBER 5,pp,9 – 13. 2013.

[12] B. Bayasut, G. Pramudya, and H. Basiron, "ULUL-ILM the Design of Web-Based Adaptive Educational Hypermedia System Based on Learning Style," J 3th International Conference on Intelligent Systems Design and Applications (ISDA). PP, 147 – 152. 2013.

[13] D. Samia, and A. Abdelkrim, "An Adaptive Educational Hypermedia System Integrating Learning Styles: Model and Experiment," International Conference on Education and e-Learning Innovations, IEEE. 2012.

[14] E. Popescu, and C. Badica, "Creating a Personalized Artificial Intelligence Course: WELSA case study," International Journal of Information Systems and Social Change (IJISSC), vol. 2, no. 1, pp. 31 -47, IEEE. 2011.

[15] Y. Mustafa, and S. M. Sharif, " An approach to Adaptive E-Learning Hypermedia System based on Learning Styles (AEHS-LS): Implementation and Evaluation," International Journal of Library and Information Science, Vol. 3(1), p. 15-28. 2011.

[16] V. Esichaikul, S. Lamnoi, and C. Bechter, "Student Modelling in Adaptive E-Learning Systems," Knowledge Management & E-Learning International Journal, Vol.3, No.3. pp.342-355. 2011

[17] D. A. Georgiou, S. Botsios, V. Mitropoulou, M. Papaioannou, C. Schizas, and G. Tsoulouhas, "Learning Style Recognition Based on an Adjustable Three- Layer Fuzzy Cognitive Map," JAISCR, Vol.1, No.4,PP.333 – 347. 2011.

[18] *M. Homsi, R. Lutfi, R. Maria, and G. Barakat, "Student Modeling Using NN-HMM for EFL Course", In Proc. of the international conference on Information and Communication Technologies: From Theory to Applications, ICTTA, 2008.*

[19] M.A. Almeida, and F. M. de Azevedo, "A Categorical Model of Adaptive Educational Hypermedia Systems in Biomedical Engineering Project," 5th International Conference on Information Technology and Application in Biomedicine, Shenzhen, China pp.611-614. 2008.

[20] I. Colak, S. Sagiroglu, and H. Tolga, "A User Modeling Approach to Web Based Adaptive Educational Hypermedia Systems," Seventh International Conference on Machine Learning and Applications, IEEE Computer Society. 2008.

[21] P. Brusilovsky, and E. Millan " User Models for Adaptive Hypermedia and Adaptive Educational Systems," Springer, Volume 4321 of the series Lecture Notes in Computer Science pp 3-53, 2007

[22] C. Froschl, "User Modeling and User Profiling in Adaptive E-learning Systems", Master's Thesis, Graz University of Technology. 2007.

[23] A. AL- Dujaily, "Personality Effect in the design of Adaptive E-Learning Systems," PhD Thesis, Massey University, Auckland, 2007.

[24] D. Kelly, B. Tangney, "Adapting to Intelligence Profile in an Adaptive Educational System," Interacting with Computers, 18, pp. 385–409. 2006.

[25] P. Maier, R. Armstrong, W. Hall, and M. Hong, "An Adaptive Web-based Learning Application," Larning, Media and Technology, volume 30, issue 3, p. 281-297. 2005.

[26] F. Coffield, D. Moseley, E. Hall, and K. Ecclestone "Learning Styles and Pedagogy in Post-16 Learning. A Systematic and Critical Review." Learning and Skills Research Centre, UK. 2004

[27] K. A. Papanikolaou, M. Grigoriadou, H. Kornilakis, and G.D. Magoulas "Personalizing the Interaction in a Web-based Educational Hypermedia System: The Case of INSPIRE". UMUAI, 13, pp 213-267. 2003.

[28] V. Momar, N. H-Bozic, and D.P. Zokovic, "Approaches to Online Testing in Web-based Educational Systems," IEEE, EUROCON Ljubljana, Slovenia, pp.343 – 346. 2003.

[29] H. Surjono, and J. Maltby, "Adaptive Educational Hypermedia based on Multiple Student Characteristic'," In Proceedings of the Second International Conference on Web-based Learning (ICWL). Springer Verlag. 2003.

[30] P. Brusilovsky, "Adaptive hypermedia," User Modeling and user Adapted Interaction, 11 (1/2), 87-110, 2001.

[31] O. Conlan, D. Dagger, and V. Wade, "Towards a Standards-based Approach to e-Learning Personalization using Reusable Learning Objects," Proc. of World Conference on E-Learning, E-Learn 2002.

[32] N.P. De Koch, "Software Engineering for Adaptive Hypermedia Systems Reference Model, Modeling Techniques and Development Process", PhD Thesis, University Munchen. 2001.

[33] P.D. Bra, L Calvi, "AHA! An Open Adaptive Hypermedia Architecture, "New Review of Hypermedia and Multimedia 4 (1), 115-139. 1998

[34] I.H. Beaumont, " User Modeling in the Interactive Anatomy Tutoring System ANATOM-TUTOR," User Modeling and User-Adapted Interaction, 4(1), p. 21-45, Reprinted in Adaptive Hypertext and Hypermedia, Kluwer Academic Publishers, p. 91-116. 1998.

[35] P. Brusilovsky, and L. Pesin, " Adaptive Navigation Support in Educational Hypermedia," An Evaluation of the ISIS-Tutor", pp. 1 – 23.1998

[36] C. Mulwa, S. Lawless, M. Sharp, I. A-Sanchez, and V. Wade, "Adaptive Educational Hypermedia Systems in Technology Enhanced Learning: A Literature Review", ACM Special Interest Group for Information Technology Education Conference, SIGITE, Central Michigan University, Midland, MI, USA, October 7th 9th, ACM, pp. 73 – 84. 2010.

[37] E. Frias-MartinezSh. Y. Chen, and X. Liu, "Survey of Data Mining Approaches to User Modeling for Adaptive Hypermedia," IEEE Transactions on Systems, MAN, and Cybernetics: Applications and reviews, VOL. 36, NO. 6, pp. 734 – 749. 2006.

IT in the Educational Programs in the Field of: "Safety of Person's Activity"

Dmitrieva I.A., Gordienko L.V., Popova O.V.
The Department of Ecology and Life Safety, The Department of Information Technology and Measuring Systems
Southern Federal University (SFedU),
Taganrog, Russia
idmitrieva@sfedu.ru, lgordienko@sfedu.ru, ovpopova@sfedu.ru

Abstract—**The article considers the additional professional education (APE) program "Occupational Safety and Health (OSH)" and the master's program "Health Safety and Environmental (HSE) Management System" for graduate students. In the article the program urgency, goals, objectives, and content are analyzed. The multidisciplinary approach towards teaching of university students and training concerning the given programs is taken into account.**

Index Terms—**The additional professional education program "Occupational Safety and Health (OSH)", OSH expert, the master's program "HSE Management System", labor and safety protection, distance learning courses and technologies, geographic information systems (GIS).**

I. Introduction

Nowadays, in the era of academic mobility and globalization, professional compatibility, development of modern technology and the appearance of new requirements for different occupation positions, the quality improvement of management of the enterprises need to respond to insurance claims in order to predict and prevent occupational safety and health risks.

It goes without saying that the education, training and practice to require occupational safety competence is actual and necessary. It is concerned with such failures of materials, machines, processes and structures as may give rise to dangerous situations, including the release of harmful agents. The aim of education in this field is to enable students to foresee the danger, both at the planning stage of projects and in existing situations, to quantify the danger and to design measures to combat it. Training in occupational safety involves students in a substantial study of selected problems from engineering and materials science, particularly those related to mechanical, civil, chemical, electrical and structural engineering.

The additional professional education (APE) program "OSH" is aimed at qualified personnel training to require necessary competencies in order to perform the professional activity in the field of health and safety at enterprises and organizations. The purpose of the master's program is to provide graduates with the ability to successfully implement the acquired knowledge and skills in the field of technological processes and production safety (considering the security of technologies for the health and life of the working personnel); and environmental safety of processes and production (concerning the safety of production technologies for the environment and for the population) in the chosen area of activity (in science, manufacturing, business and management); the formation of general cultural, general professional, and specific professional competences in the process of implementation of innovative programs and learning technologies that ensure social mobility and competitiveness of graduates in the labor market and prepare graduates for self-training and continuous professional self-improvement.

The system of additional professional training and education program "OSH" and the master's program "HSE Management System" requires multidisciplinary training in the process of the process of teaching of university students. For example, separate curricular units are to be concerned with the structure and strength of materials in mechanical engineering; with forces in structures in civil and computer engineering; with the handling and transportation of chemicals in chemical engineering; with design standards, protective equipment and the theory of preventive maintenance in electrical engineering; and with the behavior of strata in mining engineering.

Application of geographic information systems (GIS) in education is actual. GIS offers effective methods of the solution of territorial problems, and geospatial information is one of the most important elements of the solution of economic, social, political and ecological tasks.

Therefore, preparation of the qualified personnel able to use and develop GIS-technologies and effectively work with digital geospatial data for research and decision-making acquires special significance.

GIS-education has a number of peculiar features:

- interdisciplinary character (working with GIS is based on knowledge of cartography, databases, computer graphics, health and safety, labor protection, etc.);
- engineering and technological component;
- extensive information component;
- wide range of applications.

Training to manage data and what is more important, use of thematic mapping with professional models of social, economic and natural processes, receptions of the multidimensional analysis and expert-evaluation for optimization of environmental management and monitoring is a special task of student's GIS-education.

II. Problem Statement

Today in Russia we have the process of the replacement of existing inefficient OSH management system to a new system of occupational risk management in accordance with the recommendations of the standards of the International Labor Organization (ILO). In this connection, this work should be carried out by specialists who possess the appropriate knowledge and are able to develop effective preventive measures to improve working conditions [1].

The necessity and actuality of the carried out programs are maintained by the government document, which has the force from July 1, 2013 based on the order of the Ministry of Labor of Russia from May, 15, 2013, No 205. According to this order from the Qualification schedule of managers, professionals and other employees (approved by the Resolution of the Ministry of Labor of Russia from August, 21, 1998 № 37) (hereinafter - Qualification Handbook) the positions of "Head of Protection Labor", and "Safety Engineer "and their qualifying performance were eliminated [2].

Instead of the eliminated positions, by the order the Health Ministry of Russia from May, 17, 2012, No 559n, which entered into force on the 1st of July, 2013 the following positions were added:

- «Head of labor protection" (for a qualification – additional professional education (APE) in the field of occupational safety and health with work experience in the field of occupational safety at least five years);
- «Specialist in Occupational Safety" (qualification requirements – retaining in the field of occupational safety and health) [3].

These changes caused anxiety among OSH engineers and employers who have the position of an engineer for a labor protection in their staff.

In this situation, employers become interested in the program, developed by the experts of the Department of Ecology and Life Safety.

III. Methodology

Engineering Technological Academy of Southern Federal University (ETA SFedU) retraining program has *the duration* of 1 year of study, 504 hours of *the total workload*, including 404 hours of classroom work. *Form of study* is part time distant and correspondence work. *The final document of education*: a Diploma of Professional Retraining, giving the right to conduct a new activity.

The goal of the program: training of qualified personnel to require necessary competences to carry out professional activities in the field of health and safety at enterprises and organizations.

Laboratory and practical training is 404 hours of classwork.

APE program "OSH" is opened for the purpose of preparing skilled workers, specialists in the sphere of occupational safety and health to ensure the safety of employees of enterprises and organizations. It involves training and personal and professional development of specialists, enables graduates to implement a new type of professional activity successfully, to realize their potential, ensure social stability and mobility in the labor market. Professional activity of graduates is aimed at ensuring the safety of employees and the consistent improvement of working conditions and safety in a business or organization. OSH specialists solve the problems of the analysis of dangerous and harmful factors of the working environment, technical expertise, monitoring of working conditions for compliance with the regulations and safety standards, identifying and addressing violations that endanger the life and health of employees to carry out preventive work aimed at forming safe behavior.

To implement this program, we have designed the website for distance learning according to the program of professional retraining. In order to achieve this objective, the domain of the first level has been acquired. The server is intended for storage and maintenance of the websites on it.

After analyzing the existing information about the hard drives, we plan to use the solid state drive (SSD). SSD is a non-mechanical computer memory device based on memory chips. Besides them, the SSD includes a master controller. The most common type of solid state drives uses for storage NAND-type flash memory.

The designed website has a number of advantages:

- The possibility to access the website round-the-clock for more than 200 people (at a time can be on the website (IP addresses) and the website will not suspend).
- Unlimited number of subdomains – that gives the possibility to make large on volume of materials website (treelike)
- The number of placed materials is limited only to the volume of files, for example, 1 picture size 1 gigabyte or 1000 1 megabyte pictures, etc.
- Hosted by German servers, which means stable performance, high speed downloads and fast ping.
- Support of popular advanced platforms for editing the website, including WordPress.

The website is created in the WordPress editor. WordPress - is the ideal platform for the publishing focused on beauty, support of standards and usability. WordPress is free and available for distribution [8].

This design tool is really powerful with good history and achievements. After a long and arduous editing, we succeeded to create the website with reliable protection.

What has ensured security:

- Login.

- Password (letters and numbers) is long that significantly complicates probability of accidental input of the correct password.
- Function Captcha - the compulsory solving of simple math tasks for confirmation of all entered parameters that significantly complicates use of password guessing programs.
- Function Google Authenticator – it is necessary to enter the six-digit password which is generated by Authenticator Google service every 2 minutes which significantly complicates the use of Trojan viruses to steal passwords.

In case of DDoS-attack attempt the server switches off the website for some time with the subsequent notice that attack has been made that complicates successful DDoS-attack on the website and the server where it is stored.

After error-free entry of all parameters access to the site Edit menu is allowed, where You can find very diverse settings and menu programs that can contribute to the design and protection. Site Protection is further ensured by WordPress Security program.

For example, protection on verification of comments: the message comes to e-mail and it needs to be approved or ignored. Also the website has many other functions which are not disclosed for safety of the website.

The Curriculum APE OSH includes the worked out materials that are available and are at the disposal of the university members at the Department of Ecology and Life Safety in electronic form: lecture notes, guidelines for practical work and supervised independent work, and presentations of the lectures [4].

Engineering Technological Academy of Southern Federal University (ETA SFedU) Master Program "Health Safety and Environmental Management System" has *the duration* of 2 years of study, *credit hours:* 120. *Form of study* is full time. *Admission*: it is necessary to have bachelor education. *The final document of education*: a Master Diploma of "HSE Management System".

The general purpose of the program is to provide graduates with knowledge and skills in the field of technological processes and production safety, security technologies for the health and life of the working personnel, and environmental safety of processes and production; to form professional competences in the implementation of innovative programs and learning technologies that ensure social mobility and competitiveness of graduates in the labor market.

The Master Program is designed in accordance with the Federal-State Standard of Higher education. Implementation of the program will occur as a result of the joint activity of the departments of the Institute of Management in Economic, Ecological and Social systems of SFedU and representatives of the professional community [5].

Basic Courses: IT, risk management and modeling in a technosphere safety, Economical and organizational bases and monitoring of a technosphere safety.

Elective courses: Methodology of scientific researches and design / Technologies of safety processes and production/ Ecological safety as part of the international relations / Principles of environmental systems protection and resource-saving technologies development / Business Professional Foreign language /Environmental impact assessment/ Technological electrical safety and perspective power technologies / Materials, methods and means for providing a technosphere safety / Economic and legal support of nature protection activity/ Economic and legal support of nature technosphere safety.

To achieve the intended learning outcomes, the following methods of educational technology have been used: the methods of IT: the use of Internet resources to expand the information field, processing, transmitting and receiving information; the interdisciplinary training: a kind of training with using knowledge from various fields (disciplines), implemented in the study of a particular discipline; the learning from experience: activating the cognitive activity of the student association at the expense of their own experience with the subject of the study; the research methods: the cognitive activity aimed at the acquisition of new theoretical and factual knowledge through research, independently or under the guidance of a teacher.

When teaching people with disabilities distance learning will be implemented with the provision: electronic lecture notes; control tasks; test questions. Consultations of teachers and knowledge control are carried out by e-mail or via Skype.

The research and practical training is held at the end of the second semester for 6 weeks and is aimed at deepening and broadening of professional competencies (practical knowledge and skills), obtained in the study of subjects during the Bachelor level; strengthening them in practical training; maintaining the theoretical skills and practical experience during the semester; the formation of practical skills to acquire the ability to carry out an independent research. The practical research is carried out in the fourth semester for 12 weeks and is preceded by a pre-diploma practice.

The Master Research Practice is aimed at

- expanding and deepening the practical skills of conducting independent research activities;
- developing the professional competencies (knowledge, skills), which were obtained in the previous stages of studying on the base of graduate programs;
- providing targeted accumulation of material for writing the final qualifying work – Master's Thesis.

Pre-diploma practice is carried out in the fourth semester for 6 weeks before the State Final Examination. Pre-diploma practice complete the formation of general competence and research skills and is aimed at the development of the final qualifying work as a Master's Thesis on a high professional and scientific level.

All kinds of practices are carried out in the enterprises and organizations which are the partners of the SFedU,

which have signed an agreement on cooperation, as well as in the laboratories of the issuing department. The process of multidisciplinary collaboration between university departments guarantees the possibility of the program fulfillment in the field of the scientific and practical training and guiding the students during practice.

As a result of the development of graduate programs at the graduate level common cultural, general professional and specific professional competence in accordance with the direction of preparation 20.04.01 "HSE Management System" must be formed. The graduates who completed the program successfully are given the qualification "Master".

The types of professional activity of graduates can be used in the following fields: design and engineering; service and research; organization management; expert and supervisory spheres, and inspection and auditing activities.

As far as the employment of graduates are concerned they will be able to carry out their professional activities:

- in the industrial and service companies in various fields;
- in the design and scientific organizations and institutions that are developing technology and documentation to ensure technological and environmental safety departments;
- in the services and supervisory bodies of local, regional and federal levels;
- in the field of occupational safety and health, environmental and industrial safety.

The key positions of graduates after graduation are an environmental engineer, and a specialist in labor protection.

The Process of modernization of the system of Higher Education in the conditions of the modern world tendencies, which include information, communication, integration and globalization, maintain the consideration of competence-based approach in the system of professionally focused training of future specialists of Higher Education institutions. In this regard requirements to the quality of modern standards to national education have to be brought into accord with the international standards for ensuring the competitiveness of specialists in the international labor market.

Graduate "OSH" and the master's program "HSE Management System" students must have the following professional competences: the ability to optimize methods and techniques to ensure human security from the effects of various negative factors in technosphere; the ability to implement new methods of improving the reliability and stability of the technical facilities to maintain their functionality; the ability to navigate the full range of the scientific problems in the professional field; the ability to address issues of the sustainable and safe use of technical equipment; the ability to analyze and evaluate the potential danger of economic facilities for humans and the environment [6, 7].

A Safety and Health professional education requires multidisciplinary training of students. The "OSH" and "HSE Management System" professions need a broad-based educational background on the foundation of specialized competencies in physical sciences (medicine, physics, chemistry and engineering) and social sciences (behavior, motivation and communication) alongside with the principles and concepts of management.

CONCLUSION

The worked out programs are the best value for students to acquire the best safety training experience. The gained competencies will give the opportunity to reach high-career goals, be competent in the sphere of the Occupational Safety and Health field, and to develop the skills, needed to manage effectively safety in a workplace.

The rapid pace of technological change, combined with the persistence of unsafe or environmentally threatening working conditions, focuses the attention on the need to create a safe, healthy working environment and to promote a new safety culture in the workplace. Organizations increasingly seek OSH professionals with the right competences who can manage these processes.

That is why the programs "OSH", and "HSE Management System", are urgent ones to prepare qualified specialists of this type.

REFERENCES

[1] American Safety Council Online OSHA Training, June, 26, 2015. Retrieved from: http://www.americansafetycouncil.com/osha-hazard-training/index.aspx.

[2] Order of the Ministry of Labour of Russia, No 205, May 15, 2013. Retrieved from: http//www.rosmintrud.ru/docs/mintrud/orders/60.

[3] The decree of the Ministry of Labour of Russia Federation, No 37, August 21, 1998 (current version from February 12, 2014) Qualification handbook for managers, professionals and other employees" Retrieved from http://www.consultant.ru/popular/spravochnik-dolzhnostej/

[4] Bakaeva T.N, Dmitrieva I.A, Tolmachyova LV. Program Organization DPO "Occupational Safety" in the modern educational environment: challenges and prospects". Materials of of the International Forum Adult Continuing Education, St. Petersburg, 2015, pp. 129-132.

[5] Popova O.V. The Master program "Technology and environmental safety processes and production" in the direction of 20.04.01 Technosphere safety. Retrieved from http://sfedu.ru/pls/rsu/edu2015.obr_prog_doc?p_sell5_id=331

[6] Occupational Safety and Health Training. OHS Academy, June, 28, 2015. Retrieved from: http://www.oshatrain.org/pages/programs/132-hour-professional.html.

[7] Andrienko A.S., Bakaeva T.N., Popova O.V., Tolmacheva L.V. The Realization of programms: «Occupational safety and health (OSH)» and «Health, safety and environmental (HSE) management system» Application of Information and communication Technologies – AICT2015: Conference Proceedings, 9th International Conference, 14-16 October 2015, Rostov-on-Don, Russia, 2015. P. 511-515

[8] WordPress editor. Retrieved from: https://ru.wordpress.org/.

Course Scheduler and Recommendation System for Students

Suleyman Uslu, Can Ozturan
Dept. of Computer Engineering
Bogazici University
Istanbul, Turkey
{suleyman.uslu, ozturaca}@boun.edu.tr

Mehmet Fatih Uslu
Dept. of Industrial Engineering
Yıldız Technical University
Istanbul, Turkey
f0611303@std.yildiz.edu.tr

***Abstract*—Every semester, university students need to decide their course programs, and as the number of course offerings increase, it becomes challenge to form a schedule for the semester courses to be taken. In addition, students might have difficulty in finding courses that can match their schedules and course preferences. In order to address these problems, we build Mecanin, an online platform for Boğaziçi University students that provides an optimum schedule and course recommendation tools. In this paper, we propose an integer programming model to optimize course programs and also a recommendation algorithm to suggest courses to students. Both algorithms are tested and compared, and the results are presented.**

***Index Terms*—integer programming, collaborative filtering, course recommendation**

I. Introduction

Optimization and scheduling is an important aspect of many processes from timing work shifts to assigning jobs to machines. As the number of professors and courses increase, students start considering different courses and teachers for their elective courses, and this problem draws interest of researchers in the optimization field. Graduate or senior undergraduate students need to decide optional courses while preparing their course programs. There are many factors which affect the students' decision such as different times of courses, professor and course preferences, and conflicting hours of courses. These constraints become difficult to handle by hand or mind when the number of available courses reaches hundreds. In addition, students may not be aware of all the courses they can take. It also requires effort to find the best matching ones when there are many course alternatives.

For university students, we built an online platform, named Mecanin, for helping them to prepare their schedules considering their course and professor preferences as well as the time conflicts of courses. We developed an integer programming (IP) method to find the optimum schedule for students and made this tool available online as a part of Mecanin platform. We also provide another algorithm to help students find new courses which we think they might be interested in based on their and other students' course preferences. Proposed course recommendation algorithm is based on collaborative filtering (CF) and includes several adjustments that are specific to course program planning have been considered to improve the success of the algorithm. This online platform is used by Boğaziçi University students. The number of students that have used the system is between 700 and 1400 for the past 9 consecutive semesters from 2010 to 2014. Algorithms are tested on this real dataset considering confidentiality of the students.

The paper is organized as follows: In Section II, related work for both optimization and recommendation problem is presented. In Section III, we present the architecture and the usage of the Mecanin system in detail. In Section IV, integer programming formulation is presented for scheduling the student's courses by providing the objective function and constraints. In Section V, course recommendation algorithms are presented. Here, we describe the adaptation of CF method into course recommendation algorithm and show the different variations of the CF method. In Section VI, we first compare our simplex based method implementation with the well-known and widely used open source integer programming framework *Lpsolve* and present the results of both implementations. Later, we describe how the results of the recommendation algorithm are evaluated. We also show the results and compare the proposed methods with the basic techniques which are not based on similarity.

II. Previous Work

Integer programming is a widely used method to optimize processes. For example, there are several studies which use integer programming in arranging work shifts in contaminated area with radio-active or chemical materials and football league fixtures in some countries [1], [2]. Some publications focus on optimizing the time between final exams in universities [3]. More importantly, many researchers [4]-[10] has worked on finding an optimized annual academic schedule in educational institutions because it is considered as a difficult problem due to a large number of constraints such as limited number of rooms and day-hour preferences of professors. In another study [11], they try to solve a similar problem for high schools which is smaller in scale but tightly constrained in comparison with the scheduling problem in universities. Moreover, the work [12] concentrates on finding the schedule that leads students to finish their academic program in the shortest time possible. At the same time, they try to allow students to place their courses in specific days of the week and hours of days.

There are also studies proposing different approaches other than integer programming because of the high amount of time

and computational power required to solve high dimensional problems. For example, [13], [14] try to find the whole calendar of academic institutions using heuristic approaches such as genetic algorithm and particle swarm optimization. Santiago-Mozos et al. [15] proposes an approach using evolutionary algorithm to assign optimized schedules to all students in an institution. Their approach prioritizes the creation of the schedules of students who come first in their list.

For the recommendation problem, there are many studies about recommendation systems. Some well-known papers focus on movie datasets and use collaborative filtering and regression techniques [16], [17]. Furthermore, Shani et al. [18] develops an MDP-based recommendation algorithm for online bookstores.

Success of these recommendations motivated us to use the idea of these techniques to help students in their elective course selections. Vialardi et al. [19] proposes a course recommendation method considering demographics, course enrolment, number of courses per semester, average grades, and cumulative grades data of students. They used C4.5 algorithm and claimed 80% success rate. Ray et al. [20] proposes a technique for course recommendation based on the previous performance of students. They study 255 students and make item-based and user-based recommendations, and get 30% success rate. Booker [21] designs an online system to suggest courses for prospective students using k-means. Lee et al. [22] tries to find the association rules between students and courses using previous course preferences of students and course categories. They use curriculum experts to evaluate their recommendations. Sharon et al. [23] designs a recommendation system for elective courses using content-based approach considering the requirements of the job market. Bendakir et al. [24] develops a course recommendation system for students based on collaborative filtering and machine learning methods considering prerequisites and course evaluations.

In this paper, we propose a collaborative filtering technique to recommend courses to university students based on the student's and other students' preferences. We, then, adjust the algorithm considering course hour conflicts. We present the results of these approaches. Our results show that similarity based approaches like collaborative filtering are superior to other techniques like popularity based approaches. Our specialized collaborative filtering method and similarity function design give 80% correctness score using our evaluation methods.

III. Mecanin System Architecture

Mecanin is a social platform where university students can prepare their course programs and find courses that they might be interested in. It has two main functions: First, it provides a tool which helps students prepare their schedules. When users start typing the code or time of a course, or name of a teacher in the search box, the system brings the courses which have matching data with the query and their properties such as additional lab hours to the user. After adding all courses which the user has intended to add to his course program, system decides a program according to the ratings and conflict choice of the user. Second, Mecanin provides course recommendations based on current user's and other users' course preferences and since it is a social platform, a friendship mechanism is provided. Mecanin is implemented in Java, and NEO4J which is a graph based database is used as database management system.

A. Layout

After registering to the system, the first page, which is called profile, appears. In this page, users can see their current course program on a weekly timetable and course recommendations in a separate box. Second page is the editor page where the students prepare their programs. There are two timetables, one search box, and one course list on this page. First timetable is used to provide a better view of all the courses that the students intend to add to their programs, and the second timetable is the program that Mecanin suggests to the students.

B. Detailed Usage

Since all scheduling process is done on the edit page, it is necessary to select the courses first on this page. Users start by typing any property of the intended courses on the search box, and the system brings the courses that satisfy the information that is typed into the search box with its autocomplete functionality. Department, code, section, teacher, room, day and hour information could be retrieved from the database. As the users could establish multiple search, they can also select multiple courses from the results of one search.

After adding all intended courses, they appear on the left timetable as they do right below the search box as a list. If the conflict option is enabled, system chooses the program with all courses on the left and displays on the right timetable as the suggested course program. Disabling conflicts allows the system to run integer programming calculation and suggest the new course program. If there is a conflict among courses in a time slot, courses are shown and cell is displayed in red to take attention of the user.

Another option that users can benefit from is grouping. By default, there is one group of courses and all selected courses go into this group. Users are allowed to create additional groups and move courses between these groups. The advantage of the grouping mechanism is that it lets one to specify the minimum and the maximum number of courses in a group. For instance, a student might be required to take at least one and at most two of the 5 elective courses determined by his department in a semester in addition to his required courses. He can create a second group that consists of those 5 courses and specify the minimum and the maximum number of courses as 1 and 2 respectively. The system gives program suggestions considering the second group. The suggested programs include at least 1 and at most 2 of the courses from the second group.

Another criterion that the system considers while providing the optimum schedule to users is the importance of the added courses. This information is supplied by the user himself and specified as discrete numbers from 1 to 10. As the importance of a course increases, the system is more likely to put that course into the weekly schedule. Each course has an importance value of 5 by default and can be assigned importance values separately.

In addition to the scheduling options like grouping and importance, Mecanin is able to handle optional hours of a course. There could be lab hours or problem sessions assigned to a course which can be both optional or mandatory. Users can select these optional hours and add them to their course program.

C. Recommendation

Another feature of the Mecanin social platform is making course recommendations to its users. The algorithm used for this purpose is discussed in more detail in recommendation algorithm section. In general, proposed recommendation method is based on the currently selected courses of the user and course selections of other users which have common courses with the current user. Recommendations are given in the profile page.

D. Concepts

Every course having different abbreviation and code number in the university registration system is treated as a standalone course. For instance, CS 101 could be renamed as CE 101 for students of a different department having the same time slots, rooms and teachers but they are considered as different courses in Mecanin and a user can add both at the same time. However, their conflict will be taken into consideration while suggesting a schedule to the user.

In addition, courses can also have different sections as CS 101.1 and CS 101.2. They are also considered as different courses and users are allowed to add them to their programs. Since, for most of the cases, taking only one section of a course with many sections is sufficient, users are advised to create a separate group, put all different sections of a course into that group, and specify the selection limits as at least and at most one.

Apart from the regular course hours, some courses can have optional or mandatory additional hours such as laboratory work or problem sessions. These additional hours can be selected under the course options pane.

IV. Course Program Scheduling Algorithms

Mecanin system can suggest its users an optimum schedule according to their course preferences and ratings. Integer programming is used to solve this optimization problem and the Simplex method is made use of in its implementation. First, the IP model is explained and then the test system and the test cases are described. Results of the Mecanin scheduling algorithm is mentioned in the results section.

A. IP Model

IP model is described in detail by explaining decision variables, constants, objective function and the constraints of the IP formulation.

1) *Decision Variables:* Decision variables correspond to all the courses that students select and indicate the courses to be taken during the semester. There are n courses which are denoted as $[x_1,x_n]$. $x_i = 0$ means that course is not in the solution and $x_i = 1$ means i^{th} course is in the solution. A course being in the solution means that Mecanin includes that course in the suggested course program.

2) *Constants:* Constants are denoted as $[c_1,c_n]$ where n refers to the number of courses. These constants are weights that represent how much the student wants to take the corresponding course. In Mecanin scheduler, users are allowed to rate a course by giving a number between 1 and 10.

3) *Objective Function:* The objective of the IP model is to find the combination of courses forming an acceptable schedule with the highest score. Total score of a schedule is calculated by adding up the ratings of each course in the schedule, which are determined by the student. Therefore, the IP problem is a maximization problem:

$$maximize \ \sum_{i=1}^{n} x_i c_i \tag{1}$$

4) *Constraints:* There are 3 types of constraints in this model. First type is *default constraints*. All variables are either 0 or 1.

$$x_i \in \{0,1\} \tag{2}$$

Grouping constraints are used by solver to decide on the number of courses in the schedule. By default, there is one group; however, students are allowed to create additional groups. Total number of groups, which is represented by g, cannot exceed the total number of courses in the model.

$$1 \leq g \leq n \tag{3}$$

Minimum number of courses to be selected from the group i, say a_i, can be less than or equal to b_i. b_i is the maximum number of courses to be selected from the group i, and it is also less than or equal to t_i, the number of courses in the group. Combining the previous constraint with this one, a rule appears for each group.

$$1 \leq a_i \leq b_i \leq t_i \leq n \tag{4}$$

The *model constraint* is used to determine the upper and lower bounds of number of courses to be included in the schedule from a group. As mentioned previously, t_i denotes the number of courses in the group i, and the courses in the group can be denoted as $[x_m,x_{m+t}]$. Total number of courses to be selected from this group has to be equal to or greater than a_i and less than or equal to b_i, where a_i and bi denote the limitations of the group explained in the previous constraint.

$$a \leq \sum_{i=0}^{t-1} x_{m+i} \leq b \tag{5}$$

This constraint shows that Mecanin selects at least a_i and at most b_i courses from the i^{th} group to be included in the course program.

V. Course Recommendation Algorithms

Another feature of the Mecanin system is to provide its users course recommendations which are shown on the profile page. The users can add recommended courses to their programs using the recommended course links.

Mecanin recommendation engine has an algorithm based on collaborative filtering. Basically, the system recommends courses of users with similar tastes in terms of course selections to the current user. In Mecanin, similar taste is measured as the number of common courses. Recommendation engine first finds the courses of the user who will get recommendations. Then other users who selected at least one of these courses are retrieved, and then the courses of those users, other than the ones that current user already has, are selected from the database. This course list becomes the base list for recommendation. Since Neo4j, the graph based database, is used, this operation does not require table joins as in the conventional relational databases.

Mecanin recommendation engine has an internal ordering algorithm for ordering courses that are going to be recommended. For instance, suppose a current user, U1, who is to receive recommendations, has currently 5 courses, C1, C2, C3, C4, and C5, and there are 3 other users U2, U3 and U4. Suppose, U2 has C1 and C2 as a common course with U1 and has taken C6 additionally. Additionally, suppose also that U3 has both C3 as a common course with U1 and C7, C8, and C9. Lastly, U4 has C4 and C5 as common courses with U1 and C7 and C9 additionally. First of all, the system will select courses C6, C7, C8, and C9 as the base recommendation list. Since in the real world, the number of courses per user is higher than this example, and a course is selected by many users, this list becomes larger than the list in the example which has size of 4. To decide which courses to recommend from the base list, recommendation engine first assigns scores based on the total number of common courses of the current user with each user that have at least one common course.

Figure 1 shows the graph formed as a result of the aforementioned example. In this example, U2 and U4 have 2 courses and U3 has 1 common course with U1. After calculating scores for users, system calculates the scores for the courses in the base list. The score for C6 becomes 2 because it is only selected by U2 which has a score of 2. C7 and C9 would have a score of 3 because they are both selected by U3 and U4 which have scores of 1 and 2 respectively. Finally, the recommendation score for C8 would be 1 because of U3.

Our first recommendation algorithm, which is denoted as CF, sorts the courses in the base list based on their scores and recommends the first 5 or 10 courses according to the list size preference. A new variant of this algorithm, which is denoted as CF-0, takes time conflicts into account, eliminates those conflicting courses and does not include them in the recommended courses list. A conflict between two courses means that a student cannot attend all classes of both courses, he needs to make a choice between two courses. A third variant, which is denoted as CF-1, is similar to the second variant but it allows courses which has at most one-hour conflict with the current courses of the user to be recommended.

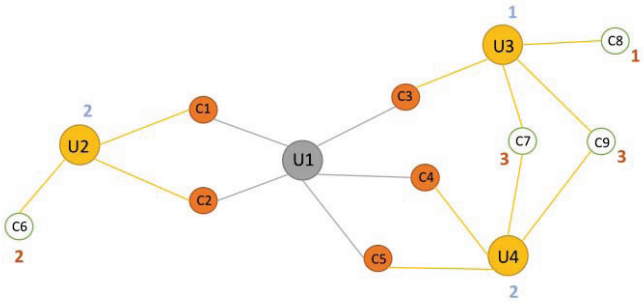


Fig. 1. Graph representation of course selections

VI. Tests and Results

A. Mecanin Scheduler

Several test cases are prepared to measure the effectiveness of the Mecanin scheduler. These cases differ in (i) the number of courses, (ii) whether they allow time conflicts between courses, (iii) course number constraints on groups as well as the number of groups. Cases are named according to these differences. Following is the format of case naming:

case<course number><conflict><group number>

The following is the name of the case having 15 courses, no conflict and 2 groups:

case15n2

The following is the name of the case having 20 courses with conflicts allowed and 1 group:

case20a1

Mecanin scheduler is compared with *Lpsolve,* the well-known open source integer programming solver which is implemented in ANSI C. The test system used is Early-2013 Macbook Pro with Intel Core i7 3.0 processor (3540M) and 8GB memory. Only the times for solving the IP problem is considered. Reading input files, parsing them or generating required classes are not included in the compared times.

To measure the speed of the solvers, all cases are run on both solvers. Results shown in the table is the run time results in milliseconds of running each case 1000 times.

As seen in Table I, Mecanin scheduler runs faster than *Lpsolve* up to 20 courses, but when the number of courses goes up, *Lpsolve* starts solving the linear problems faster. However, in the real life, a reasonable estimation could be that an average student probably does not attempt to add more than 20 desired courses to generate a schedule. Furthermore, Mecanin scheduler works faster when the courses are divided into groups. In other words, it works slower when all desired courses are put into one group and user does not create additional groups to inform the system further. Comparison of the two implementations is shown in the chart in Figure 2.

Additionally, Mecanin is designed as a web based social platform and Mecanin scheduler is implemented in javascript, but *Lpsolve* is written in ANSI C. Therefore, Mecanin scheduler could be faster compared to the current results when it is implemented in C.

B. Recommendation Engine

Mecanin recommendation engine is tested using data collected from Bogazici University students. Students' course statistics are gathered from the Soppus platform, which is the first attempt to build an online course-based social platform for university students [25]. Soppus provides users to prepare their schedule quickly but manually. Test is done on the data between the years 2010 and 2014. Due to the small of number of courses, users, and courses per user, summer semesters are disregarded and were not used during the test.

1) *Statistics:* In total, 9 semesters are used for testing the recommendation engine. All the semesters are either spring or fall semester. No summer semester data is used. The first used semester is Spring 2010 and the last one is Spring 2014. Table

TABLE I. SOLUTION TIMES IN MILLISECONDS

Cases	Mecanin	Lpsolve
Case5a1	26	346
Case5a2	27	257
Case5n1	27	325
Case5n2	27	334
Case10a1	54	367
Case10a2	70	333
Case10n1	87	340
Case10n2	81	378
Case15a1	111	219
Case15a2	82	261
Case15n1	151	377
Case15n2	119	283
Case20a1	219	363
Case20a2	122	379
Case20n1	322	371
Case20n2	197	300
Case25a1	339	395
Case25a2	189	454
Case25n1	493	496
Case25n2	321	355

II shows various statistics of the data that is used. Note also that total number of course selections (of all 9 semesters) is a little over 50K which is also the total number of course recommendations that is made for each algorithm.

2) *Test Methods:* As described in the recommendation algorithms section, main and modified versions of Mecanin recommendation engine are tested. In addition, results of random course recommendation and recommendation based on popularity of courses are also provided. The test method is based on removing an already selected course of a user and trying to estimate the course based on the remaining courses of the user and the other users who have common courses with the current user. Success of a recommendation is based on finding a course that is intentionally dropped for testing. If the course is found, corresponding recommendation is counted as successful. In total, the success of the algorithm is measured as a fraction of course selections that is successfully estimated. Tests are performed using variable size of recommendation lists.

3) *Test Results:* RND is random course recommendation algorithm. RWH is an algorithm in which courses are put in a roulette wheel according to their number of selections by users and courses are recommended according to their probability of being selected. POP sorts the courses according to their number of selections and recommends first n courses. CF is the novel algorithm developed based on collaborative filtering as described in Section V section. CF-n is the modified version of CF where courses causing conflicts are eliminated. Here, n is the maximum number of total allowed conflicts in the program. In this paper, only CF-0 and CF-1 variants are considered.

TABLE II. SOLUTION TIMES IN MILLISECONDS

Semester Statistics	Min	Max	Avg
Active courses	776	1009	1203
Users	438	898	1448
Users with more than 1 course	423	484	1428
Courses selected by users	5.9	6.3	6.4
Total course selections	2576	5639	9291

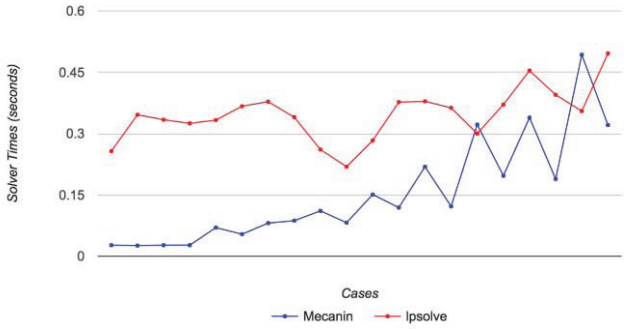


Fig. 2. Comparison of solving times

Results are shown in Table III. First, the scores for the non-similarity based algorithms are given. RND, POP, and RWH algorithms are tested using the list sizes 5 and 10, whereas CF, CF-0, and CF-1 are tested using the list sizes 1 to 4 in addition to 5 and 10.

TABLE III. COMPARISON OF RECOMMENDATION ALGORITHMS

	Size of the recommended courses list					
Methods	**1**	**2**	**3**	**4**	**5**	**6**
RND	-	-	-	-	0.031	0.062
POP	-	-	-	-	0.260	0.350
RWH	-	-	-	-	0.076	0.140
CF	0.610	0.706	0.748	0.783	0.807	0.874
CF-0	0.613	0.706	0.750	0.784	0.808	0.875
CF-1	0.612	0.707	0.810	0.810	0.810	0.876

VII. Conclusion

In this study, an integer programming based solver is proposed to provide university students optimized course programs based on their ratings on courses and course timing conflicts. An online web platform, Mecanin, is built for the use of this optimum schedule tool by Boğaziçi University students. It is observed that this tool which is implemented in javascript is sufficiently responsive to suggest a schedule for students based on their course ratings and time conflicts of courses.

Additionally, recommendation algorithm based on CF method is proposed which considers the current courses of the user and the course preferences of other users which may be similar to the user in term of course programs. Several variants of this algorithm is also provided, and the results of these proposed algorithms and basic methods which are not based on similarity techniques are compared. It is observed that CF, which is a similarity based method, outperforms all other methods. In addition, it is shown that by considering the conflicts, the score of the CF method could be improved.

References

[1] T. Sawik, "An integer programming approach to scheduling in a contaminated area," *Omega*, vol. 38, no. 3, pp. 179–191, 2010.

[2] D. Recalde, R. Torres, and P. Vaca, "Scheduling the professional Ecuadorian football league by integer programming," *Computers & Operations Research*, vol. 40, no. 10, pp. 2478–2484, 2013.

[3] S. MirHassani, "Improving paper spread in examination timetables using integer programming," *Applied Mathematics and Computation*, vol. 179, no. 2, pp. 702–706, 2006.

[4] J. Fiallos, "Integer programming model for course scheduling at Unitec," in *IIE Annual Conference. Proceedings*. Institute of Industrial Engineers-Publisher, 2009, p. 2048.

[5] S. Daskalaki and T. Birbas, "Efficient solutions for a university timetabling problem through integer programming," *European Journal of Operational Research*, vol. 160, no. 1, pp. 106–120, 2005.

[6] S. MirHassani, "A computational approach to enhancing course timetabling with integer programming," *Applied Mathematics and Computation*, vol. 175, no. 1, pp. 814–822, 2006.

[7] M. Dimopoulou and P. Miliotis, "Implementation of a university course and examination timetabling system," *European Journal of Operational Research*, vol. 130, no. 1, pp. 202–213, 2001.

[8] S. Daskalaki, T. Birbas, and E. Housos, "An integer programming formulation for a case study in university timetabling," *European Journal of Operational Research*, vol. 153, no. 1, pp. 117–135, 2004.

[9] R. J. Acosta-Amado and M. Villa-Marulanda, "An integer programming model for the academic timetabling problem," in *IIE Annual Conference. Proceedings*. Institute of Industrial Engineers-Publisher, 2013, p. 2712.

[10] M. A. Bakir and C. Aksop, "A 0-1 integer programming approach to a university timetabling problem," *Hacettepe Journal of Mathematics and Statistics*, vol. 37, no. 1, 2008.

[11] N. Boland, B. D. Hughes, L. T. Merlot, and P. J. Stuckey, "New integer linear programming approaches for course timetabling," *Computers & Operations Research*, vol. 35, no. 7, pp. 2209–2233, 2008.

[12] M. H. Patel, H. J. Tsao, and Z. J. Abdollahi, "Integer programming for student program design and scheduling," in *IIE Annual Conference. Proceedings*. Institute of Industrial Engineers-Publisher, 2011, p. 1.

[13] Y.-Z. Wang, "Using genetic algorithm methods to solve course scheduling problems," *Expert Systems with Applications*, vol. 25, no. 1, pp. 39–50, 2003.

[14] D.-F. Shiau, "A hybrid particle swarm optimization for a university course scheduling problem with flexible preferences," *Expert Systems with Applications*, vol. 38, no. 1, pp. 235–248, 2011.

[15] R. Santiago-Mozos, S. Salcedo-Sanz, M. DePrado-Cumplido, and C. Bousoño-Calzón, "A two-phase heuristic evolutionary algorithm for personalizing course timetables: a case study in a Spanish university," *Computers & operations research*, vol. 32, no. 7, pp. 1761–1776, 2005.

[16] A. Ansari, S. Essegaier, and R. Kohli, "Internet recommendation systems," *Journal of Marketing research*, vol. 37, no. 3, pp. 363–375, 2000.

[17] A. Mild and M. Natter, "Collaborative filtering or regression models for internet recommendation systems?" *Journal of Targeting, Measurement and Analysis for marketing,* vol. 10, no. 4, pp. 304–313, 2002.

[18] G. Shani, R. I. Brafman, and D. Heckerman, "An mdp-based recommender system," in *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 453–460.

[19] C. Vialardi, J. Bravo Agapito, L. S. Shafti, and A. Ortigosa, *Recommendation in higher education using data mining techniques.* Barnes, T., Desmarais, M., Romero, C., & Ventura, S., 2009.

[20] S. Ray and A. Sharma, "A collaborative filtering based approach for recommending elective courses," in *Information Intelligence, Systems, Technology and Management.* Springer, 2011, pp. 330–339.

[21] Q. E. Booker, "A student program recommendation system prototype."

[22] Y. Lee and J. Cho, "An intelligent course recommendation system," *SmartCR*, vol. 1, no. 1, pp. 69–84, 2011.

[23] Y.-J. L. Sharon and K.-W. Lien, "Careermaker: An adaptive elective courses recommendation system."

[24] N. Bendakir and E. Aïmeur, "Using association rules for course recommendation," in *Proceedings of the AAAI Workshop on Educational Data Mining,* vol. 3, 2006.

[25] E. Sadiyev, M. Akboyun, and S. Uslu, "Course schedule planner," 2010, senior project, Dept. of Computer Engineering, Boğaziçi Universit

Rewriting, academic fraud and falsification: analyzing cases of academic practice in St. Petersburg, Russia

Rekhtina L.S., Pankratova L.S., Sokolov N.V.
Saint Petersburg State University
Saint-Petersburg, Russia
lisabet-09@mail.ru, l.s.pankratova@spbu.ru,
nvsspb@yandex.ru

Karimov R.N.
Institute of Geography of National Academy of Sciences
Baku, Azerbaijan
rovshan_karimov@yahoo.com

Samedov F.R, Yunisov J.K.
Research Institute of Aerospace Information
Baku, Azerbaijan
farid.samedov@mail.ru, cyunisov@yandex.ru

***Abstract*—The article discusses phenomenon of plagiarism forms in academic end education practice and its detection and reaction system.**

***Index Terms*—rewriting, academic fraud, falsification higher education.**

I. Introduction

Information technologies are an essential part of the education system. E-learning is gradually developed and implemented in education process, not only as a conveyor of knowledge and information, but also as a communication buffer between faculty, administration, and students. This system should save resources and optimize learning process, reduce time budget of education agents and formalize communication between them. The article considers some complex issues in the work of certain elements of the implemented E-learning system and science.

Due to the overproduction of texts, as well as the demand to quickly produce new texts, the probability of plagiarism in a paper increases. A wide spread of the Internet, information technologies, and communication tools give to a vast audience access to a huge amount of information, which also increases the probability of textual borrowing in academic papers. Thus, according to various studies, the percentage of plagiarism in student papers increased from 23% to 90% during the period from 1941 to 2001 [1]. For example, the number of students in Russia from 1940 to 2015 has increased from 478,000 [2] to 14, 191,000 people [3].

Scientists, participants and managers of major research projects, supervisors and even students are facing the need to check texts for plagiarism, as they are no longer confident in their own texts [4]. Previously, textual borrowing was possible to find out only by manually comparing the two texts - the source and its replication. The information society offers a wide range of tools to facilitate this task. But the attitude to plagiarism has become more severe and serious. Simple, even automatic, comparison is no longer sufficient to define plagiarism in a text. There is a need for a serious, professional expertise of each individual case of textual borrowing to identify whether it is plagiarism or not. Increasing proportion of plagiarism toughens and complicates the procedure of inspection, and that requires more costs. Requirements for texts and their expertise are considerably growing. Now it is necessary not only to compare two texts, but also to conduct semantic analysis of the papers. That is poorly automated and requires manual comparison and the work of experts.

II. Research methods

This article is based on the thorough analysis of 9 typical cases of plagiarism detection in student papers (there are more than 30 cases overall) in Saint Petersburg, Russia. These are papers submitted by the students of the final year of undergraduate program in Sociology on their major courses.

First of all, these cases illustrate different examples when the work of one student is stolen by the others and presented as their own paper. In 2011 there was a case when the successful work of one student was brought in a week three times by his/her co-students with minimal changes. But lately, cases of complete repeat of someone's work are rare. As a rule, the original text of the first paper is edited: parts of the text are put in different order, and words are replaced by synonyms. Among such cases the compilation of several works is a usual practice. These texts may belong to their co-students or senior students who studied earlier.

Primary this applies to text assignments (course programs and analytical reports) and practical tasks.

Plagiarism from the Internet sources (total or partial borrowing and their compilation) is almost a global problem. Such kind of plagiarism could be found in every second job. Most often it is detected in the section of a paper that presents methodological and theoretical bases of the work. This is a common case, but quite problematic for a check as plenty of materials are used from sources that are not indexed (social media, books or journals from private, closed databases or archives).

We consider separately cases of practical works downloaded from the Internet or fabricated, as there are no any plagiarism detection technologies for such papers. They are usually checked manually and it is effective only if part of the material is in the public domain.

III. RESULTS

Rewriting is a natural component of the information explosion. It allows enormously increase the amount of information while reducing its cost and resource intensity [5]. The academic environment does not involve the use of rewriting as a source of new original texts. Rewriting is a skill that is necessary for abstracting of scientific sources, but it is not allowed in the final, original (especially research) texts. In the situation of "the tyranny of the moment" a student dwells on rewriting rather than trying to continue to work creatively with information from the initial source. This "tyranny of the moment" [6]. is felt not only by students who do not have time (motivation, desire) to work thoroughly with the original paper, but also by teachers (lecturers) who do not have time to educate, develop skills of independent, critical and creative writing in the class. Moreover, students are already immersed in an environment where rewriting is a commonplace. For this generation it is obvious that many of the tasks do not require the creation of a new text, it is enough just compiling it from materials previously created by other authors.

Plagiarism detection of students' papers gradually becomes a mandatory procedure in the educational practice in the post-soviet space. In Russia a lot of universities have introduced requirements for the verification of the final qualifying papers for plagiarism. In Azerbaijan in 2016 began the development of a unified system for plagiarism checks of the texts in the spheres of education and science [7].

The main specialized system of plagiarism detection recognized by the Russian scientific and academic community is called "antiplagiat.ru". It appeared in year 2007 and is still in the process of development. This is a basic tool to check texts. This system (a free version) allows you to create a personal profile where it is possible to download text files, and then the system searches for a match with its database of papers, educational materials, and public sites. These systems hide the algorithm of the work, but presumably the "antiplagiat.ru" system uses "w-shingling based on N-grams with an additional normalization of texts in accordance with dictionary of synonyms, and finding of plagiarism locally has a trigger effect to protect against false positives" [8].

In several Russian universities are introduced e-learning and internal communication systems right now. In Saint Petersburg State University, it is the system named Blackboard (https://bb.spbu.ru), which consists of three platforms - Blackboard Course Delivery, Blackboard Content Management, Blackboard Community Engagement. To check downloaded texts Blackboard uses such component as SafeAssign. The Higher School of Economics University uses the LMS system (Learning Management System - http: //lms.hse.ru) which allows to download texts in an extended version of the "antiplagiat.ru" service. Modification and enhancement of applications comprising these systems still require a lot of time. Their users constantly address to support division and are forced to bypass the limitations of the systems in order to complete their professional tasks. For instance, key problems of the Blackboard are following: basic functionality, not enough flexibility, clunky interface, technical difficulty or inability to organize work in groups/in pairs, sensitivity to the bibliographic data and common phrase.

As practice shows, the automatic plagiarism detection of papers is not enough. On the first stage of the check all works were tested by "antiplagiat.ru" system which is the leading one among existing in Russian. Such services warn that they are looking for the plagiarism, meaning matches in the text. But that is not enough for plagiarism detection to know what percentage of matches were found by the program. It is essential to understand what is behind found textual borrowing. This may be a list of references, properly executed quotes, set expressions and common phrases that simply cannot be formulated otherwise. That means that a match found by automated tools may be not plagiarism. It is important to explain this nuance to students. Each year students raise the question about the acceptable percentage of plagiarism in the work. Whilst according to the documents they signed before their admission to the university state that writing works (in accordance with curriculum) prepared by student must not contain plagiarism. Works may contain a certain percentage of such textual borrowing, as quotations and illustrations if they are correctly formatted. Apparently it is up to 20% of textual borrowing that is allowed. Experience has shown that level of understanding of the differences of these phenomena among students is very law.

So, on the second stage there is a necessity to carry out further check based on the results provided by automated services. If the software allows you to view a report, you need to go through it and see each case matches. Internet is a very mobile system and in the case of the anti-plagiarism report the links, which should show the alleged source, often do not work. Then teachers have to carry out manual checks of the matches, using public search engines (like "Google", "Yandex"). This procedure requires form a lecturer time and resources, research skills and

knowledge of different software systems that allow fixing the fact of plagiarism.

Thus, the first level of such "forced" plagiarism is a rewrite without references. It seems to be a natural reaction of the generation to a task that is considered daunting and excessive. (That is the result of "the tyranny of the moment" when resources are limited.)

Let us consider the example of a good student paper containing a rewriting without causing any damage to its content. It is a work of group of students which is, according to the "antiplagiat.ru" service, original on 96.43%, meaning the amount of borrowing in the text is 3.57%. The work regarded issues of identifying and detecting signs of social identity in everyday life. In this study there were used valuable and correct scientific sources expanding and enriching the conceptual and basis of the paper. And the references were listed. The problem was in the lack of quotation marks ("_") on the both sides of the borrowed text. The necessity to use quotation marks was explained to the students and they solved the problem by adding them shortly afterwards. (Although there were cases when other students' teams could not fulfill this requirement because they did not manage to detect borrowing text in their own paper and arrange everything properly.) This case can be seen as a rewriting which allows students to complete the task successfully. In fact, while preparing a research proposal there is absolutely no need to create new theoretical text, since it will be written worse by students then in the original version used by them. But academic rules exclude the possibility for students to use only someone else's text. Students are required to complete a very intensive, consuming a lot of resources tasks, meaning to prepare their own academic products under the conditions of resource shortage [9] and integration in the environment where there are obvious easy solutions in the form of rewriting, for example.

Interestingly this paper was also submitted by other students from the same course as their own original work. The students who tried to cheat submitted this paper before the real authors. Actually it was one of the previous (unfinished) versions of the final paper that was given to them by real authors as an example of the task. This case, as well as others, shows that the students do not have a sense of the value of the original author's work and their own creative work.

The second level of academic dishonesty is an attempt to create academic papers from other texts trying to imitate the scientific activities. Thus, the fabrication and falsification of meanings of academic activities happen. If rewriting does not lead to loss of previously created meanings, such compilation of texts leads to destruction of these meanings. Unfortunately, rewriting often transforms into plagiarism when there is not only absence of quotation marks of a text which is appropriate to be used, but instead a consciously created paper from fragments of other texts. An example of such case is a student's essay (8 000 characters) devoted to the problem of homeless children. According to the "antiplagiat.ru" service the originality of this text is just 9.88%, everything else is different pieces of other texts. At the second stage of plagiarism detection it turned out that this essay is an example of dual plagiarism. The essay was a downloaded text of student paper from the Internet, which is in turn, was a compilation of parts of texts written by other authors ("patchwork text"). There were only a few references to the sources used. This is an example of academic fraud [10]. But a passing glance given by a teacher at the paper may not raise any questions as the text has logic, and text pieces fit together, and the problem of homelessness and research approaches to it are well presented.

The third example illustrates a case where a set of "pieces" of a text are combined chaotically to form a meaningless, false and sophisticate text. The task was to prepare a research design to study some social issues of migration. The originality of the text according to the "antiplagiat.ru" service is 67%. The second stage of the check has shown that the submitted paper consists of different pieces of various texts mixed together. There were some references in the paper which were not related to the borrowing from the source text. The submitted paper contained references that were falsified. Drawbacks of this paper: borrowed pieces in the text are copied with misprints and errors; the text is replete with trivial phrases (at the level of common sense). There were selected most common and expatiative arguments for the compilation. They do not indicate a problem, do not justify the relevance of the work, and do not allow justify the sequence of researcher's activities. The extracts from statistical compendiums are given without any reference to them. Interestingly, the manual search with the help of "Yandex" (search engine) showed that almost the whole paper consists of pieces from other texts. A high percentage of originality shown by the "antiplagiat.ru" service could be explained by the fact that the database of this service did not contain some original sources.

The third level is the partial or complete falsification of the academic process and academic product. It is criminal activity that must be stopped in the early stages of education, as it can completely discredit and devalue the whole scientific spheres. Below are presented examples related to the problem of falsification of papers/practical exercises by students. In one case, the students falsified the interview – they role played the interview which was published in "LiveJournal" (service for journals and blogs). Another example of falsification of the same task: some tracks that were not recorded during the interview were "glued" in it. It was possible to identify this because of the difference in recording quality, specificity of students' conversation style. It was clearly heard that the text is read on the recording and there are places where the text was "glued". This is very time consuming practice. Falsification of the interview takes more time and resources than the fulfillment of the task (to conduct an interview). Another instance is when the interview was completely borrowed from the article of the Russian internet media "BaltInfo.ru". Students changed only names of the interviewee and the city of origin.

If rewriting and academic fraud can be checked with the help of technical means, falsification of practical exercise may be identified only by manual comparison of two files. But this requires creation of a database that contains all students' papers that could be compared with each other. In Russian universities there have been already introduced the system of communication between teachers and students which should solve this problem. It is a social network for internal use at the university supplemented with special software. Students submit all their papers only through this system, and lecturers check them their (upload reviews and comments). Thus, communication between students and lecturers become more formalized; implemented more strict time management - if a submission deadline has passed, a student will not be able to download a paper; formed an internal database of all student papers. When the same paper is uploaded into the database, the system should automatically identify and report such cases. But in fact, such databases are just in the process of formation at the moment. The database check file attributes and, so far, is not good in identifying repeating parts of the file or compilation within it. Internal systems of plagiarism detection have not proved to be effective yet. They create an additional burden on a lecturer for its maintenance, and thus taking his/her attention from checking of students' papers. Another effect of such systems concerns bibliographical references which are seen by the system as borrowings. Thus, bibliographical references increase the total percentage of plagiarism detected by the system. Students are determined to minimize the percentage of "plagiarism" by reducing a list of bibliographic references, as they do not understand the difference between plagiarism and borrowing from a text. There is a major controversy - the system is designed to check and teach students how to work with texts, to follow copyright laws and to save hyperlinked apparatus of science, but in fact it achieves the opposite effect as it destroys the habit of students to fix all the sources used. Another alternative was introduced in Azerbaijan in 2016. It is a development and implementation of a general system of plagiarism detection in texts for universities and research centers. Yet it is too early to evaluate its effectiveness [11]. Such idea has its benefits, as texts and papers can be spread between students not only from different courses, but also universities.

IV. Conclusion

Informational and software technologies for borrowing detection in papers designed to maintain and preserve the norms and rules of academic work with texts become more complex. The development of these systems requires very serious resources, as well as development of their interaction with other information systems (libraries, search engines). The development and refinement of these systems, and expertise in case of serious precedents requires work of a number of specialists: programmers who are capable of implementing the sophisticated algorithms, lawyers that deal with the legal aspects, marketing consultants [12], sociologists, economists, and certainly linguists that are able to assess and understand the features of word usage and carry out semantic analysis of the text [13]. Complexity of these systems is already very high and their potential is not sufficient to eliminate the problem of manual plagiarism detection of the texts and difficult professional expertise. Rewriting technologies, plagiarism and falsification of academic papers are ahead of the development of these systems. When the "antiplagiat.ru" service was released a counter service "antiplagiat-killer.ru" appeared. That is a kind of arms race. New communication technologies have created endless amounts of information that are available to a large number of consumers. That requires constant replenishment of this information. Rewriting is an answer to this demand. The need to link new texts with original sources and to suppress theft of papers makes it necessary to develop new systems of sophisticated search and detection of borrowing. And they, in turn, force to look for new ways to "create" materials and/or opportunities to bypass the check. The dialectics of this process makes to constantly increase the resource intensity of these systems, thus distracting resources from the main goal - teaching students the norms of academic writing and development of their writing skills. The era when a student brought someone else's work ends with the emergence of the opportunity to compare student papers with papers from previous years. But there are technologies that are beginning to produce texts from materials that are not indexed by databases of plagiarism detection systems.

The practices of plagiarism are dangerous, because they destroy the frame of hyperlinking in science. But total detection of plagiarism and rigid copyright protection leads to the situation when information is "locked" in the narrow frames of citations. And that is even more dangerous and harmful for science. This anti-plagiarism war shadows the main purpose of the educational process. Students are no longer taught how to write a text or to conduct an interview, no one cares about that. What is important is whether a student is caught or not caught plagiarizing. And not to be caught plagiarizing is enough for a successful continuation of a career.

V. Discussion

Plagiarism in academic texts is a social phenomenon which is a reaction of the actors of educational process to the spread of modern information technologies. Problem of plagiarism occurs in the following cases: 1) students do not understand the boundaries of legal and illegal usage of someone else's information product, 2) students do not understand the purposes of educational tasks assigned to them or believe these purposes are irrelevant, 3) students believe that rewriting technologies, fraud and falsifications are reliable and safe. But plagiarism is widely distributed in student texts also because of lecturers, meaning that: 1) lecturers are not interested in the process of creating a student text and control only the result, 2) lecturers give tasks that are not relevant to the modern level of development of information technologies and provoke plagiarism, 3) lecturers do not provide the declared level of control of student papers and texts. There is third party responsible for the problem of plagiarism it is academic

management which: 1) reduces the resources needed for the effective and close cooperation between lecturers and students, 2) formalizes the academic interaction, 3) tends to massification and simplification of educational processes. Plagiarism is the result of economy of resource from three sides - students, lecturers and management. The social problem is the latent criminal convention of these three groups on the admissibility of plagiarism in academic texts within the margin error of informational and technological tools that are used to monitor it at any given moment. But these tools do not solve the problem, but only reduce amount of attention paid to primary tasks of education and science. Moreover, they create alternative problem of artificial restrictions on the free dissemination of information that is necessary for the academic environment. Only the transition to social and information tools will allow moving the focus of control from overprotection of information to the motivation of its operators. The aim is to reduce benefits that come from the exploitation of someone else's intellectual labor. Solution of the plagiarism problem is in the field of social technologies, meaning reducing the benefits of using fraud and falsification in academic sphere.

The key point is the redistribution of educational procedures in accordance with the modern status of the student and correction of academic status itself. For centuries, students have created academic texts in order to achieve the participation in the elite club of intellectuals that are free from heavy physical labor for the sake of doing science. Modern society is close to the total displacement of manual labor by intellectual, but the latter has completely lost elitism. Overcoming the lack of information allows us to solve most of the intellectual problems without creating an original text. This means that the competent rewriting that implies respect for the law and science ethics, as well as the preservation of the original source of meaning, and provision of access to it via hyperlinks, is the necessary and sufficient level of competence for a large part of students, future professional practice of which does not imply intellectual creativity. Creative education demands significantly more resources than the one dominant today in the mass market. Teaching to write a text is much more expensive than doing rewriting. Therefore, social solution of the plagiarism problem implies restoring of the elite status of the creative intellectual labor and education.

VI. REFERENCES

[1] L. A. Jensen, J. J. Arnett, S. S. Feldman & E. Cauffman. "It's wrong, but everybody does it: Academic dishonesty among high school and college students", Contemporary Educational Psychology, 27, pp. 210, 2002.

[2] G.V. Balashov. "The scale of higher education in Russia", Economics of education, issue 2, pp. 125-130, 2014.

[3] "Russia 2015" Statistical Handbook URL: http://www.gks.ru/free_doc/doc_2015/rus15.pdf (the date of access 30/05/2016)

[4] N. S. Nikiforov, S. V. Lavrinenko. "Tests in education and their application in the electronic environment", Application of Information and Communication Technologies (AICT), 2015 9th International Conference on 14-16 Oct. 2015, pp. 482 – 485.

[5] R. Hilliard. "Rewriting" in Writing for Television, Radio, and New Media, Cengage Learning, 2014, pp.176-178.

[6] T. H. Eriksen. Tyranny of the moment. Fast and slow time in the information age, London: Pluto, 2001.

[7] Azerbaijan intensifies combating plagiarism in education. Veliev. G. Baku network: International policy and security // URL: http://www.trend.az/azerbaijan/society/2516303.html (the date of access 30/05/2016)

[8] E.S. Chirkin. Use of Antiplagiat in education Vestnik Tambov University. Series: Natural and Technical Sciences Issue number 6 (2) / volume 18/2013

[9] L.S. Rekhtina, N.V. Sokolov. "A one-dimensional man in Gutenberg galaxy: the professionalization of the young scientist", 11 Materials of the international scientific-practical conference "Modern scientific potential" - February 28 - March 7, 2015.

[10] M.A. Eckstein. Combating academic fraud: Toward a culture of integrity. International Institute for Educational Planning. UNESCO, 2003, p. 101.

[11] E. Ostapenko. "Azerbaijan begins to deal with academic plagiarism", URL: http://1news. az/society/20140611052049665.html (the date of access 03/06/2016).

[12] J. Bailey. "Why Marketers Should Care About Copyright and Plagiarism. Plagiarism Today", URL: https://www.plagiarismtoday.com/2012/09/13/why-marketers-should-care-about-copyright-plagiarism/ (the date of access 05/06/2016).

[13] A. Kravchenko,. Vl. Kureichik, D. Zaporozhets. "Translation of natural language queries to structured data sources", Application of Information and Communication Technologies (AICT), 2015 9th International Conference on 14-16 Oct. 2015. P 57-5

MATLAB and Heat Transfer in Solid State

Martin Vozár
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
mvozar@ukf.sk

Peter Kuna
Department of Informatics
Constantine the Philosopher University in Nitra
Nitra, Slovakia
mvozar@ukf.sk

***Abstract*—The article deals with the problems of heat transfer through solids. The calculation of the thermal resistance of the system consisting of several layers of solid materials having different thermal resistances on the intensive calculations and the time. In the first part we law of calorimetry equation. On this basis, we can build a theory of one-dimensional and multi-dimensional differential equation of thermal diffusion. The derived equations are then used to simulate the heat transfer in building applications. We had to adjust the thermal diffusion equation in such a shape that it can be addressed by using PDE toolbox.**

***Index Terms*— MATLAB, Simulation, Heat Transfer, Partial Differential Equation.**

I. Introduction

In the article, the authors deal with the issue of heat transfer and simulation capabilities in the process. The focus is on heat transfer in solids - heat conduction (thermal conduction). Theoretical findings are summarized in the form of formulas serving the mathematical expression of specific factors and phenomena in heat transfer. The article also describes the procedure for adjusting a combination of the laws of physics, which was derived probably the most important known formula of the issue - partial differential equation of thermal diffusion.

On the theoretical basis of thermodynamics are the steps to creating a simulation of heat transfer in solids. For the development was used MATLAB interactive environment usable for various purposes such as numerical computation, data visualization itself, or application programming with high-quality user interface. The authors describe the logic of computer code applications, as well as its own user interface for defining multiple layers of different materials through which the heat is transferred.

II. Related Works

A. Thermodynamics

The term heat means a quantity respectively portion of the internal energy that does not have the nature of work or cloth transmission. This is the energy which system exchanges (received / transmitted) when dealing with systems of different temperatures. Then we talk about heat exchange or heat transfer. If between two bodies having different temperatures, there is a physical link permeable, heat will flow from the warmer to the colder body.

Heat transfer is pervasive and extends to some degree in virtually all. The heat is continuously transferring from our blood stream through a layer of connective tissue to the skin surface and from there to the ambient atmosphere. Heated air then flows from our bodies and change the temperature of the room in which we find ourselves. Even if you leave the room, the air in it will continue to flow slightly, because the walls of the room are never completely insulating. Similar processes occur in all plant and animal world, and in the air around us. They are taking place across the Earth, whose core is hot and cooler surface.

The driving force behind these thermal processes is cooling (equalizing) of temperature gradients in space. The sun is the main sources of the heat flow in our world - conduction of heat from the Earth's core and radiation from distant stars is in terrestrial natural processes secondary.

Thermodynamics (also thermals) is the branch of physics dealing with the heat and temperature and their relation to energy and labor. Examines and describes patterns of heat and thermal processes, defines parameters characterizing the macroscopic thermal systems (internal energy, entropy and pressure) and changes in these variables over of physical phenomena related to the exchange of heat between the system and surroundings [1]. It also sets out the general laws which these variables are subject, and which do not differ according to the attributes of a particular material but are common for all materials. These laws are expressed in four laws of thermodynamics, which will be described in the next chapter.

There are four laws of thermodynamics defining the basic physical quantities of thermodynamic systems (temperature, energy, entropy). These laws describe the behavior of these quantities in various circumstances and conditions, and exclude the existence of certain phenomena (eg. a perpetual motion machine) too.

The zero law of thermodynamics states that if any two thermodynamic systems in thermal equilibrium with a third system are also in thermal equilibrium with each other as well. Between these two systems, there is thermal equilibrium, where they are connected by a wall that is permeable only by heat, and these systems do not change over time. Two systems, however, may be in thermal equilibrium, even if not connected, and therefore, heat transfer is not possible, but in their case we

can say with certainty that the thermal exchange did not go, or if it were linked.

Most precise representation of zero-law is thinking of thermal equilibrium as the session equivalence between pairs of thermodynamic systems. "When the body A is in thermal equilibrium with the body B and body C, then units B and C to each other in the thermal equilibrium." [2].

III. Heat and Calorimetric Equation

The heat transfer can take different forms and processes. This process will be address in the next subsection. We also make the mathematical model used in the calculations related to the transmission of heat as well as a model equation which is used in building applications.

Calorimetric equation is elemental equation describing the heat exchange on the basis of the law of conservation of energy. On this basis it is possible to determine the specific heat capacity of materials, determine the amount of heat to be supplied or withdrawn to changing the temperature of a certain value, or determine the final temperature during the heat exchange. The amount of heat that the body or come forward depends on the mass of a body m, specific heat capacity c, and the difference of the initial and final temperatures t. Of equal received and transmitted of heat in the system of bodies, that we consider ideal isolated, causes the resulting temperature relationship

$$t = \frac{c_1 m_1 t_1 - c_2 m_2 t_2}{c_1 m_1 + c_2 m_2} \tag{1}$$

where body mass m1, specific heat capacity c1 and t1 is in thermal contact with body weight m2 heat capacity c2 and t2; t is the temperature reached by the system of bodies after a certain time.

By the heat transfer material of the particles with a higher temperature transmit part of its internal energy by mutual collision particles in places with a lower temperature, ie. particles with lower internal energy. During this process, the particles are not moved, but vibrate around their equilibrium positions. The propagation of heat through the line most often we can meet in the housings formed solid which different parts have different temperatures. The heat can be spread the leadership also in liquids and gases. Here, however, it applies mainly convective heat transfer [3]. When creating model situations, however, each particular case tends to describe using a method which will most likely be in the heat exchange prevail. Our models will therefore be simplified, but only enough to make the most accurately describe the situation and comply with the results observed from real measurements.

Because heat transfer reminds the water flow pipe or net charge, there can be found a number of analogies. In this regard, we divide the thermal conduction:

- stable (stationary) heat conduction - the steady leadership is the difference in temperature between the different parts of the body stable, ie. independent of time,
- unsteady (time-varying) heat conduction - under dynamic leadership occurs over time the gradual equalization of temperature differences between different parts of the body.

The Basic Law of heat transfer by convection is Fourier's law, which says that the magnitude of the heat flux q (W/m2), generated during thermal conduction is proportional to the magnitude of the temperature gradient and he is of opposite sign [4]. For one-dimensional problem of thermal diffusion can be Fourier law to enroll in a scalar form:

$$q = k\frac{\Delta T}{L} \tag{2}$$

where in L is the thickness of the body in the direction of heat flow, q and ΔT are both positive values. Using this formula, we must keep in mind that heat always flows from high to low temperatures.

The fundamental problem of calculating heat transfer by Fourier law is that the law has two dependent variables q and T. It is therefore necessary to eliminate the size of the heat flux q and solve math problems for a single unknown, the temperature T. As already indicated in the chapter on heat conduction, to eliminate the variable we use the law of

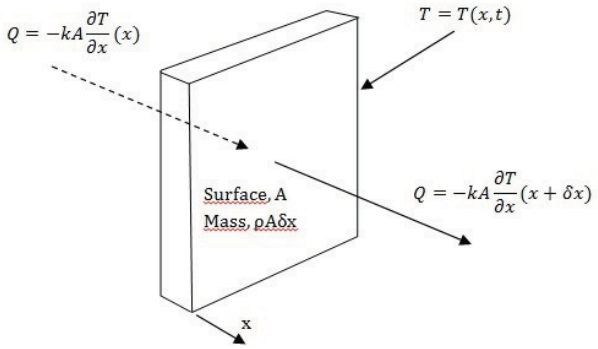


Fig. 1. One-dimensional heat conduction

conservation of energy or the first law of thermodynamics. Consider a one-dimensional element of the system Fig. 1 with Fourier's law applied to both sides.

Total heat transferred by convection heat during steady heat transfer can be expressed as follows:

$$qA = Q = -kA\frac{\partial^2 T}{\partial x^2}\delta x \tag{3}$$

Then we can replace heat loss Q using sentences of the first law of thermodynamics in the form:

$$-Q = \frac{dU}{dT} = \rho cA\frac{dT}{dt}\delta x \tag{4}$$

wherein ρ is the density of the object, and c the specific heat capacity. Combining relations (3) and (4) we get the equation of one-dimensional thermal diffusion:

$$\frac{\partial^2 T}{\partial x^2} = \frac{\rho x}{k}\frac{\partial T}{\partial t} \tag{5}$$

It is necessary to emphasize the importance of this expression, which is precisely the removal of an unknown quantity Q by combining Fourier law and the first law of thermodynamics. The resulting differential equation allows us to determine the temperature distribution T (x, t) in the system during the heat exchange line. It is also a fundamental equation of the whole theory of conduction heat transfer [4].

IV. Three-dimensional Equation of Thermal Diffusion

In the three-dimensional body, the temperature of which is for some reason (e.g. by heating one side) varies in space as well as in time, the temperature can be defined as an array

$$T = T(x, y, z) \text{ or } T(\vec{y}, t) \tag{6}$$

The size of the heat flux has not write as a scalar quantity, but as a vector. It has the size specified in addition to already specific direction. Based on Fourier law can be expressed as

$$\vec{q} = -k\nabla T \tag{7}$$

After the distribution of the three components we get the form:

$$q_x = -k\frac{\partial T}{\partial x},\ q_y = -k\frac{\partial T}{\partial y},\ q_z = -k\frac{\partial T}{\partial z} \tag{8}$$

Now we have Fourier's law in three dimensions. Below is the actual derivation of the diffusion equation of heat conduction in the three-dimensional shape.

As background, it was only the expression of the first law of thermodynamics. This relationship is applied to three-dimensional control volume, which represents some finite region of the head body heat. Surface marked as S and the volume as *R*. Identifies a certain element of surface *dS*, and two vectors corresponding *dS* - normal vector $\vec{n}$ (Price $|\vec{n}| = 1$) and heat flux vector $\vec{q}$ at a given point of the surface. For the general case of in the equation we include the possibility distribution of heat flow volume $\dot{q}(\vec{y})\ W/m^3$ across the region. This heat flow can be the result of a chemical or nuclear reaction or radiation from an external source and so on.

According to the model described in this way can the heat removed from the element *dS* calculate the expression

$$(k\nabla T)(\vec{ndS}) \tag{9}$$

Heat generated (or absorbed) in the region *R* must be added to total heat flow toward the *S*, to obtain the total heat gain in *R*.

$$Q = -\int_S (-k\nabla T)\cdot(\vec{n}dS) + \int_R \dot{q}dR \tag{10}$$

Rate of increase of energy in the region R is

$$\frac{dU}{dt} = \int_R \left(\rho c\frac{\partial T}{\partial t}\right)dR \tag{11}$$

where the derivative T is in partial form, because it is a function of $\vec{y}$ and *t*. Finally, we combine Q expressed in Equation 10 and *dU/dt* given by Equation 10 in equity and after treatment markedly profitable relationship

$$\int_S k\nabla T\cdot\vec{n}dS = \int_R \left(\rho c\frac{\partial T}{\partial t} - \dot{q}\right)dR \tag{12}$$

To adjust the left side of the equation more satisfactory form, we use Gaussian sentence that converts integral surface of the integral volume [5]. Theorem says that if $\vec{A}$ is any continuous function position, the following applies:

$$\int_S \vec{A}\cdot\vec{n}dS = \int_R \nabla\cdot\vec{A}dR \tag{13}$$

Whereas the region R is freely given, can be removed along with the integral of the equation. It gives you final form of the differential equation of thermal diffusion in three-dimensional space in the form:

$$\nabla\cdot k\nabla T + \dot{q} = \rho c\frac{\partial T}{\partial t} \tag{14}$$

V. Simulation in Matlab

MATLAB contains a number of additional products that provide important tools and features expanding our capabilities in the development or design applications. The primary tool in the development of our simulation was in terms of programming PDE Toolbox (Partial Differential Equation Toolbox), which provides functions used to solve partial differential equations in 2D and 3D space or over time [6]. It includes commands that allow to specify a two or three-dimensional geometry model of the analysis, to formulate the boundary conditions and the parameters of the equations themselves [7]. The results of equations can be further processed and visualized in the form of graphs.

A. Transcription equations for thermal diffusion PDE Solver

Mathematical models in the previous chapter will be administered in the form usable in the context of the centre point of MATLAB. As we already mentioned, the functions

relating to the calculation of partial differential equations arch over toolkit called PDE Toolbox. Differential equation of thermal diffusion, which we derived in the previous chapter in complete form, expressed in Equation 14, rewrite with some minor adjustments to the form in which it will use for the method of PDE Toolbox:

$$d\frac{\partial u}{\partial t} - c\nabla^2 u + au = f \tag{15}$$

Unknown quantity u, which we want to express, that represents just variables temperature T. Variables d, c and f are be classified as the primary coefficients in MATLAB they are used as input data of used PDE solver, thus functions for solving partial differential equations. Variable d is therefore a product of density and heat capacity of the material (ρ.c) and variable c represents the coefficient of thermal conductivity k. Using the coefficients a and f to describe additional processes affecting heat flow thermodynamic system (eg. heat flow volume $\dot{q}$ in Equation 14 shown for illustration). Therefore, they should in this case zero. Finally, the term $\nabla^2 u$ (resp. $\nabla^2 T$) is called the Laplace operator and it is the differential operator indicating the divergence of the gradient field (in our case the temperature T):

$$\nabla^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \tag{16}$$

B. Calculation of equations and its visualization

In the initial part of the MATLAB programming was created the basic code of the program, which will take care of the actual equation calculation and visualization of the process of heat transfer. First, a case has been processed elementary of a thermodynamic system with a single layer and the base material. On the left wall of the body is heat source of a fixed temperature, while the right wall of the heat is discharged out of the system. The remaining edges of the body be considered thermally insulated, so no heat enter them into the system or leaking from it.

We have analyzed the steady state, as well as a unsteady (continuous) state. In the case of steady state, we are interested in solutions of final temperature gradient of body at steady state under these conditions. Under unsteady state, the result of heating (or cooling) is obtained from the initial body temperature over time. At the beginning, it is necessary to initialize variables storing the basic physical parameters of the body and define variables representing coefficients of partial differential equations Equation 15.

An essential part of setting the specification of boundary conditions for differential equations. There are two basic types of boundary conditions Dirichlet boundary conditions and Neumann boundary conditions. MATLAB has to specify the boundary conditions pdeBoundaryConditions() built-in functions, by which we can enter both types of boundary conditions. After settings input of all variables we can proceed to the actual calculation. In order to enter the input values more simply and more transparent, we created a Graphic User Interface with the required control elements Fig. 2.

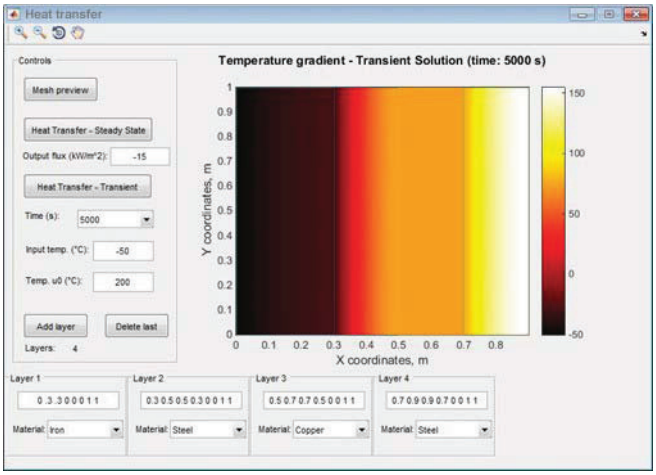


Fig. 2. Final application – computing unstable state

The Textbox *Output flux* means the supply of heat from the delivery side to the right. Number with a negative value represents a heat sink. Textbox *Input temp.* represents the temperature at the input (left) side of the first layer of the system. For each layer the coordinates are put in the form [x1 x2 x3 x4 y1 y2 y3 y4]. *Decsg(GDM, f, n)* function perform decomposition geometric model constructed in a narrative matrix GDM. After analyzing the input model it will break it down to the basic incoherent primes - a minimum area defined boundary segments. Use the command *initmesh()* gives a triangular model of two-dimensional geometry using Delaunay triangulation algorithm. Object properties defines itself input geometry *g*, but also other attributes entered as parameters in the feature name-value pairs. In our model, we used the attribute *Hmax* determining the maximum size of the edges generated triangles. Function *PdeGeometryFromEdges(g)*, created from decomposing solid geometry geometric shape, making it easier to define the boundary conditions of differential equations, which can then specify Modular (for individual model edges. Then we decomposed geometry portray as a model divided into triangular elements shown in Fig. 3. Defining the boundary conditions in MATLAB functions we are realized through *PDEBoundaryConditions ()* with parameters extracted from the input parameters *Output flux* and *Input temp*. From the Fig. 2 at the edge number 4 (left side) we applied Dirichlet boundary conditions having a temperature of 400 ° C, and the opposite edge of the bearing 2 in turn Neumann boundary conditions the value of the variable heat flow g=-15 kW/m^2 (flowing to the outside of the system). With this set of conditions we start solving partial differential equations of heat transfer.

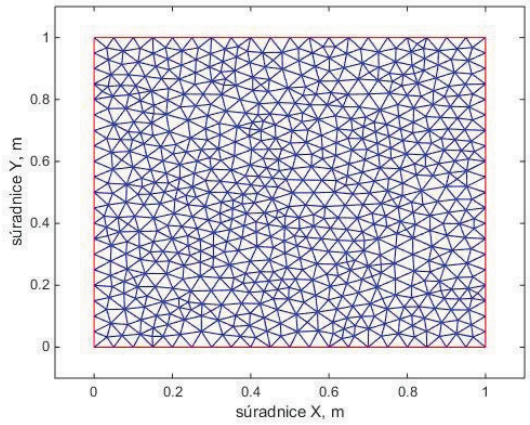


Fig. 3. Graf of triangulated model

VI. Conclusion

In the first part of the paper we dealt with process related to heat transfer in solids. We have introduced a cross-section of knowledge and physical laws describing the heat and his replacement, we are dedicated to scientific notation these laws. The main objective, however, was algorithmization mentioned mathematical physics calculations in MATLAB.

The final version of the application allows user to create two-dimensional geometric model of the body with multiple layers of different materials through which can then run simulations of heat transfer. The simulation takes place either with the specified outlet respectively inlet of the heating system in case of the steady state solution, or simulation runs with the given initial temperature of system in the case of solution the current status of the selected finite time. When specifying individual layers the model has a number of material with which it comes into contact normally either at construction sites, in a variety of instruments and equipment, the thermal insulation/management, etc.

Acknowledgment

This paper is published thanks to the financial support of the project VEGA 1/0559/14, Named: Evaluation of machine translation and the project KEGA 011UKF-4/2014, Named: Education and Counselling in Telenursing in Home Health Care.

References

[1] Koprda, Š., Balogh, Z., Turčáni, M. 2011. A design and simulation of fuzzy PID controller for the optimization of temperature and humidity in the thermodynamic system. In: Recentresearches in neural networks, fuzzy systems, evolutionary computing and automation: 12th WSEAS International Conference on fuzzy systems (FS '11). Brašov: WSEAS Press, 2011. ISBN 978-960-474-292-9, P. 119-126.

[2] Planck, Max. 1914. Vorlesungen über die Theorie der Wärmestrahlung. [transl.] Morton Masius. The theory of heat radiation. Philadelphia: P. Blakiston's Son & Co., 1914. LC Control Number: 14008862.

[3] Rajan, K.S., Pitchumani, B., Srivastava, S. N., Mohanty, B. 2007. Two-dimensional simulation of gas–solid heat transfer in pneumatic conveying, International Journal of Heat and Mass Transfer, Volume 50, Issues 5–6, March 2007, Pages 967-976, ISSN 0017-9310. [online] Available at http://www.sciencedirect.com/science/article/pii/S0017931006004807 [Accessed 2 February 2016].

[4] Lienhard, John H.,V., Lienhard, John H., IV. 2011. A Heat Transfer Textbook: Fourth Edition. 4. Cambridge : Dover Publications; Fourth Edition edition (March 17, 2011), 2011. p. 768. ISBN-13: 978-0486479316.

[5] Antonakakis, T., Maglioni, C., Vlachoudis, V. Closed form solutions of the heat diffusion equation with a Gaussian source, International Journal of Heat and Mass Transfer, Volume 62, July 2013, Pages 314-322, ISSN 0017-9310. [online] Available at <http://www.sciencedirect.com/science/article/pii/S0017931013001853> [Accessed 20 January 2016].

[6] Vozár, M. 2010. Differential equations in MATLAB. In : Biometric Methods and Models in Current Science and Research : proceedings of the XIX. Summer School of Biometrics, 6.-10.9.2010. Brno: Mendel University, 2011. ISBN 978-80-7401-028-6, P. 1-9.

[7] Brečka, P., Magdin, M., Koprda, Š. 2011. Two-state regulation in MATLAB for the comparison of some parameters (damage power conpsumption) by PSD a regulation. In : Lecture Notes in Electrical Engineering : advances in computer, communication, control and automation. ISSN 1876-1100, Vol. 15 (121) (2011), p. 369-376.

Video in Teaching

Veronika Pisarenko
Foreign Languages Department
Institute of Management in Economic,
Ecological and Social Systems of Southern Federal University
Taganrog, Russia
jetaime2014@yandex.ru

Galina Krasnoshchekova
Foreign Languages Department
Institute of Management in Economic,
Ecological and Social Systems of Southern Federal University
Taganrog, Russia
krasnoshokova@tgn.sfedu.ru

***Abstract*— Teaching Internet information resources are slowly conquering the system of education, which gives them their methodical adjustments. The use of information technologies aims at developing students' self-study work and at the ability to think logically, at the formation of creative ability, which is a condition for the formation of communicative competence and readiness for intercultural communication. The development of new technologies and techniques of teaching based on information technologies is considered to be one of the most important problems of education. Education technologies have been a research focus for many years in higher school pedagogics. At the same time, foreign language is considered one of the most important subjects not only for the future specialist training but also for the student's personality successful development.**

***Index Terms*—Video, training, information technologies, audio-visual technologies, grammar.**

I. Introduction

In the context of technical universities in the study discipline "Foreign Language", where a small amount of hours is given, it is especially necessary to organize the self-study work of students rationally. And we have the invaluable help of modern information technologies in this process.

The scientists have tried to find the ways to optimize the learning process in technical universities for a long time. Various options of organization and methods of teaching foreign languages were offered, but significant positive results have not been received yet.

This paper is about seeing, making, and acting in and through a foreign language. The object of attention in the activities is the exploration of the world of video and the way it relates to oneself and to others in the class.

Audiovisual Technology represents a new approach to teaching methods and techniques and this is the current reality in education. Audiovisual Technology dramatically improves efficiency and availability of teaching resources.

There are many definitions and concepts of *technology* in literature. *Audiovisual Technology* is a term for the abstraction of the process of constructing a set of operations, methods and techniques based on video resources and resulting in competence formation [1]. This technology is based on use of human audio and visual channels of information perception. It permits to organize the perception in maximal value. This is very important for teaching foreign languages. Video resources allow to demonstrate the object of studies – a foreign language – in real conditions. Video permits to demonstrate not only a language but the whole environment where it exists including cultural, social, ethnic, historical and many other aspects [2, 3, 4].

The following scheme of knowledge mastering process presented on Fig.1 has been used for the construction of algorithms of the foreign languages training by means of video materials [16].

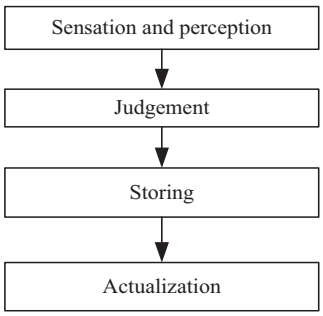


Fig. 1. The scheme of knowledge mastering process when studying foreign languages

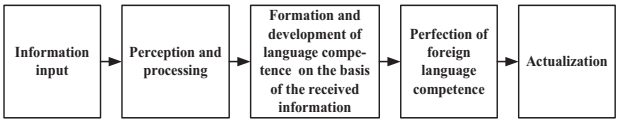


Fig. 2. The scheme of knowledge mastering process when studying foreign languages

The kind of camera use we are proposing here differs radically from the video work we have seen in some colleagues' classrooms, where the students prepare long stretches of language, the teacher videos them producing it, and then spends time going over their mistakes. In this sort of work the focus is largely on the language the students produce in front of the camera. The camera is being used as a passive recording device, not as a maker of reality. We are proposing that the students should control the interaction between the camera and the scene being portrayed. Much of the real language work will come in the process of discussing, brainstorming, planning, videoing, reformulating, and trying out again. Language may be produced on camera, but the main use of language is in the videoing process itself. We propose that language produced in front of the camera should be used as text suitable for teacher-led correction sessions. When a small group of students presents the class with a video they have just made, the focus is naturally on the whole viewing experience not on the fragments of recorded inter-language that do not conform to the norms of standard English. There are other times and places for correction.

The scheme presented on Fig. 2 shows the specificity of process of foreign languages training with use of audiovisual technologies.

A special video activity book has been developed. It contains the description of different language training activities based on video. The activity book falls into three sections: *Active viewing, Making your own videos* (*Whole class video work* and *Small group video work*) and *Video your course book.* For example: two people leave the group for half an hour to make a three-minute video in which one of them describes his or her own face while the other videos the features being described. The result is a word and picture evocation, which is subsequently shown to the whole class (see activity *The camera as mirror)***.**

If you want students to practise descriptive language, this is an exploratory, personal, and real way of doing it. It is real in the sense that the speaker is relating to the image of himself or herself on the monitor as they are being videoed, to the presence of the camera person, and to the rest of the class who will see the video later. In this type of activity people are saying important things to themselves and each other via the camera.

A central concern of this paper is to bring the common sense of humanistic thinking to the area of foreign language video work in accordance with some simple principles:

- the learner is centre-stage as actor, doer, creator, and interpreter;
- the teacher offers the learner initial situations or frames as short cuts to meaningful self-expression;
- since learners (and teachers) are very different from one another, they need to be provided with a wide range of activities and stimuli;
- all text is produced by the students - when viewing they produce their own very particular, internal version of what is on the screen; when videoing they make all the decisions both behind and in front of the lens;
- even coursebook material can have life breathed into it by letting the students take it over and re-fashion it in their own special way.

II. Materials and Methods

A. Active Viewing

Video is a supercharged medium of communication and a powerful vehicle of information. It is packed with messages, images, and ambiguity, and so represents a rich terrain to be worked and reworked in the language learning classroom. This section of the paper presents activities on the use of the VCR and the monitor as teaching aids to exploit this wealth of information. However, these are not activities in video deconstruction. We simply propose approaching video differently from the way it is conventionally used in the home, work place, or in video training packages. For example, even the verb 'watch', in using video differently, needs an extended definition. We invite students and teachers to interact with a powerful and very versatile medium, which is the major, pervasive, opinion-influencing cultural object of our age.

To practise manual focusing zoom onto any distant object, such as a lamp-post, and focus as quickly as possible. Repeat quickly, selecting objects at different distances.

The first section aims to free the student from the passive, consumerist feeling that some people bring to viewing. The activities in this section allow students to focus on the different strands that make up the video message, to look at them separately, and then in various combinations. The section deals extensively with video soundtrack and includes activities that pick out the following areas:

- the use of sound effects;
- the relation of Language 1 subtitles and Language 2 speech;
- the written script behind much soundtrack;
- how a written script becomes spoken language;
- the relationship between the soundtrack and the pictures;
- the use of key words and phrases within a videoed text.

By unraveling, the strands that go to make up a video message students become aware of what the TV is feeding them each night. They take what for some students are the first steps towards video/ TV literacy and they do this via the medium of the foreign language. In other words, the language work is part of a process that is wider and more interesting than itself.

You can, of course, teach any subject using the target language as the vehicle. However, video literacy combines particularly well with work on a foreign language since the relationship between the pictures and the language directs powerful attention to the language. We know of few more successful ways of getting students to want to listen to language than to invite them to first watch the pictures of a video sequence without the soundtrack. This time-honored EFL exercise prepares students for the listening by contextualizing the language and giving them a reason for listening. An incomplete message provokes most people to want to get their hands on the missing part of the communication.

This section also focuses the students' attention on the visual part of the video message. For example activity *1.2. Count the cuts***,** asks students simply to view a short sequence and then to count how many cuts they noticed and where these occurred. The activity will often generate lively discussion with plenty of reference to both the soundtrack and the pictures. The students are the doers and the judges, not merely the recipients, of the message. The message from the screen is the material they act on and judge.

B. Making Your Own Videos

This section offers you ideas for involving your class in videoing activities. It begins with some short camera familiarization activities (Part 1) to get your students used to being watched by a camera, to holding one, and to using various camera techniques.

The aim of this section is to arouse people's creativity and allow them to discover skills that in some cases they do not know they possess. The activities are designed to stimulate a range of emotions and states of mind such as aggression, humour, concentration, curiosity, playfulness, and so on. The activity frames offered are often unexpected and students find themselves doing in English something they have probably never done before in any language. Enough activities of this sort will gradually

change your students' internal relationship with the target language. We are a long way from the railway timetables and rehearsal for the 'real world' that fill many 'communicative' coursebooks. The activities here are not rehearsals for some future situation - they are here and now situations that generate their own immediate reality.

The second part of this section (Part 2) suggests activities that are best done by small groups of students rather than by the whole class. While the presence of the whole class enhances the activities in the first part of this section, a student looking at his or her own image on the monitor screen, for example, and carefully describing his or her own features does not want to have the whole class staring at them during the making of the video. For these activities the videoing team needs to be made up of two or three people (a camera-operator, the main actor, and maybe a director).

C. Video Your Course book

This section suggests allowing the students to use the camera to take over part of the course book and make it their own, and so learn the language it sets out to teach them.

The course book will almost certainly include dialogues on the page and on tape. Missing from these will be the time of year, the time of day, the sort of light, the setting, and the way the characters look. Even if there is a picture next to the dialogue in the book it can only be a 'still'. You can ask a camera team to video the dialogue the way they imagine it and then use their video to present the dialogue to the class.

The course book is a massive piece of external text that is dropped on the table in front of the student. Videoing grammar presentations, dialogues, reading passages, and revision of past units is one way of making the course book less remote and inanimate. It follows the same principle as getting the students to draw a picture on the board for a grammar presentation rather than relying on a five color picture in the book. The book picture is dead, while the blackboard picture belongs to the class.

This section of the book draws on the inspiration of Hans Eberhard Piepho of Giessen University. When he has to teach a course book unit to a class he asks the students to go through the unit picking out things they find good, bad, or intriguing, at the level of words, phrases, sentences, explanations, and pictures. He then builds his teaching of the unit round what the students bring up. It is one way of turning an off-the-peg book into something more tailor-made. Asking students to video parts of their book is one way of getting them to express their vision of what is presented to them by the distant, absent course book writer.

D. How Each Activity Is Organized

Each activity is presented under five main headings: *Level, Time, Preparation, In class* (in Active viewing) and *Procedure* (in *Making your own videos* and *Video your course book*), and *Comments***.**

Level indicates the minimum level at which the activity can be carried out. Sometimes a range of levels is given to show that, with suitably adapted materials, the activity can be used with different levels. *Time* is a rough guide to the amount of time the activity will take in class. *Preparation* tells you what kind of preparation and/or materials you will need for the work to be carried out in the classroom. *In class /Procedure* - under either of these headings you will find a step-by-step guide to carrying out the activity in the classroom. Also included under *In class! Procedure* are variations on the approach or the texts used in the activity. The *Comments* are there to help you and your students to understand the purpose of the activity, to provide you with useful tips, and to point out some of the problems that might arise.

E. Choosing Activities

The activities with video can be used in a wide variety of situations.

1. You may be the sort of teacher who makes a lot of use of educational videos and off-air material. In this case you have to use Active viewing with plenty of viewing techniques in that build on ideas you are already familiar with.
2. You may be teaching in the sort of school where using the course book is obligatory and nothing is allowed outside this framework. In this case Video your course book may be your natural way into the book.
3. If you are teaching a foreign language for specific purposes it will be interesting for your students to make some video on some experiments or practical workshops on different disciplines.
4. You may work in an adult education system in which a course around video literacy via the medium of English would be popular with students.
5. If you are teaching a literature-weighted course to upper secondary students, you may find one or two of the text-to-video sequence ideas are what serve you best. There may be just one format that gives you a term's work.

F. Familiarizing Yourself with the Camera

1) *How to Hold It.* When you are not using a tripod the main problem is how to keep the camera steady and not to suffer too much from eyestrain. To help keep the camera steady, place one hand under the elbow of the arm which is actually holding the camera. Inexperienced camera operators, not knowing what to do with this hand, often attach it to the camera's handle on top. By keeping this hand under the elbow, the camerawork will become much steadier to the viewing eye. The next tip takes a little practice. Most people, when videoing, squint through the lens with one eye and keep the other eye shut tight. Not only is this a strain on the closed eye, it is not even necessary. When videoing, try to get in the habit of keeping both eyes open. One advantage in doing this is that, while doing any action-videoing, your open, non-videoing eye can see things you might want to include in your video that you might otherwise miss.

2) *Automatic and Manual Focusing* Most modern camera-recorders have automatic focusing, with a choice of automatic or manual. The advantage of autofocus is that the camera 'thinks' for you, just like a still camera with automatic focus. The disadvantage is that the camera does not know what you want to focus on. For example, if you are videoing a person speaking four metres away and the camera is fixed on that person, the lens automatically

focuses on this subject. However, if another object is introduced between the camera and the subject, for example a bottle placed on the table, then the lens automatically focuses on the new object. To avoid such video misunderstandings, it is useful to be able to use manual focusing. Without prefocusing, zoom in on any object and try to hold focus as you move towards it. Track in on any object and try to hold the focus as you track. Then track out, holding focus. Practise varying the speed of the zoom or tracking as you go.

3) *Zooming.* For videos you plan to use as part of a video archive, avoid using the zoom function too often while recording. The effect of zooming during a sequence distracts from the subject. Instead, try videoing a subject, putting the camera on pause, then zooming in or out while on pause. Take the camera off pause and go on with your shooting. When you view this sequence, you will notice that the effect is a natural transition and gives your video a more professional quality. Another disadvantage of zooming while videoing is that on some cameras the simultaneous use of the built-in microphone will pick up a hum on your soundtrack. The use of an extension microphone will avoid this problem.

4) *Lighting.* Not being lighting experts, all we have ever done when we have had access to studio lighting is simply to position the lighting in order to get a pleasing effect. Most cameras nowadays have automatic light regulation so that the camera operator need not be a lighting technician. Try to use natural daylight whenever possible, or good-quality artificial lighting, and make sure the light source is behind you. Cameras should not be pointed at direct lighting sources.

5) *Meet the Mike.* Videos are good for making 'silent movies' as well as 'talkies'. But if your video comes out silent or muffled, and you really wanted a proper talkie, you probably did not use the microphone to its fullest advantage. A typical video recorder has two microphone possibilities, the microphone built-in the video recorder and an external microphone. The first is useful when the camera operator needs to move around and wants to worry less about tripping over wires. This microphone is usually very sensitive and is intended for active videoing projects. However, because the microphone is not always the same distance from the speakers, the sound quality on the video may vary. There are several ways of representing the passage of time in a few seconds of video 'real' time. A cue card can be marked with phrases like 'six months later' and these can be videoed. A title generator, an accessory available for some cameras, will do the same thing but you should place it directly onto a desired shot. Alternatively, select images to symbolize time passing, such as an egg timer. An old Hollywood standby is simply fading out then fading back in, which can create the effect of almost any amount of time having passed. A Chaplinesque device is to zoom in on a clock which is being held and at the same time wound forward. The viewer will see only the clockhands going round and round. The second microphone is perfect for scenes that do not require a moving camera. Normally this microphone is plugged into the video recorder or into the VCR along with the recorder. Check in the camera and VCR manuals where to plug in these cords. For videoing, set up the microphone so that it is close to the speaker(s) yet not included in the screen image. The resulting on-screen effect is natural and the sound quality is even and clear. We suggest you get to know the extent of your microphones' effectiveness before you try out some of the activities presented in this book. For example, try out *The heckler* (activity 2.6) with some colleagues and use the built-in microphone. When you play back the video, you will have some amusing feedback on this microphone's capabilities. Likewise, experiment with *Video talks* (activity 2.12) and test setting up and using your off-screen external microphone. The results of these trial runs may satisfy you enough to add them to your video archives!

6) *Buttons.* Many of the buttons on the newer cameras are automatic function options, such as lighting and focus. In other words, there are added buttons in order *to cut down* on camera operator stress. Many buttons have a common ancestry with the audiocassette recorder. A short, practical session with a couple of your colleagues deciding which buttons you ***really*** need for your video purposes will be time well spent.

7) *Camera Shyness.* During the normal course of a lesson have the video camera switched on, but not actually recording. Turn the camera towards the class so that the picture displayed on the TV monitor is what the camera 'sees'. Participants' first reactions are sometimes to admire themselves on the screen, make funny faces or grin like Cheshire cats. If the 'live' picture is permanently displayed, students soon come to treat the screen more like a group portrait which happens to move, and will gradually ignore it. The result is that when you do record something, the participants do not feel self-conscious. Seeing oneself on the monitor in a class- made video will seem quite natural.

Note: it is a serious error to make 'secret' recordings of students when they are unaware of being videoed. A trust relationship between the subject and the camera is vital.

8) *The Camera Sees the World.* Before participants actually use the camera, let them experience seeing what the camera sees in order to highlight the camera's limitations compared to the self-focusing, wide-angled human eye. They can each make a card, A4 size with a square about 6 mm by 6 mm cut out of the middle. Let them walk around, inside and outside the classroom, looking through the square, which simulates the camera lens.

III. Grammar in Videoing

Grammar represents a very big problem in training foreign languages. Trying to fill this process with new content, we have developed and tested an innovative multimedia mean for methodical training and self study English for students «Explicator» [3].

Multimedia «Explicator» is a grammatical reference composed of two sections:

a) Linguistic comments of grammar phenomena;

b) Complex of language training exercises and tests.

«Explicator» of grammatical phenomena is based on the results of the typical difficulties of acquisition analysis. Comments of grammatical rules, instructions are formulated in terms of functionalities of English grammar phenomena. Complex of linguistic exercises activates the grammar phenomena in the training of foreign speech.

Explaining and primary training are carried out in self-learning mode, so we may release the time for the formation of foreign language communicative competence based on communicative exercises.

Thus, using of «Explicator» presentation of a linguistic phenomenon is carried, accompanied by an explanation sufficient to understand and appreciate its meaning, necessary for further training [1].

Using of «Explicator» will greatly increase the academic autonomy of students. Methodically advisable to reorganize training through self-supported formation of computer language skills outside the classroom in the absence of a teacher. In this regard, the majority of classroom time is devoted to the development of foreign language communicative competence by performing tasks and exercises of interactive nature (conferences, debates, projects, role plays, etc.).

Testing computer programs provide objective control the level of communicative competence formation. In addition to training programs, a lot of multimedia resources of the Internet and telecommunications networks are used in the process of teaching, including multimedia computer dictionaries, reference books, encyclopaedias, computer courses for the development of social competence and expanding background knowledge of students.

Innovative learning means, as experience shows, and special studies by themselves do not have any significant effect on learning results. It requires a system-forming basis, what presents the developed and introduced into the practice of teaching English in technical universities explicative - communicative method, which provides a rational combination of explanatory and illustrative and communicative teaching, based on the integration of cognitive-oriented and context-competence approach.

Explicative-communicative method is based on the methodological principle of philologization and aimed at teaching foreign language communicative competence of students based on linguistic competence; provide the fundamental process of mastering a foreign language communicative competence [2].

To master foreign language communicative competence effectively explicative - communicative method provides training algorithm of successive steps:

- Explication – linguistic explanation of grammatical phenomena, full acquisition of its meaning (self-study work with the first part of the explicative textbook «Explicator»);
- Initial training in the use of grammatical phenomena in speech, language self performance of the second part of the exercise «Explicator»;
- Pre-communicative training (linguistic revitalization of students' activities during the execution conditionally communicative exercises of professional training and speech - oriented situations in classroom under the guidance of a teacher;
- Communicative practice in all types of speech activity on the basis of the studied material (activation of linguistic activity of students in the implementation of communicative exercises in the classroom under the guidance of the teacher).

Thus, having understood the form and function of grammatical phenomena, further training under the guidance of teacher students perform communicative exercises, learning to recognize and understand mastered grammatical phenomenon in video, in the text, in audio texts, use in monologue and dialogue speech in cooperation with partner communication, as well as writing.

IV. Results and Discussion

In the course of working with video and multimedia explicator students perform the following self-learning cognitive-communicative activities:

- Understanding of grammatical phenomena, processing and acquisition of their contents;
- Activation of grammatical phenomena by repetition, the analysis of this material in the tables, in language training exercises, use learned grammatical material in other forms;
- Action to use in the speech grammatical material individually, paired with a partner in the training group;
- Action-control, self-monitoring and self-evaluation on the development of grammatical material.

Organization of self study work using multimedia explicator with video is a reliable mean of optimizing the process of learning a foreign language.

Studying with the help of explicative-communicative approach by using multimedia explicator with video, students move from the minimum of autonomy on the first stage through an intermediate level on the second stage to the level of maximum autonomy on the third stage of training speech in the learning process.

V. Conclusion

The media of video training and self - training explicator may be qualified as information and communication technologies as it meets the basic requirements for information and communication technologies, in particular:

- Focus on the development of the student individuality, his autonomy;
- Domination of self- study work over training by the teacher;
- Openness of the process of mastering a foreign language communicative competence of each student;
- Unity of teaching, learning and development;
- Providing opportunities of nonlinear organization of the foreign language learning process;
- Providing opportunities for students to choose their individual educational route in mastering foreign language communicative competence and the timing of its passing;
- Modelling in the educational process of communicative situations and cognitive activity;
- A combination of individual, pair and group classroom work and self extracurricular activities;

- Providing training for different interconnected verbal communication;
- Ensuring productive character of educational communication activities trainees.

REFERENCES

[1] V. Pisarenko, "Audiovisual technologies of foreign languages teaching in technical high school". Moscow: Izdatelskiy dom MPA-Press, 2004.178 p.

[2] V. Pisarenko, "Video resources use in modern technologies of foreign languages teaching" In IV Russian scientific and practical conference "Modernization of System of Professional Education Based on Regulated Evolutioning", V 6. Ch.1. Yuzhno-Uralsk State University. – Chelyabinsk: «Obrazovanie», 2005, pp. 65-69.

[3] G. Krasnoshchekova, "Information technology in solving a problem of quality language training" // Open Education. - Moscow, 2010. - № 6. - S. 68-75.

[4] V. Pisarenko, "Communicative problems in foreign languages teaching based on audiovisual technologies", Izvestiya TRTU. Taganrog: Izd-vo TRTU, 2005. № 5 .pp. 121 – 127.

[5] O.I. Barmenkova, "Videolessons in system of teaching foreign speech". In "Foreign Languages at School" 1999, № 3. 20-25 pp.

[6] V. Pisarenko, "Pedagogic algorithm of video materials use" In "Perspektivnyie informatsionnyie tehnologii i sistemyi". Vol. 1(13). Taganrog. 2003. pp.77–83.

[7] V. Pisarenko, "Audiovisual technologies and their use in teaching". In «Intellektualnyie sistemyi. Problemyi vyisshego obrazovaniya». Moscow: Fizmatlit, 2004. pp.433 – 442.

[8] I. Platonova, "Television in foreign languages teaching". Minsk: Vyisshaya shkola, 1980. – 88 p.

[9] I. Platonova, "Teaching television and problem of creation of telesituations of communication": Dissertatsiya na soiskanie uchen. step. kandidata ped. nauk. Minsk, 1976. 167 p.

[10] I. Platonova, A. Nyashina, E. Tobolich, "English on TV". Minsk, 1974. 145 p.

[11] "Studying Process: Perception". Moscow: Pedagogika, 1982. 336 p.

[12] E. Polat, "Use of foreign languages teaching tutorials in language lab". 2-e izd. Moscow: Vyisshaya shkola, 1988. 158 p.

[13] E. Polat, "Telecommunications at school". Informatika i obrazovanie. 1993. № 1. pp. 55–57.

[14] N.V. Baryshnikov, "Learning French in high school: questions and answers". - M.: Education, 1992. - 139 p.

[15] V.V. Korotenkova, " Philologisation of foreign language teaching" // Foreign languages at school. - 2001. - № 4 . - S. 22-28.

[16] I. Salistra, "Methods of foreign languages teaching". Moscow: Vyisshaya shkola, 1966. 252 p.

[17] M. Allan, « Teaching English with video". London: Longman. 1986.

[18] J. Alinc, M. Clay, P. Léon, « Le laboratoire de langues et la classe audio-visuelle ». – Langues modernes, № 5. Paris. 1965.

[19] Yu. Babanskiy, "Problems of Effectiveness Of Pedagogical Research", Moscow: Pedagogika, 1982. 192 p.

[20] R. Armes, "On Video". London: Routledge. 1988.

[21] J. Berger, "Ways of seeing". London: Pelican. 1972.

[22] J. Berger, "About looking". New York: Pantheon. 1980.

[23] E. Beyer, « La perception humaine ». Paris, Hachette, 1986. 343 p.

[24] R. Cooper, M. Lavery, M. Rinvolucri, "Video". Oxford University Press. 1996. 112 p.

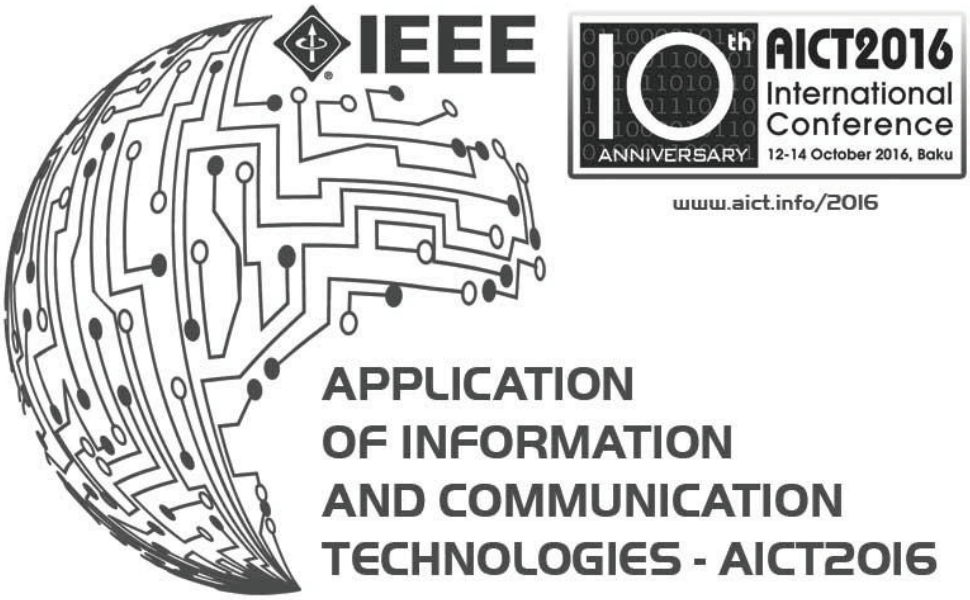
[25] A. Hunt, "The Language of television". London: Methuen. 1981.

[26] K. Kennedy, "Film in teaching". London : B.T. Batsford Ltd. 1972.

[27] V. Pisarenko, "Perfection of technical specialists language training at technical university". In "II All Russia Conference "Problems of Modern Philology and Methods of Foreign Languages Teaching", Penza, 2003.pp. 166-168.

[28] Voinova A.V. "Education informative and dynamic reading in the higher technical school (in the English language)" : dis . ... Candidate .ped. Sciences: 13.00.02. - Pyatigorsk , 2003 . - 182 c.

[29] V. Pisarenko, "Computer technologies use in foreign languages teaching", Otkryitoe obrazovanie. 2004. № 1 (42).



CONFERENCE PROCEEDINGS

SESSION 8.

ICT IN MEDICINE AND HEALTH CARE

12-14 October 2016, Baku, Azerbaijan

Dichotic Listening Test in case of Small Stimulus Set: Random Walk and Sequential Wald Analysis

Elena Popova[1], Evgeny Wasserman[2], Nikolay Kartashev[3]
[1,2,3]Saint Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, Saint Petersburg, Russia
[2]Saint Petersburg State University, Saint Petersburg, Russia
[2]Herzen State Pedagogical University of Russia, Saint Petersburg, Russia
[2,3]St. Petersburg Scientific Center of Russian Academy of Sciences
[3]St. Petersburg Clinical Hospital of Russian Academy of Sciences
[1]*eap@spiiras.ru*, [2]*ewasser@ev7987.spb.edu*, [3]*kolq@spiiras.ru*

***Abstract*— Procedures of dichotic listening of speech are considered. Random walk theory and sequential Wald analysis are proposed as a means to minimize the necessary number of stimuli presentations, keeping statistical significance of the results at an acceptable level. An algorithm of choosing of the earliest possible stopping time of experiment, when providing conclusive results is proposed. The algorithm can be propagated to all cases of binary variable, observed at medium or small sets. A number of formulae useful for prediction of the laterality index, including ones implying absolute exactness of results, are presented.**

***Index Terms*— dichotic listening, random walk, lateralization, sequential Wald analysis.**

I. Introduction

Where in the cerebrum are dominantly located speech functions? This is the question neurologists and neurosurgeons typically face when diagnosing neurological and psychiatric disorders or planning treatment on human brain structures. The simple and affordable method to preliminary answer it is to conduct dichotic listening (DL) of speech test (see [1, 2] for the review). DL test is a means to find differences of capability to process auditory information between the left and right human cerebral hemispheres. The method allows to ascertain not only laterality but also some other psychophysiological traits of the subject [3, 4]. The term "dichotic" by itself means simultaneous listening of two different signals, one of which is being presented to the left ear, the other – to the right one.

They often use 6 syllables consisting of one of six stop-consonants paired with the vowel /a/ as stimuli [5, 6, 7]. The most popular version of DL test consists of 36 (all the possible 2-tuples) dichotic stimuli presentations, or 72 ones, if we want to conform to Bergen paradigm [6, 8]. However when DL test is conducted in children or specific adults, quite often we have to cancel test when the due number of stimuli is not yet presented – as a reason of quick attenuation of attentional control of the subject. It causes a number of special difficulties; also resulting test conclusions require additional grounding [9, 10]. So when we need to predict the laterality index (LI), which is a principal result of DL test, using incomplete data, the random walk theory [11] comes in handy.

The aforementioned reason makes the sequential Wald analysis [12] interesting to us, which allows us in specific cases to determine LI using less DL test iterations than the classic statistics requires. This is possible because the set volume is not chosen a priori but being determined during DL test.

II. Material and Methods

This work is based on the analysis of DL results of 40 boys, 4–16 years old and 8 girls, 4–13 years old (patients of the children neuropsychiatric clinic) and of three healthy adults 30, 37, and 49 years old respectively. DL was conducted in clinical and laboratory conditions using the software we developed previously [13].

We presented to the subject as dichotic stimuli the following six Russian syllables: /ba/, /da/, /ga/, /pa/, /ta/, /ka/. Subject's responses were collected and preferably recognized left ear (L) and right ear (R) stimuli were counted. Resulting LI was calculated using the following formula

$$\mathrm{LI} = 100\% \cdot (R - L)/(R + L)\,. \qquad (1)$$

According to traditional psychophysiological interpretation the sign of LI determines laterality quality (left or right) and absolute value of LI determines its degree.

III. Random Walks as a Model of Dichotic Listening

When the investigator fails to present predetermined number of stimuli to the subject, short-term prediction of LI can be conveniently made using random walk theory. The easiest explanation of this idea is a graphical one.

During the DL test we will build a two-dimensional graph, after the presentation of each pair of stimuli putting the value $R + L$ on the horizontal axis and $R - L$ on the vertical one. Value of $R + L$ – is an integer greater than or equal to zero. Each step of DL test is increasing $R + L$ by 1, at the same time increasing or decreasing $R - L$ by 1. This graph can be considered as a kind of trajectory of wandering particle D($R + L$, $R - L$) [11]. The particle does move up or down on each step with probabilities p and $q = 1 - p$, respectively.

We assume that the case of the presentation of the full stimuli set matches the final position of the particle – point

B($R_N + L_N$, $R_N - L_N$). Point A($R_n + L_n$, $R_n - L_n$) is an intermediate position reached on the way from (0, 0) to B($R_N + L_N$, $R_N - L_N$).

On Fig. 1 the solid line is a trajectory formed by points representing the DL results; the dashed lines are boundaries for possible $R - L$ values.

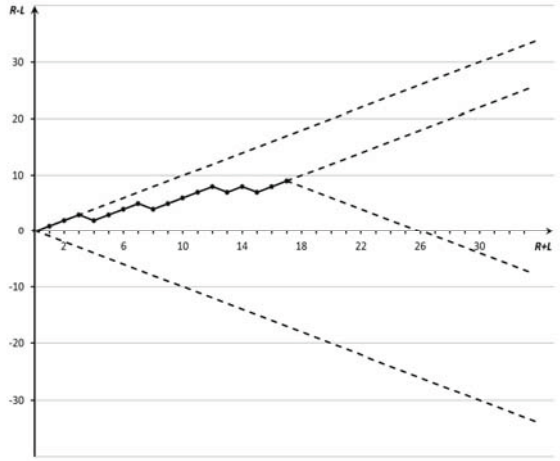


Fig. 1. Dichotic listening results of 6 year old boy. The solid line is a trajectory formed by point representing the DL results; the dashed lines are boundaries for possible $R - L$ values.

None of the possible trajectories will go beyond the drawn dashed lines going from the (0, 0) point. If at the moment $R_n + L_n$ the difference $R - L$ reached a value of $R_n - L_n$, then further movements are only possible within the sector formed by the rays going from the point ($R_n + L_n$, $R_n - L_n$) at an angle of ±45°. Value $R_n - L_n$ during k additional observations will increase or decrease by no more than k units. Then there is 100% probability that LI at the step ($R_n + L_n + k$) lies within the following interval:

$$100\% \cdot \left[\frac{R_n - L_n - k}{R_n + L_n + k}; \frac{R_n - L_n + k}{R_n + L_n + k} \right]. \quad (2)$$

The interval size does not depend on p and stimuli set size. The longer prognosis is required, the greater the interval size is.

If for some person p is 1/2, then we can calculate other probabilities, useful for LI prediction.

Let $R_n + L_n = n$, and $R_N + L_N = N$. The probability that DL in the point $R_n + L_n$, is equal to LI in the final point B($R_N + L_N$, $R_N - L_N$) is

$$P\left(\frac{R_n - L_n}{n} = \frac{R_N - L_N}{N} \right) = \frac{C_{N-n}^{0.5 \cdot (N-n)(1+(R_n - L_n)/n)}}{2^{N-n}}. \quad (3)$$

The probability that all points of the path from the starting point (0, 0) to (R_n+L_n, R_n–L_n) will be above zero is

$$(R_n - L_n)/(R_n + L_n). \quad (4)$$

if in the result $R_n \geq L_n$.

That means that module of Eq. 1 can be considered as a probability that R dominated over L during whole DL test. Also the probability that at the step $R_n + L_n$ value $R - L$ for the first time becomes equal to the positive $R_n - L_n$ is in Eq. 4.

The probability that LI reaches 0 in the length from $R + L = R_n + L_n$ to $R + L = R_N + L_N$ is

$$(R_n - L_n)/(R_N + L_N - 2L_n). \quad (5)$$

The probability that $R - L$ for the first time reaches level $R_k - L_k$, $R_k - L_k > 0$ at the moment $R_N + L_N$ is

$$(R_k - L_k)/(R_N + L_N). \quad (6)$$

The probability that $R - L$ reaches the negative $R_k - L_k$ is

$$(R_N + L_N - 2R_n)/(R_N + L_N - (R_n + L_n) - (R_k - L_k)). \quad (7)$$

In the general case when the particle D($R + L$, $R - L$) moves up and down with probabilities p and $q = 1 - p$, but p not necessarily equal q, the above probabilities are calculated using more complex formula. The particle will primarily move up ($p > q$) or down ($p < q$).

In case of walk similar to Bernoulli scheme with probabilities p и q, the probability that in the point $R + L = k$ the value $R - L$ is equal y can be calculated the following way:

$$P(R + L = k; R - L = y) = C_k^{0.5 \cdot (k+y)} p^{0.5 \cdot (k+y)} q^{0.5 \cdot (k-y)}. \quad (8)$$

Theoretical probabilities coincide with those calculated from the experimental data (in this case the DL test), only when the experiment is repeated to infinity (see The law of large numbers). In case of DL, this requirement makes irrevocable sampling of stimulus pairs impossible, as the number of the latter is limited.

IV. Sequential Wald Analysis in Dichotic Listening Test

Classical statistical analysis procedure is based on a fixed sample size: according to the results of the pre-planned number of observations the hypothesis to be tested is accepted or rejected. Pre-fixed sample size sometimes leads to unnecessary efforts to obtain a large sample. Application of adapted sequential Wald analysis [12] allows us to get a statistically significant result from the same or smaller sample sizes.

To do this it is necessary to select the acceptable probability of errors of the first α and second β kind as well as the values of p_0 and p_1 before the DL procedure. The decision on the form of lateralization is made based on the values of p in population. In the case where p is equal to or greater than p_1, the right hand lateralization hypothesis (H_0) is accepted. If $p \leq p_0$, left hand lateralization hypothesis H_1 is accepted (right hand lateralization hypothesis is rejected). In the case when $p_0 < p < p_1$, no conclusion is made and the next pair of stimuli should be presented to the subject. There p is the proportion of

trials in which the random variable X takes the value +1. A discrete random variable $X = (R_{k+1} - L_{k+1}) - (R_k - L_k)$ is obtained by $k+1$ step $(R + L = k + 1)$. It takes value $x_{k+1} = 1$ with probability p and $x_{k+1} = -1$ with probability $q = 1 - p$. As a result of the DL we obtain sample $(x_1, x_2, x_3, \ldots, x_n)$ of volume $R + L = n$ and, respectively, the sample value p, from which it possible to make assumptions about the value of p in population. It is much more convenient to be based on a sample value R. Thus, according to the method of sequential analysis, the null hypothesis H_0 is accepted, provided that

$$R \geq \left(\ln \frac{1-\beta}{\alpha} + n \ln \frac{1-p_0}{1-p_1} \right) \Bigg/ \left(\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1} \right), \quad (9)$$

and rejected (i.e. H_1 is accepted), when

$$R \leq \left(\ln \frac{\beta}{1-\alpha} + n \ln \frac{1-p_0}{1-p_1} \right) \Bigg/ \left(\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1} \right). \quad (10)$$

In case of

$$\frac{\ln \frac{\beta}{1-\alpha} + n \ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1}} < R < \frac{\ln \frac{1-\beta}{\alpha} + n \ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1}} \quad (11)$$

DL is going on and the next pair of stimuli is presented.

Note that, in contrast to the standard approach to the statistical analysis of the data it is necessary to choose not only α, but also the acceptable level of error of the second kind, i.e. β.

Using this procedure, for example, in DL of 12 years old girl yielded significant results in 16 instead of 36 presentations. The presets for the procedure are $\alpha = 0.02$; $\beta = 0.03$; $p_0 = 0.4$; $p_1 = 0.6$. That is, in particular, the probability that the results may be a mistake may not exceed 0.03.

V. Results and Conclusion

All the introduced probabilities can be used during DL as auxiliary data in the analysis of its results. For LI prediction the convenient and 100% reliable boundaries are obtained by considering all the possible trajectories of the particle D($R + L$, $R - L$). Where sequential Wald analysis allows us to reduce the necessary number of stimuli presentations, comparatively to the classical statistical analysis procedure.

The principal practical result of this study is an optimized algorithm of DL which allows the researcher to get reliable results in some conditions more quickly than the standard DL technique. Idea of the algorithm was approved by clinical practitioners we collaborate with and is about to be implemented in the end-user software.

In a theoretical plane, the approach we presented can be propagated to all cases when binary variable is observed at small and medium samples. It makes possible real-time reporting to the researcher the potential information content of the remaining iterations of the experiment.

References

[1] K. Hugdahl, "What can be learned about brain function from dichotic listening?" Rev. Esp. de Neuropsicología, vol. 2, no. 3, pp. 62–84, 2000.

[2] I. Pelletier, H. C. Sauerwein, F. Lepore, D. Saint-Amour, and M. Lassonde, "Non-invasive alternatives to the Wada test in the presurgical evaluation of language and memory functions in epilepsy patients," Epileptic Disord., vol. 9, no. 2, pp. 111–126, 2007.

[3] S. de Bode, Y. Sininger, E. W. Healy, G. W. Mathern, and E. Zaidel, "Dichotic listening after cerebral hemispherectomy: methodological and theoretical observations," Neuropsychologia, vol. 45, no. 11, pp. 2461–2466, 2007.

[4] F. Takio, M. Koivisto, L. Jokiranta, F. Rashid, J. Kallio, T. Tuominen, S. J. Laukka, and H. Hamalainen, "The effect of age on attentional modulation in dichotic listening," Dev. Neuropsychol., vol. 34, no. 3, pp. 225–239, 2009.

[5] K. Hugdahl, "Auditory laterality and brain function for speech sounds in a dichotic listening paradigm," (in Russian), Psychology. J. of the Higher School of Economics, vol. 3, no. 2, pp. 102–110, 2006.

[6] K. Hugdahl, "Fifty years of dichotic listening research – still going and going and…" Brain Cogn., vol. 76, pp. 211–213, 2011.

[7] K. Hugdahl, "Considerations in the use of dichotic listening with children," in Advances in Child Neuropsychology, vol. 1, M. G. Tramontana and S. R. Hooper, Eds. NY: Springer-Verlag, 1992, pp. 137–167.

[8] K. Hugdahl, "Dichotic listening in the study of auditory laterality," in The Asymmetrical Brain, K. Hugdahl and R.J. Davidson, Eds. Cambridge, MA: MIT Press, 2003, pp. 441–478.

[9] E. Wasserman, and N. Kartashev, "Dichotic listening: computer implementation, methodical problems, and modeling prospects," in 3rd Int. Conf. "Problems of cybernetics and informatics," (PCI'2010), Sept. 6-9, 2010, Baku, Azerbaijan, vol. 1. Baku: Elm, 2010, pp. 106–109.

[10] E. L. Wasserman, N. K. Kartashev, E. A. Popova, "Dichotic listening in children: some methodic problems and formal ascertainment of laterality significance of its results," (in Russian), in 5th Jubil. Baltic Congr. on Child Neurol., supported by ICNA [St. Petersburg, june 8–10, 2015]: Conf. Proc., V.I. Guzeva, Ed. St. Petersburg: Chelovek i ego zdorove, 2015. pp. 74–76.

[11] A. N. Shiriaev, Probability – 1, Probability – 2. (in Russian). Moscow: MTSNMO, 2007.

[12] A. Wald. Sequential Analysis. NY: John Wiley and Sons, Inc, 1947.

[13] E. L.Wasserman, and N. K. Kartashev, "Dichotic 2" [computer program], St. Petersburg (Russia): St. Petersburg Institute for Informatics and Automation of RAS, 2013.

Assistive Healthcare Home Monitoring System for Elderly People

Ali Mubarakov
School of Engineering
Nazarbayev University
Astana, Kazakhstan.
Email: amubarakov@nu.edu.kz

Yestay Zhengis
School of Engineering
Nazarbayev University
Astana, Kazakhstan.
Email: yzhengis@nu.edu.kz

Yau Hee Kho
School of Engineering
Nazarbayev University
Astana, Kazakhstan.
Email: yauhee.kho@nu.edu.kz

Abstract — **The problems faced by most senior citizens living alone at homes along with their poor health conditions have motivated the development of a home monitoring system which aims at improving their wellness and health-care. The purpose of this system is to monitor the physiological parameters in real time, together with the help from a computer vision module. Several research works on this area have been conducted, however, the combination of image processing and wearable sensors have not been widely reported. The daily activities and health parameters are to be further analyzed. In case of any abnormal or emergency situations, alarm will be triggered and urgent message will be sent to a medical doctor.**

Keywords—image processing, wearable sensors, heartbeat, blood pressure, body temperature.

I. INTRODUCTION

Technology has always been one of the key enabling factors of humanity being able to carry out all the useful initiatives and inventions throughout the ages in order to bring progress to our society. This paper is about methods and techniques that engineering can bring in order to aid elderly people in receiving qualitative medical care. One may question the significance of focusing on the elderly in terms of medical treatment troubles whilst there exists a plethora of problems spread around the world related to healthcare. The answers to this come to mind as people start realizing the fact that today every ninth person on our planet is considered as old – at age above 65 years old [1] and by the year 2050 the amount of such senior citizens will consists of 15% of the total population, thereby outnumbering the numbers of children at age under 5 [2]. More importantly, the amount of elderly people living alone across the world is quite considerable and still growing [2].

Although nowadays telemedicine can offer a wide range of health care services like on-line consulting and high-tech monitoring of a patient's health, it is still very costly and hence not affordable for most elderly people, and especially for those who live a solitary life. The cost of being inspected by a doctor via telemedicine is quite high: device and service both combined can cost the patients up to $1,600 [3] annually.

The purpose of this project is to develop an assistive system for elderly people that will be installed at their homes at a much lower price than those available commercially. The system's main features include automatic monitoring and classification of activities of elderly people, thereby the methods and approaches are based on image processing as well as sensors and actuators fields of work. The final product will be the integration of both fields aimed at producing a single unit performing a synergized work of sensory network and computer vision.

II. IMAGE PROCESSING SUBSYSTEM

The aim of the image processing subsystem is to extract the information from the video streams obtained via cameras and conduct monitoring on elders in order to determine their postures and motions. These are essential for the classification of the activity types so that alert can be sent to health giver if there are any activities that impose potential danger to the person.

The elements of computer vision in present days include image acquisition, preprocessing, feature extraction and registration [4]. Image sequences obtained from the imaging system, i.e. video cameras, are to be calibrated before putting in use. This is done in the preprocessing step: "low-level" operations are conducted on the initial images to reduce the noise levels as well as the total vast amount of incoming data. Continuous reduction of information on the image can bring out to the surface the set features uniform to disturbances, or in other words noise, distortions, ambient lighting and camera location. Last level, registration involves creating a correlation or match between those sets of features and features of known objects in a database and/or features of the following image; thereby carrying out the outcome or final hypothesis [4].

After being programmed, the image processing system could be able to handle following tasks:

- Object recognition: spotting the presence and/or position of known objects in an image;
- Tracking: coursing known objects through a sequence of images.

Object detection methods have been developing for years by programmer scientists and software engineers for the wide range of purposes. By that, today, object detection is generally achieved using the following techniques or their combinations:

- Background subtraction: detection of moving areas in an image by means of figuring out the pixel-by-pixel difference between current image and a background of reference;
- Temporal differencing: determining the absolute difference between two or three consecutive frames from the vide translation;
- Optical flow: estimating the approximation to a counted movement vector map of the pixels collected from the two consecutive frames [5].

As this project consists of assistive monitoring system at home for elderly people, the image processing module will be reading data streams from the cameras installed in rooms that are permitted by the patient. This subsystem will be outputting information about the location, posture and/or activity type. However, to be able to distinguish one feature from another, the subsystem should be provided with a database of various possible models of behavior. Although the behavior can be very unique for different kinds of persons, the approach is similar. The method compels the computer vision module to compare the components from an obtained information with the ones that are stored in the database. With an addition of some characteristic variables such as time and coordinates of the captured object the image processing part of the project could be capable of deciding whether to trigger an alert or not. The most widespread programming languages used to create image processing systems are e.g. Java, C++, python and Matlab; there are also open source computer vision libraries provided by OpenCV.

Some of the issues considered in the image processing subsystem include, for example:

- Lighting: this affects performance in dark ambiance or during night time: not all video cameras support night vision or thermal sensing. Implementation with a high cost thermal camera can increase the cost of the project dramatically;
- Facial detection: this may seem optional, however sometimes it helps to know whether the person is just resting on a chair with his eyes wide open, in case no motion is detected;
- Image quality: as this project is aimed to be low cost we have chosen to use web camera.

As mentioned a database of postures will also be developed. At the moment, we focus on the following:

- Standing
- Sitting
- Lying

After the database is created, another essential thing to do is to sort them by specific venue within the house and time slots. In other words, it means context in relation to the posture will be used in terms of location and time of occurrences. For example, consider an unlikely scenario of a person lying in the kitchen. This probably means that an accident has occurred, so the system will send an alarm and notify the care giver. The same is true with reference to time where the day can be divided into several time slots to help the system detect unusual activities that possibly may not occur during the given time slot. For example, sitting motionlessly at 3:00 AM in the bedroom is unusual (as shown in Figure 1), so the image processing module with check with sensory module to determine if help is needed.

Figure 1. Detection of a sitting person with no motion detected.

III. Sensors Subsystem

Wearable sensors are necessary for the collection of physiological data from elderly in our system. There are three main health parameters to be measured: heartbeat rate, body temperature and blood pressure. Figure 2 shows a diagram of the subsystem architecture.

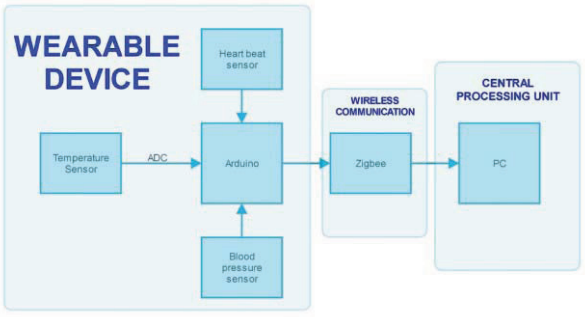


Figure 2. Sensor subsystem block diagram.

The sensor network is designed as a star structure. In such a structure the architecture is centralized at one node, where all data are manipulated. In this project information from sensors are collected by an Arduino Nano microcontroller board.

Heart beat

In medical field there are several methods to measure heartbeat. Conventionally, sound and pressure sensors are used to detect the pulses. In this project, an optical technique known as photoplethysmography (PPG) method was implemented. This is a non-invasive technique, and uses infrared light emitter to illuminate the body surface and photodetector to determine small changes in the transmitted light [6]. These changes are related to the variation of blood volume inside the tissue. To obtain synchronous signal with the heartbeat, the signal is filtered and amplified to get a clean PPG waveform [6].

An IR emitter and detector TCRT5000 sensor is applied as shown in Figure 3. As a detector, this sensor uses phototransistor, a semiconductor device that converts light into current. Therefore, whenever the light intensity changes, the output current will also changes. Partition between the emitter and detector is crucial in the case of reflective sensor. In addition, this sensor is placed on to a leaded package which blocks visible light.

Figure 3. An TCRT5000 sensor is embedded onto a wearable fabric for measuring heartbeat.

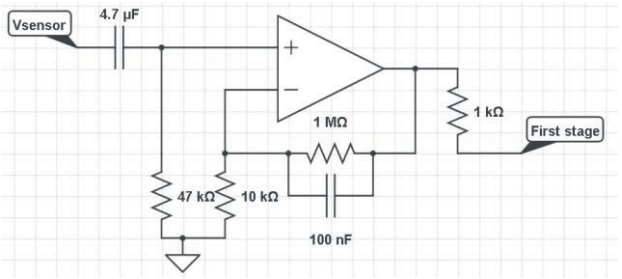


Figure 4. First stage of filtering and amplification.

The output from the sensor is very weak and the oscilloscope could not even detect any pattern of pulses due to noises. Therefore it was necessary to amplify that signal and filter it. As shown in Figure 4, the sensor signal enters the first stage of filtering which is a passive high-pass filter. This is done to block DC component and suppress low frequency noises. The cut-off frequency of this filter is at 0.346 Hz. The needed pulse information is carried by the weak AC components of this wave, and must be amplified. Therefore signal enters an active low-pass filter (LPF) based on Opamp (operational amplifier) configured as a non-inverting state and has a gain of 100. Finally, the value from first filtering stage versus time plot is obtained, as shown in Figure 5 below.

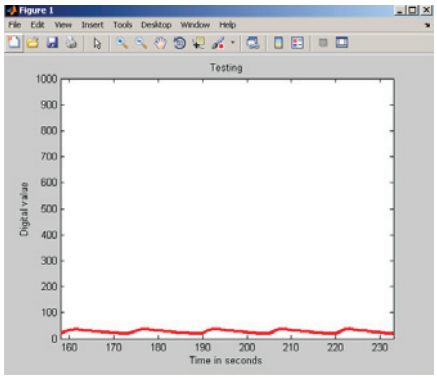


Figure 5. Digital value versus time graph.

Before connecting to the Arduino, signal from the first stage filtering enters the second stage. The circuit is a replica of the first filtering stage. It again clears the signal from DC components and amplifies it. The resultant waveform from second stage is shown on Figure 6.

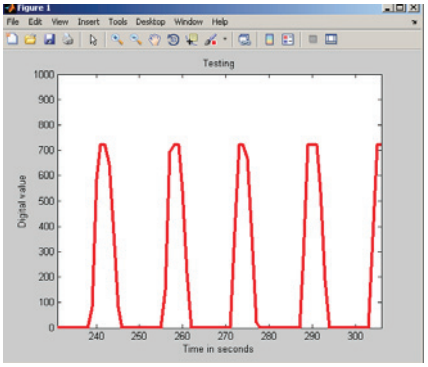


Figure 6. Second stage of filtering.

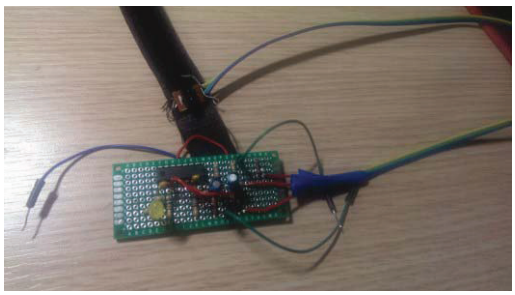


Figure 7. Sensor for measuring heartbeat rate.

Blood pressure

There are two main methods to measure blood pressure, using aneroid and oscillometric devices. The latter one is considered as automatic device where measurements are done automatically and results are displayed [7]. Therefore, this method was chosen to monitor the health parameter.

The working principle of a blood pressure monitor is simple. The cuff is inflated and deflated electronically. The pressure wave generated by the brachial arterial wall is sensed by a transducer in the device. As the cuff pressure is released the pressure wave amplitude increases and peaks at the mean intraarterial pressure (MAP), then decreases again. The point of maximum amplitude (the MAP) is detected by the oscillometric device [8]. A commercial blood pressure monitor as shown in Figure 8 was used where reading from the unit is extracted.

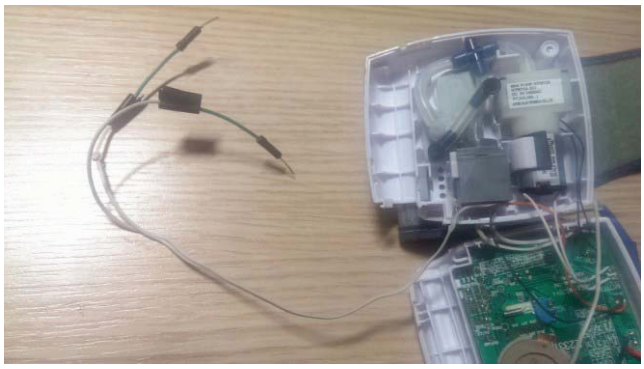


Figure 8. Extraction of data from the solenoid valve.

By default the blood pressure monitor is switched on by pushing an ON button mechanically. A relay circuit as shown on Figure 9 was used to automatically control the monitor. The two ends of the switch were connected to the 5V relay where the relay itself was connected to a Arduino digital pin. The logic states that whenever the pin goes HIGH, relay switch is closed, going from NO (normally open) to NC (normally closed) state. This by itself connects two wires from the tonometer switch, "pushing" the power button on. This way of solution gives an opportunity to automatically control the switch using Arduino digital pins.

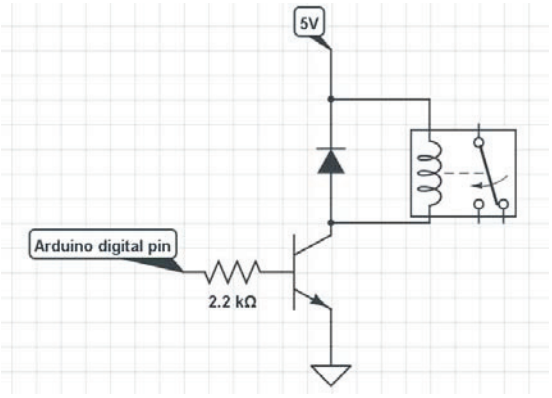


Figure 9. Relay connection circuit.

Temperature

Thermistor is a resistor which changes its resistance with changing temperature [9]. Technically all resistors are considered as thermistors. However, their changes are minor and it is difficult to measure the changes in temperature. Therefore, thermistors are a special resistor that has significant changes in resistance as temperature varies [9]. In order to measure the temperature value, resistance must be measured first. Arduino has no ohmmeter function, however, this can be done by detecting changes in voltage drop with the help of an analog-to-digital convertor (ADC). Therefore, it is needed to convert changes in resistance of thermistor to the changes of voltage. This was executed by using a potential divider circuit as shown in Figure 10.

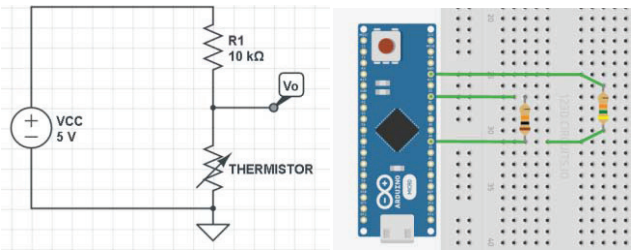


Figure 10. Schematics and connection to Arduino Nano

The voltage drop at V_0, that will be transmitted to the Arduino analog pin is:

$$V_0 = R / (R+10K)*Vcc \quad (1)$$

where Vcc is the voltage source from Arduino 5V pin. From the formula above we can relate changes in resistance to the voltage changes. Next step is to connect V_0 to the Arduino ADC. Converter of Arduino remaps the input values from 0 to 1023 range. Therefore, 5V corresponds to 1023, and 0V to 0 respectively. Finally it can be written as (V_0=Vi from Arduino perspective):

$$ADC\ value = Vi*1023 / Vcc \quad (2)$$

Combining equations (1) and (2), and solving for the needed resistance, we have:

$$ADC\ value = R / (R+10K) * 1023$$

$$R = 10K / (1023/ADC - 1) \quad (3)$$

The value obtained from equation (3) for R gives us varying resistance of thermistor that can be read by Arduino. Next step is to find dependence between these changes in resistance with temperature variations of body. For this reason a simplified version of Steinhart-Hart equation is used:

$$\frac{1}{T} = \frac{1}{T_0} + \frac{1}{B}\ln\left(\frac{R}{R_0}\right) \quad (4)$$

where T_0 is the normal body temperature (36.6˚C=309.75K), B is the thermistor coefficient, R_0 is the resistance at T_0, and R is the thermistor resistance. Figure 11 shows the reading of room temperature using these equations.

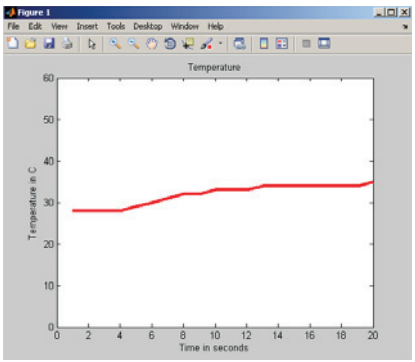


Figure 11. Temperature versus time graph results.

Communication

The communication of PC and Arduino board was done with the help of a Bluetooth HC-05 module. Bluetooth is great for transmitting data in a 10-100m range [10]. The most important part is that Arduino treats Bluetooth as serial data connection. This implies that Serial Library is also usable for Bluetooth, which is very convenient and simple. After pairing, completed data transfer can be started. In order to read data on PC, TeraTerm terminal was used, connected to a COM port of the Bluetooth adapter. The results are exactly the same as when using a cable connection because both of them were using serial communication. Next step is to save this data. In TeraTerm software there is a useful function called 'Log', which saves incoming data to a text file instantaneously, and updates it if new serial data is received as shown in Figure 12.

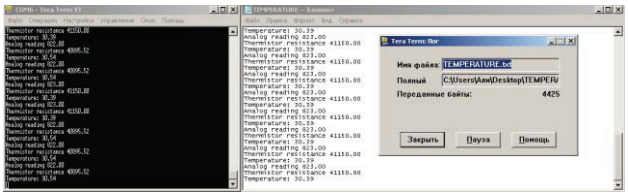


Figure 12. Saving the sensor data onto a text file.

IV. System Integration

In this part outputs from two subsystems are now considered in one loop of Python code. As it was mentioned before, from sensors subsystem there are three outputs (heartbeat, blood pressure, temperature). They are saved in a constantly updated text file, indicating the current health condition of the patient. From image processing subsystem there are also three outputs, indicating the current posture of the patient (standing, sitting and lying), and how long the patient is in a particular position.

The Python loop itself comprises of IF/ELSE statements where it monitors the vital physiological signs. The normal temperature value is set to 36.6˚C, heartbeat is considered normal between 60 and 85bpm, systolic value between 100 and 140 and diastolic from 50 to 90. These threshold parameters can be adjusted differently for specific patient. If one of these parameters will go beyond the acceptable range, immediately an email message will be sent to an appropriate medical service. This is executed by importing smtp library to the Python main code.

Besides this, the second parameter being monitored is the posture and its duration. Messages will be sent if the person:

- is lying on floor for more than 10 seconds;
- is sitting without standing for more than 1 hour.

In such cases, the message also includes the last measured parameters of the patient.

V. Discussion and Analysis

There are some clear advantages of the proposed system in terms of cost and availability. The entire project costs not more than $170. The equipment and components can be easily found in local markets, making the project available. Cost and availability factors can have a huge impact on elderly. There are indirect benefits to the aged society as travelling to the hospitals and queue issues can be eliminated.

The main problem of this project is its functionality. The outputs from integration part strongly depend on the outputs from image processing subsystem. This subsystem has a very high potential, and can be improved a lot: detecting more actions and tracking more objects.

Communication with medical service through emails was an obliged design specification. There was a trial to connect 3G modem to the computer, and send short messages (SMS) to a doctor's individual phone number. This solution would be more effective. However, one obstacle encountered was no access to an AT commands of this modem. Access could be opened only by paid hacks or buying special expensive GSM device (~$70) which was not cost efficient for this project. However, if AT commands function could be accessed, it would give chance for doctor to request latest health parameters, see the history of vital signs and activity done.

There were also interesting findings at the completion of this project. As it was discovered, blood pressure can be measured straightly from heart beat output signal waveforms. The output signal from first filtering, where DC component still exists, also carries information about systolic and diastolic pressure.

VI. Conclusion

The developed assistive healthcare home monitoring system can monitor current health condition and prevent elderly from long-term diseases. The dramatically increasing population of aged people makes this project topical in current situation. Taking in fact the financial conditions of local elderly in Kazakhstan, main focus in designing the project was to make low cost and available for aged people assistive system.

This project has a very high potential and can be improved. In this project there was shown that main health parameters can be monitored in a real-time basis. The simple working method of communication was implemented with further data analysis (IF/ELSE loop).

There are also several future works and challenges. The first is image processing part, to obtain clear results and track as many objects with different interference and relations between them. The second is sensors part, it can be added more health parameters to be measured, as ECG and blood sugar level. Then IF/ELSE loop can be improved, depending on outputs from image processing part. These improvements can then even detect diseases and send possible diagnoses to the doctor. Other sensing techniques using radio frequency can also be incorporated [11, 12]. Finally, communication system with doctor can be done in more effective ways.

VII. References

[1] Population Ageing and Development, United Nations, Sept. 2012. [Online].Available: http://www.un.org/esa/population/publications/2012WorldPopAgeingDev_Chart/2012PopAgeingandDev_WallChart.pdf

[2] Global Health and Aging, 1st ed, United Nations, 2011, pp. 1-25. [Online]Available: http://www.who.int/ageing/publications/global_health.pdf?ua=1

[3] 'An Innovative Technology Profile: Home Telehealth', NEHI, 1, pp.10-13, 2015.

[4] El-Zabadani, H., Helal, S., Mann, W. and Schmaltz, M. 'PerVision: An integrated Pervasive Computing/Computer Vision Approach to Tracking Objects in a Self-Sensing Space', in Proceedings of the Fourth International Conference on Smart Homes and Health Telematics (ICOST), 2006.

[5] Ramon, R., et. al. 'Image Processing to Detect and Classify Situations and States of Elderly People', pp.164-170, 2011.

[6] Mazima, K., et al. 'Design of low cost blood pressure and temperature interface', IJESE, vol. 1, issue-10, 2013.

[7] Graves JW., 'Clinical decision-making in hypertension using an automated measurement device', Journal of Human Hypertension, vol.17, pp. 823-827, 2003.

[8] Subhash, O., Vaidya R.J. 'Arterial blood pressure measurement and pulse wave analysis', IOSR-JECE, vol. 4, pp. 41-46, 2012.

[9] Recktenwald G., 'Temperature measurement with a thermistor and arduino', 2013

[10] Monson, H., 'Bluetooth technoglogy and implications', 1999.

[11] Kho, Y.H, et. al. 'A passive wireless tracking system for indoor assistive monitoring', IEEE Int. Conf. Sensing Technology, pp. 138-142, 2015.

[12] Kho, Y.H., et. al. 'Exploiting RF signal attenuation for passive indoor location tracking of an object', IEEE Int. Conf. Computer, Communications and Control Technology, pp. 152-156, 2015.

An Adaptive Network-Based Fuzzy Inference System for Estimating the Duration of Medical Services: A Case Study

İhsan Hakan Köksal
Department of Computer Engineering
Yaşar University
Izmir, Turkey
hakan.koksal.7@gmail.com

Mete Eminağaoğlu
Department of Computer Science
Dokuz Eylül University
Izmir, Turkey
mete.eminagaoglu@deu.edu.tr

Buse Türkoğlu
Department of Business Engineering
Yaşar University
Izmir, Turkey

***Abstract*—The aim of this study is twofold. First one is to derive a feasible numerical prediction model for the duration of medical services in a medical institution that could be used by the hospital management within their quality assurance and improvement processes. The second aim is to develop a freeware software tool, which implements an adaptive artificial neural network-based fuzzy inference system with a user-friendly interface that can be effectively used by both technical and non-technical people. Some promising results have been obtained which show that both of these objectives have been successfully achieved to some extent.**

***Index Terms*—ANFIS, duration estimation, medical services, numerical prediction.**

I. Introduction

In recent years, information and communications technology (ICT) has proven to be successful in various fields of medical science and healthcare. ICT plays an important role in improving the quality of the healthcare services, the effectiveness and efficiency of hospital management and also aids in the improvement of social health. The continual progress in ICT such as hardware, software, digital systems and networks, electronic devices and their integration with business processes in medicine provides new solutions for healthcare services, such as electronic health cards [1]. There are also a great amount of scientific studies focusing on the improvement of the quality of services and business processes in clinics, laboratories and similar medical organizations such as the analysis and improvement of duration or turnaround times [2, 3, 4]. Reliable prediction or estimation of the durations or turnaround times among such services is also crucial, which might help the management to make proper decisions in order to decrease these durations or to find out the underlying causes that affect such durations. This has been one of the primary reasons that motivated us to carry out this study.

There are methodologies and theoretical models for numerical estimations or predictions that are used among different areas of applied sciences, business, economics and daily life. In the recent decades, such methodologies are also used in data mining and machine learning where new alternative models, algorithms and techniques have also been developed. Data mining is a globally accepted discipline that can be defined as the extraction of useful information from large data sets or databases. It lies at the intersection of statistics, machine learning, data management and databases, pattern recognition and artificial intelligence [5]. On the other hand, machine learning provides the technical basis of data mining and it is accepted as a discipline that is used in cooperation with data mining where both disciplines have intersection points [6, 7]. Some of the machine learning algorithms are based on heuristic models and some are based on statistical theories and methodologies or a mixture of both [7, 8].

Artificial neural networks (ANNs) are one of the well-known heuristic models that are used in machine learning and soft computing. They are computational models inspired by an animal's central nervous systems and the brain. ANNs are used to estimate or approximate functions that can depend on a large number of inputs and they can be used either for classification, clustering or numerical prediction [9]. ANNs are generally presented as systems of interconnected neurons or nodes that compute values from inputs, and they can be adaptively used for supervised, semi-supervised or unsupervised learning. There are various models among ANNs such as feedforward networks, recurrent networks, radial basis functions, self-organizing maps, convolutional neural networks and deep learning [10, 11]. ANNs can be trained to learn from the input data and they are being used to solve a wide variety of difficult tasks such as computer vision, time-series prediction, forecasting, computer aided design, decision making, time-series prediction, pattern and speech recognition [11, 12, 13].

Fuzzy sets and fuzzy inference systems are known as another type of computational methodology that are used in expert systems, optimization, robotics, decision-making, artificial intelligence, medical researches, image processing and soft computing. They can also be used in conjunction with machine learning models and methodologies. Fuzzy sets and fuzzy systems are based on fuzzy logic theory, which manifested the concept of fuzziness [14]. Fuzzy logic model combines the concepts of crisp logic and the Lukasiewicz sets by defining membership function [15] [16]. A fuzzy set extends the notion of standard set by enabling degrees of membership of an element

in the standard set where this extension is measured by real numbers having values between 0 and 1. A fuzzy system accepts continuous or discrete numerical values as input and then translates these into linguistic terms such as “high”, “cold”, “slow”, etc. This process is known as “fuzzification”. Fuzzy rules map these linguistic terms onto similar output linguistic terms. The output linguistic terms are translated into an output number which is known as “defuzzification” [17]. Fuzzy inference systems can be described simply in three stages; fuzzification, rule evaluation and defuzzification.

Adaptive network-based fuzzy inference system (ANFIS) is a kind of artificial neural network model that is based on Takagi–Sugeno fuzzy inference system which combines artificial neural networks and fuzzy logic in a single framework [18]. The inference system in ANFIS corresponds to a set of fuzzy rules that have learning capability to approximate non-linear functions. Since ANFIS is used and implemented as the basic model in this study, it is explained in more detail in sections Section 2 and Section 3.

II. Adaptive Network-Based Fuzzy Inference System

Artificial neural networks can be integrated with fuzzy inference systems in order to establish hybrid models. One of these models in hybrid neuro-fuzzy systems is adaptive network-based fuzzy inference system (ANFIS), which is usually constructed in five layers. A five-layer ANFIS architecture with two inputs, one output and four rules is denoted in Fig. 1 [19].

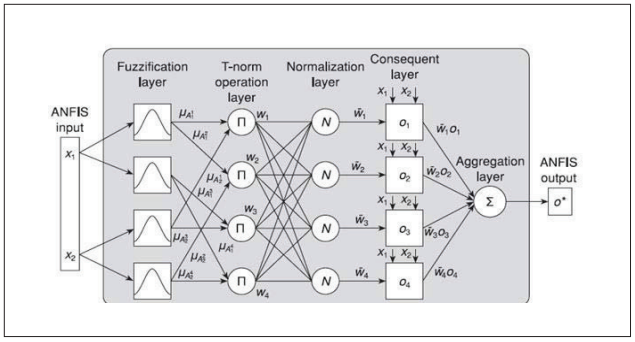


Fig. 1. Example of a five-layer ANFIS architecture.

ANFIS is an architecture for tuning fuzzy system parameters from input/output pairs of data. The fuzzy inference process is implemented as a generalized artificial neural network, which is then tuned by gradient descent techniques [20]. However, it could also be tuned by least squares method or some other hybrid methodologies [21]. ANFIS represents a Takagi-Sugeno type fuzzy system, where the fuzzy rules take the following form [19]:

$$R_p: If\ x_1 is A_1^p \dots and\ x_n is\ A_n^p\ Then\ o_p = \alpha_0^p + \alpha_1^p x_1 + \cdots\ \alpha_n^p x_n \quad (1)$$

In Eq.1 above, x_i is ith input linguistic variable in the antecedent part of the pth rule with i = 1… n and $A_i{}^p$ is the linguistic label associated with it in that rule. $A_i{}^p$ has its associated fuzzy membership function given by μ ($A_i{}^p$) (x_i). In the consequent part of Eq.1, o_p is the output of the pth rule and $\alpha_0{}^p$, $\alpha_1{}^p$ … $\alpha_n{}^p$ are the Takagi-Sugeno parameters.

In the first layer of the ANFIS architecture (fuzzification layer), nodes are calculated by multiplying inputs with their respective fuzzy membership values using a fuzzy membership function as follows [20]:

$$O_i^p = \mu_{A_i^p}(x) \quad (2)$$

In Eq.2, x denotes the input to the node, $A_i{}^p$ is the linguistic label associated with this node function and $O_i{}^p$ is the output function of that node. The fuzzy membership values can be calculated by various different of membership functions such as Trapezoid, Gaussian bell and Gaussian curve. Gaussian bell membership function is given in Eq.3, where c is the center, a is the width and b is fuzzification factor for the pth rule [19]. This function is also used in this study.

$$\mu_{A_i^p}(x_i) = \frac{1}{1 + \left(\left(\frac{x_i - c_p}{a_p}\right)^2\right)^{b_p}} \quad (3)$$

T-norm operation layer stores the rules where each rule is represented by one node. Each node is connected to those nodes in the previous layer that forms the antecedent of the rule. For a node r, the inputs are degrees of membership functions, which are multiplied through a T-norm fuzzy multiplication operator $\otimes$ to determine the fire strength w_p (degree of fulfillment) of that rule. This operation is shown in Eq. 4 where total number of rules is set to four.

$$w_p = \mu_{A_i^p}(x_1) \otimes \mu_{A_j^p}(x_2)\,, i = 1,2;\ j = 1,2; p = 1,..4 \quad (4)$$

Normalization layer is the layer where the nodes are calculated by dividing each node by total output. These nodes are called normalized firing strengths. In other words, the relative degree of fulfillment for every rule R_p is calculated [19]. This calculation is shown in Eq. 5 where $\overline{w_p}$ denotes the normalized firing strengths.

$$\overline{w_p} = \frac{w_p}{\sum_{p=1}^{4} w_p} \quad (5)$$

The nodes in the fourth layer (consequent layer) are connected to all input nodes and with exactly one node in the previous layer. Each node in the fourth layer computes the output for the rule. In other words, it computes the consequent part of the fuzzy rule. This is denoted in Eq. 6 where $\overline{w_i}$ is the output of third layer, and $\{p_i, q_i, r_i\}$ is the parameter set.

$$O_p^4 = \overline{w_p} O_p = \overline{w_p}\left(\alpha_0^p + \alpha_1^p x_1 + \alpha_2^p x_2\right) \quad (6)$$

The overall output is calculated in the final and fifth layer by the summation of all nodes’ signals coming from the previous layer which is given in Eq. 7. This single output node is the aggregated output of all the nodes.

$$O^5 = \sum_{p=1}^{4} \overline{w_p} \left(\alpha_0^p + \alpha_1^p x_1 + \alpha_2^p x_2\right) \quad (7)$$

The learning mechanism and the update of weights in ANFIS is established by the feed forward neural network mechanism. The differential error and corresponding weight updates in feed forward neural network is usually achieved by backpropagation with gradient descent methodology. The linear combination or input function for each of the nodes / units within the hidden layers and the output layer is given in the Eq. 8 which is generally used in feed-forward neural networks [22].

$$I_j = \sum_i W_{ij}\ O_i + \theta_j \quad (8)$$

In Eq. 8, for each unit j in the hidden / output layer, the net input I_j will be calculated where W_{ij} is the weight of the connection from unit i in the previous layer to unit j; O_i is the output of unit i from the previous layer; and θ_j is the bias value of the unit. It should be noted that the weights and biases in the neural network are usually initialized to small random numbers ranging between −1.0 to 1.0, or between −0.5 to 0.5.

A non-linear activation or output function is used to calculate the output value within each unit in the hidden layer or the output layer. This output activation function is usually chosen as sigmoid function, which is given in Eq. 9, where the net input to unit j is denoted by I_j, the output of unit j is denoted by O_j [11].

$$O_j = \frac{1}{1 + e^{-I_j}} \quad (9)$$

When multilayer feed-forward neural network models are used for numeric prediction, a single node / unit is used in the output layer and the last output value derived within the output layer produces the predicted value of that instance. In fact, this output node corresponds to the ANFIS output node which is explained previously. The predicted value might usually be different from the original value and this is defined as the error within the prediction. There might also be some errors within unit in each hidden layer where these errors must be calculated and the weights must be updated iteratively after finding out the errors. Back-propagation with gradient descent is one of the methods that can be used in multilayer feed-forward neural networks to calculate the errors and new weights [9, 11]. This method is also used in the neural network mechanism within ANFIS. The back-propagation error calculation for the output layer is given in Eq. 10 where j denotes the output node, O_j is the predicted value and A_j is the actual value for that record or instance.

$$Err_j = O_j\left(1 - O_j\right)\left(A_j - O_j\right) \quad (10)$$

The weighted sum of the errors of the units connected to unit j in the next layer must be included in order to calculate the error of a unit j within any hidden layer. The error of a hidden layer unit j is shown in Eq. 11 where W_{jk} is the weight of the connection from unit j to a unit k in the next higher layer, and Err_k is unit k' s error.

$$Err_j = O_j\left(1 - O_j\right) \sum_k Err_k W_{jk} \quad (11)$$

The weights and biases are updated within each round of epoch (iteration) within the training process of an artificial neural network by using these error values. The weight update is given in Eq. 12 where $\Delta W_{ij(k-1)}$ is the change in weight W_{ij} from the previous iteration *(k-1)*, W'_{ij} is the updated weight value for the k^{th} iteration, Err_j is unit j' s error, O_i is output value of unit i and l stands for the learning rate of the neural network.

$$W'_{ij\,(k)} = W_{ij\,(k)} + (l) Err_j O_i + \alpha\left(\Delta W_{ij\,(k-1)}\right) \quad (12)$$

It should be noted that in Eq. 12 there is also a constant α that is known as the momentum parameter. This momentum parameter is used in order to increase the learning rate while avoiding the danger of instability in artificial neural networks [9].

III. Case Study

The design, implementation and tests of this research was carried out as a case study within one of the private hospitals in Istanbul, Turkey. Due to the confidentiality and legal concerns of hospital management, the name of the hospital is not explicitly given in this study. The data were collected from the health information system's database. It was made up of a two-year period of data that were entered to the system whenever the patients made an appointment. It should be mentioned that only a random subsample of the whole data set was provided to us was. There were a total of 28182 records with 9 different attributes (fields) such as "Patient ID", "Department Name", "Time and date of patient admission", "Time and date of preliminary medical examination", "Specimen collection time and date", "Laboratory sample receipt time and date", "Time and date of report for test results", and so on. A sample of the data set is given in Fig 2.

Patient ID	Department Name	Patient registration Date / Time	Request for a physician service Date / Time	Specimen collection Date / Time	Sample receipt Date / Time	Report of test results Date / Time	Delivery of report to patient Date / Time	Patient's visit to physician Date / Time
104333	Emergency	1.6.2013 07:08	1.6.2013 07:08	1.6.2013 07:09	1.6.2013 07:13	1.6.2013 07:18	1.6.2013 07:18	1.6.2013 07:18
29210	Cardiology	1.6.2013 08:52	1.6.2013 09:32	1.6.2013 09:49	1.6.2013 10:08	1.6.2013 11:12	1.6.2013 11:13	1.6.2013 11:14
104167	Cardiology	1.6.2013 08:54	1.6.2013 09:05	1.6.2013 09:05	1.6.2013 09:24	1.6.2013 09:52	1.6.2013 10:20	1.6.2013 10:21
27047	Urology	1.6.2013 10:09	1.6.2013 10:25	1.6.2013 10:26	1.6.2013 11:22	1.6.2013 12:06	1.6.2013 12:07	1.6.2013 12:09
15497	Psychiatry	1.6.2013 10:39	1.6.2013 11:32	1.6.2013 11:36	1.6.2013 12:06	1.6.2013 13:10	1.6.2013 13:12	1.6.2013 13:13
57158	Cardiology	1.6.2013 10:56	1.6.2013 11:18	1.6.2013 11:19	1.6.2013 11:28	1.6.2013 12:13	1.6.2013 12:21	1.6.2013 12:25
104425	Emergency	1.6.2013 18:00	1.6.2013 18:08	1.6.2013 18:45	1.6.2013 18:50	1.6.2013 18:57	1.6.2013 19:02	1.6.2013 19:02
53548	Emergency	1.6.2013 18:21	1.6.2013 18:21	1.6.2013 18:22	1.6.2013 18:31	1.6.2013 18:37	1.6.2013 18:37	1.6.2013 18:37
89423	Nephrology	3.6.2013 08:43	3.6.2013 08:47	3.6.2013 09:04	3.6.2013 09:33	3.6.2013 10:36	3.6.2013 11:09	3.6.2013 11:09
104571	Emergency	3.6.2013 08:52	3.6.2013 08:54	3.6.2013 08:58	3.6.2013 09:15	3.6.2013 09:26	3.6.2013 09:26	3.6.2013 09:26
104264	Check-Up	4.6.2013 08:12	4.6.2013 08:34	4.6.2013 08:40	4.6.2013 09:25	4.6.2013 11:39	4.6.2013 11:34	4.6.2013 11:35

Fig. 2. An excerpt from the original data set.

A. Materials and Methods

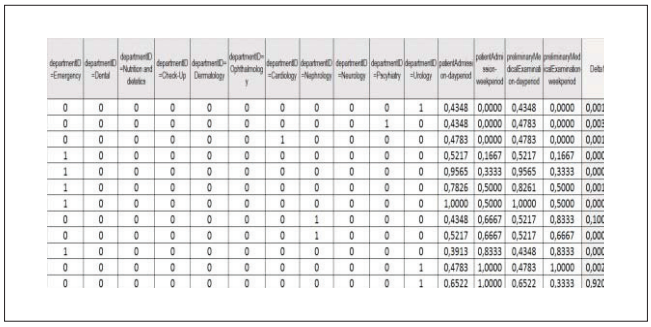
The primary objective of quality assurance department of medical services was to analyze and obtain a reliable model to estimate the duration between preliminary medical examination and patient's admission. Hence, during the data analysis process, some of the attributes and instances were excluded from the data set. Some of the records had null or duplicate data due to data entry errors and these records were also discarded from the data set. As a result, a total of 5717 records were obtained and used in this study.

Some necessary data preprocessing methods were also used in order to make input data appropriate both for ANFIS and the numerical prediction. "Department name" attribute was originally a nominal attribute having eleven different types of categorical values such as "Urology", "Cardiology" and so on. Due to the fact that ANFIS and artificial neural network algorithms can only accept and use numerical attributes, "Department name" attribute was transformed into dummy variables. This method is known as dummy coding where each different value of a categorical attribute, one input unit was generated and encoded with 0 and 1 values [22]. The duration between "Time and date of patient admission" and "Time and date of preliminary medical examination" is calculated by converting their values into terms of seconds and then calculating their difference and deriving a new attribute which is named as "Delta1". This attribute shows the duration between "Time and date of patient admission" and "Time and date of preliminary medical examination" in terms of seconds for each record. Hence, this "Delta1" is the numerical attribute that needs to be accurately estimated by ANFIS and other machine learning algorithms.

"Time and date of patient admission" and "Time and date of preliminary medical examination" were also preprocessed. The day of the week and the hourly periods were valuable information for the hospital management. Thus, these two attributes which were originally in date time format (MM.dd.yyyy hh:mm) were converted into "day of week" (ranging between 1 and 7) and "hour of day" (ranging between 1 and 24) integer values. In order to normalize the numerical variables and also to make it suitable for ANFIS and ANN models, min-max normalization [22] was used and all of the numerical attributes were min-max normalized ranging between 0 and 1. Hence, the final data set was established with 5717 instances and 16 numerical attributes. A sample from the final version of the data set is shown in Fig 3.

It should be noted that a subsample data set with 900 instances was produced out of 5717 instances and this subsample was also used in the experiments. Random sampling without replacement method was used to generate the subsample set [23]. The sampling size was chosen according to the following criteria:

- confidence interval (accepted margin of error) = 3%
- confidence level = 95%



departmentID =Emergency	departmentID =Dental	departmentID =Nutrition and dietetics	departmentID =Check-Up	departmentID= Dermatology	departmentID= Ophthalmology	departmentID =Cardiology	departmentID =Nephrology	departmentID =Neurology	departmentID =Psychiatry	departmentID =Urology	patientAdmission-dayperiod	patientAdmission-weekperiod	preliminaryMedicalExaminati on-dayperiod	preliminaryMedicalExamination weekperiod	Delta
0	0	0	0	0	0	0	0	0	0	1	0,4348	0,0000	0,4348	0,0000	0,001
0	0	0	0	0	0	0	0	0	1	0	0,4348	0,0000	0,4783	0,0000	0,003
0	0	0	0	0	0	1	0	0	0	0	0,4783	0,0000	0,4783	0,0000	0,001
1	0	0	0	0	0	0	0	0	0	0	0,5217	0,1667	0,5217	0,1667	0,000
1	0	0	0	0	0	0	0	0	0	0	0,9565	0,3333	0,9565	0,3333	0,000
1	0	0	0	0	0	0	0	0	0	0	0,7826	0,5000	0,8261	0,5000	0,001
1	0	0	0	0	0	0	0	0	0	0	1,0000	0,5000	1,0000	0,5000	0,000
0	0	0	0	0	0	0	1	0	0	0	0,4348	0,6667	0,5217	0,8333	0,100
0	0	0	0	0	0	0	1	0	0	0	0,5217	0,6667	0,5217	0,6667	0,000
1	0	0	0	0	0	0	0	0	0	0	0,3913	0,8333	0,4348	0,8333	0,000
0	0	0	0	0	0	0	0	0	0	1	0,4783	1,0000	0,4783	1,0000	0,001
0	0	0	0	0	0	0	0	0	0	1	0,6522	1,0000	0,6522	0,3333	0,920

Fig. 3. An excerpt from the final data set.

B. Design and Implementation

The architecture and design of the tool used in this study was mostly based on the ANFIS model that was described in the second section of this article. The tool was implemented and coded with Microsoft Visual C# programming language in MS Visual Studio 2015 environment. ANFIS was designed to operate with five layers, which is similar to the model given in section 2. Takagi-Sugeno type fuzzy system was used for the fuzzy rules. Since the actual input data consisted of features showing the hourly period within a day and the day of week, three linguistic levels were designed and used for fuzzification: "Early", "Medium" and "Late". Two alternative rule models were implemented and used in the experiments. One of the models was constructed with 5 different constraints for the antecedent part of the rules and the other with 10 different constraints. For instance, some of the antecedent parts used in the rules were as follows:

- IF Department name is "Cardiology" AND patientAdmission-dayperiod is "Late" AND preliminaryMedicalExamination-dayperiod is "Early" AND preliminaryMedicalExamination-weekperiod is "Early" THEN …
- IF Department name is "Neurology" AND patientAdmission-dayperiod is "Late" AND patientAdmission-weekperiod is "Medium" AND preliminaryMedicalExamination-dayperiod is "Late" AND preliminaryMedicalExamination-weekperiod is "Early" THEN …

Gaussian bell function was implemented and coded for the fuzzy membership function and T-norm operators were used. Least squares and backpropagation with gradient descent were both used for the weight updates. All the necessary data preprocessing operations such as min-max normalization and nominal-to-numeric transformation by dummy coding were also implemented in our ANFIS tool. In addition, most of the parameters used in ANFIS architecture can be interactively and easily set or changed from the graphical user interfaces provided with the tool. Some of the screen shots of the tool developed in this study is given in Fig. 4 and Fig. 5.

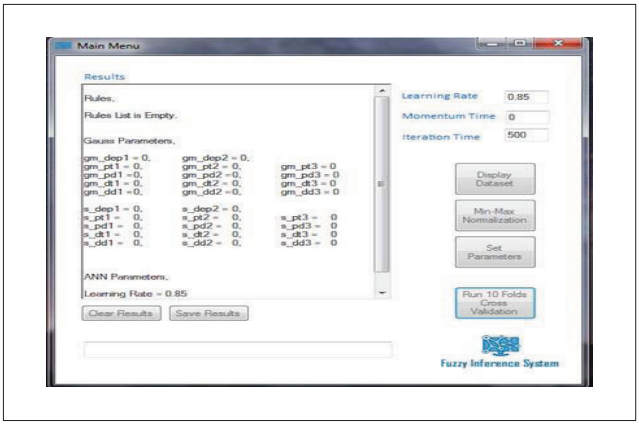


Fig. 4. Main screen of ANFIS tool.

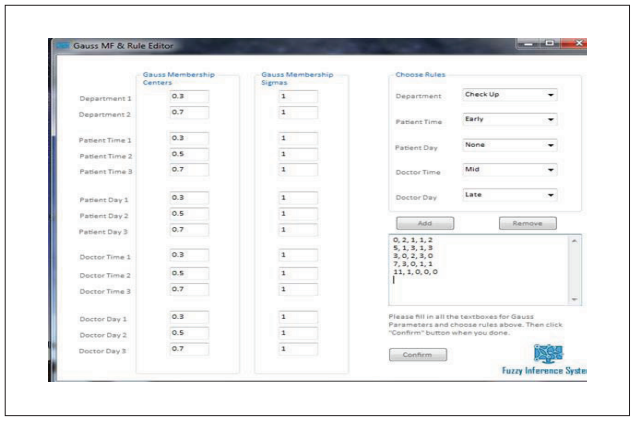


Fig. 5. Fuzzy membership function and rule editor screen.

IV. Experiments and Results

All of the experiments and tests were conducted on the same hardware platform with a 64-bit architecture, Intel Core i7 2.60 GHz microprocessor and 8 gigabytes of RAM. Weka version 3.7.17 data mining software was used for all of the other available machine learning algorithms that were comparatively tested with our ANFIS model. In all of the experiments, 10-folds cross-validation methodology was used. Two different data sets, one with 5717 instances and one with 900 instances were used. Root mean square error (RMSE) was used to measure and compare the estimation performances of the algorithms for "Delta1" attribute. It should be noted there are several options for tuning the parameters of the algorithms in Weka and we also tested these algorithms with different parameters as well as their default values where the best results are included in the tables.

The algorithms in Weka that were included in the experiments and their abbreviations are as follows:

- MLP: Multilayer Perceptron. It is a feedforward artificial neural network algorithm using backpropagation with gradient descent to balance the errors and to update the weights.
- IBk: k nearest neighbor (k-NN) algorithm. k was set to 1 and weighted distances were not used.
- KStar: It is an instance-based learner algorithm which uses some similarity function and it uses an entropy-based distance.
- Linear Regression: Multi-linear regression.
- Random Forest: It is a type of ensemble learner and it constructs a forest of random trees.
- M5P: M5Base algorithm that implements base routines for generating M5 Model trees and rules [24].

It was observed ANFIS produced the best estimation performance (lowest RMSE value) for "Delta1" when tested with the data set with 900 records. Random Forest algorithm produced the best estimation performance when tested with the data set with 5717 records. The comparative results obtained by all of the algorithms using 10-folds cross-validation are given in the tables Table I and Table II.

TABLE I. Comparative performance results for the data set with 900 instances.

Algorithm name:	**Test methodology: 10-folds cross-validation**	
	Parameters / settings	***RMSE***
Ibk (1-NN)	default values in Weka	0.1748
KStar	default values in Weka	0.1467
Linear Regression	default values in Weka	0.1656
Random Forest	default values in Weka	0.1581
M5P	default values in Weka	0.1602
Multilayer Perceptron	learning rate: 0.3 momentum: 0.72 500 iterations Hidden layer1: 7 nodes Hidden layer 2: 2 nodes	0.1726
ANFIS (with 5 fuzzy rules)	learning rate: 0.32 momentum: 0.68 500 iterations	0.1237
ANFIS (with 10 fuzzy rules)	learning rate: 0.32 momentum: 0.68 500 iterations	0.1232

TABLE II. Comparative performance results for the data set with 5717 instances.

Algorithm name:	**Test methodology: 10-folds cross-validation**	
	Parameters / settings	***RMSE***
Ibk (1-NN)	default values in Weka	0.0981
KStar	default values in Weka	0.0588
Linear Regression	default values in Weka	0.0684
Random Forest	default values in Weka	0.0573
M5P	Minimum number of instances: 5	0.0608
Multilayer Perceptron	learning rate: 0.28 momentum: 0.7 1000 iterations Hidden layer1: 8 nodes Hidden layer 2: 2 nodes	0.0661

Algorithm name:	Test methodology: 10-folds cross-validation	
	Parameters / settings	***RMSE***
ANFIS (with 5 fuzzy rules)	learning rate: 0.32 momentum: 0.68 500 iterations	0.0955
ANFIS (with 10 fuzzy rules)	learning rate: 0.32 momentum: 0.68 500 iterations	0.0894

V. Conclusion

The results show that ANFIS model could be a promising numerical prediction alternative for the duration of medical services in medical institutions. Both 5-fuzzy rules and 10-fuzzy rules versions of ANFIS provided the lowest root mean squared error values among all of the well-known numerical prediction algorithms for the sample test set with 900 instances.

On the other hand, more accurate results were observed by some other algorithms after the tests with 5717 instances. This could be due to several reasons. The fuzzy rules that were implemented in the study might not be optimized for the problem. In addition, momentum, epoch and learning rate parameters might have been reconfigured in order to achieve better results. Instead of using both methods (least squares and backpropagation with gradient descent) together for the weight updates, each might have been implemented and tested separately. In order to build a more appropriate model for the problem in this case, some other fuzzy membership functions might have been used instead of Gaussian bell function. Our research plan in the near future is to design, implement and test all of these alternatives and improve the model.

A freeware ANFIS tool with a simple interface that can be used easily by both technical and non-technical people has been developed successfully. This might also be an important outcome of this study since it has been observed that all of the commercial licensed ANFIS software products might be costly for some companies or organizations while non-commercial freeware alternatives seem to be sophisticated for non-technical people. It has been planned to develop a second version of this tool and bring it into service for free public use.

References

[1] H. R. F. Najafabadi and A. H. Navin, "Electronic health card: opportunities and challenges," IEEE Proceedings of 6th International Conference on Application of Information and Communication, pp. 1-5, October 2012.

[2] B. Goswami, B. Singh, R. Chawla, V. K. Gupta and V. Mallika, "Turnaround time (TAT) as a benchmark of laboratory performance," Indian Journal of Clinical Biochemistry, vol. 25(4), pp. 376–379, 2010.

[3] M. Fieri, N. F. Ranney, E. B. Schroeder, E. M. Van Aken and A. H. Stone, "Analysis and improvement of patient turnaround time in an emergency department," Proceedings of the 2010 IEEE Systems and Information Engineering Design Symposium, pp. 239-244, April 2010.

[4] B. Breil, F. Fritz, V. Thiemann and M. Dugas "Mapping turnaround times to a generic timeline: a systematic review of TAT definitions in clinical domains," BMC Medical Informatics and Decision Making, vol. 11(34), pp. 1-12, May 2011.

[5] C. Hand, H. Mannila and P. Smyth, Principles of Data Mining. London: The MIT Press, 2001.

[6] N. Nedjah, M. M. Luiza and J. Kacprzyk, Innovative Applications in Data Mining. Berlin: Springer-Verlag, 2009.

[7] I. H. Witten, E. Frank and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques 3rd ed. Burlington, MA: Elsevier Inc., 2011.

[8] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Data Mining, Inference and Prediction, 2nd edition. New York: Springer, 2009.

[9] S. Haykin, Neural Networks and Learning Machines, 3rd edition. New Jersey: Pearson Education, Inc., 2009.

[10] E. Alpaydın, Introduction to Machine Learning, 2nd edition. USA: The MIT Press, 2010.

[11] C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer Science Business Media LLC, 2006.

[12] H. R. F. Najafabadi and A. H. Navin, "Enhancement and cleaning of handwritten data by using neural networks and threshold technical," IEEE Proceedings of 5th International Conference on Application of Information and Communication, pp. 1-4, October 2011.

[13] F. V. Celebi, S. Tankız, R. Yıldırım and L. Gökrem, "Modeling quantum cascade lasers by multilayer perceptrons," IEEE Proceedings of 3rd International Conference on Application of Information and Communication, pp. 1-3, October 2009.

[14] L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, pp.338-353, 1965.

[15] F. M. McNeill and E. Thro, Fuzzy Logic, A Practical Approach. USA: Academic Press, 1994.

[16] N. Nedjah and L. M. Mourelle, Fuzzy Systems Engineering: Theory and Practice. Netherlands: Springer-Verlag, 2005.

[17] W. Siler and J. J. Buckley, Fuzzy Expert Systems and Reasoning. USA: John Wiley & Sons, 2005.

[18] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," IEEE Transactions on Systems, Man and Cybernetics vol. 23 (3), 1993.

[19] F. O. Karray and C. W. De Silva, Soft Computing and Intelligent Systems Design: Theory, Tools and Applications. UK: Pearson Education Ltd., 2004.

[20] R. Fuller, Neural Fuzzy Systems. Turku: Abo Akademi University, 1995.

[21] H. Li, C. L. P. Chen and H.-P. Huang, Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering. USA: CRC Press, 2001.

[22] D. T. Larose, Discovering Knowledge in Data - An Introduction to Data Mining. New Jersey: John Wiley & Sons Inc., 2005.

[23] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2006.

[24] J. R. Quinlan, "Learning with continuous classes," Proceedings of 5th Australian Joint Conference on Artificial Intelligence, Singapore, pp. 343-348, 1992.

Using Fuzzy Logic Concepts in Creating the Decision Making Expert System for Cardio - Vascular Diseases (CVD)

Assemgul Duisenbayeva, Lyazzat Atymtayeva
Department of Information Systems Management
Kazakh-British Technical University
Almaty, Kazakhstan
a.duisenbayeva@gmail.com, atymtayeva@gmail.com

Iskander Beisembetov
Kazakh-British Technical University
Almaty, Kazakhstan
i.beisembetov@kbtu.kz

***Abstract*— This paper is devoted to the problems of the design of Fuzzy Expert System for cardiovascular disease diagnosis. Such type of systems may be efficiently constructed by using the fuzzy variables as inputs and outputs and generation of results in terms of the presence of the disease in human.**

***Index Terms*— expert systems, fuzzy logic, cardiovascular disease.**

I. Introduction

Cardio - Vascular Diseases (CVD) are the leading cause of death worldwide: for no other reason, every year more people die from CVDs than before. According to estimates in 2008-2012, 17.3 million people were died from CVDs that represent 30 % of all deaths in the world. Of this total, 7.3 million people died from coronary heart disease and 6.2 million died due to stroke. And this is a problem not only for the elderly population, but unfortunately equally for young generation. The early prophylaxis of CVDs may help to decrease the mortality rate. For prediction of the state of human's health the expert systems are often used. By combination of different parameters and analyzing its impact we can identify the level and degree of the disease presence in human.

Very often heart disease diagnosis is hard to determine. Many factors may serve as a base for identifying the state of human. But actually not all of these discoveries are real features of disease. The vague and uncertain boundaries between the state of health and state of disease make the process of identifying the risk of CVDs more complicated. [1]

In this case by using the intelligent systems we can facilitate the work of CVD's experts.

Main reasons of the heart disease usually include the following characteristics: risk of blood pressure, cholesterol, diabetes, smoke, gender and age that affect on heart disease risk [2]. By identifying the level of each factor above we can define the risk level of disease. Thus, they can serve as an input factors for our Fuzzy Expert System which allows to create the special rules in order to generate the information about human state.

II. Overview of Existing Technologies to Development An Intelligent System

Conduct studies on the review of existing technology Intelligent Systems are the following aspects. Classes of expert systems. According to the degree of complexity of tasks of the expert system can be classified as follows: According to the method of formation of the decisions of the expert system are divided into two classes: analytical and synthetic. Analytical systems include solutions from the set of known alternatives, and a synthetic system to generate the unknown solutions. The analytical expert system is ES carrying out the evaluation of solutions (hypothesis testing). Synthetic expert system is ES engaged in the generation of possible solutions (hypothesis generation).

Part of the provisional characteristic of an expert system can be static or dynamic. Static systems solve tasks with immutable in the process of addressing data and knowledge. Dynamical permit such changes. Static expert system is ES, the crucial tasks in conditions that are not time-varying source data and knowledge. The dynamic expert system is ES, the crucial task in terms of time-varying source data and knowledge.

The types of data used and knowledge of the expert system are classified into a system with deterministic (well-defined) knowledge and uncertain knowledge. Under the uncertainty of knowledge (data) shall be defined as incompleteness (no), inaccuracy (measurement uncertainty), ambiguity (ambiguous concepts), fuzziness (qualitative assessment instead of quantitative).

The number of used sources of knowledge expert systems can be built using one or multiple knowledge sources.

Also expert systems classified. The standard classification of expert systems is missing, however, is most often expert systems differ in purpose, subject area, methods of knowledge representation, dynamism and complexity: For the purpose of classification of expert systems can carry out the following:

diagnostics of the systems, including monitoring (continuous monitoring of the current status) forecasting systems based on modeling past and present planning and development of organizational and technological management; the design or formulation of clear regulations for the construction of objects that meet the requirements; production expert system designed to improve the efficiency of the various stages of the production process (planning, design, production, control) automatic control (regulation).

By subject area the greatest number of expert systems used in military, Geology, engineering, computer science, aerospace engineering, mathematics, medicine, meteorology, industry, agriculture, process control, physics, Philology, chemistry, electronics, law.

Classification of expert systems on the methods of knowledge representation divides them into traditional and hybrid. Traditional expert systems are used mainly empirical models of knowledge representation and predicate calculus of the first order. A hybrid expert system using all available methods, including optimization algorithms and concepts of databases.

According to the degree of complexity of the expert system is divided into superficial and deep. Surface expert systems represent knowledge in the form of rules "IF-THEN". Classification of expert systems for dynamic divides expert systems for static and dynamic. The subject area is called static, if it describes the original data is not changed in time. Static field means describing the immutability of its source data. Thus derived data (derived from source) and may appear again, and be changed (without changing, however, the original data). If the original data describing the subject area, change during the decision task, the subject area is called dynamic. In the dynamic architecture of the expert system compared to static, two components are introduced: the subsystem modeling the external world; the communication subsystem and the external environment.

Most famous ES and their application.

- CLIPS - a very popular shell for the construction of ES (public domain)
- OpenCyc is a powerful dynamic ES with a global ontological model and support independent contexts WolframAlpha, a knowledge base and a set of computational algorithms, intelligent “computational knowledge engine”
- MYCIN was an early expert system developed over 5 or 6 years in the early 1970-ies at Stanford University. It was written in Lisp as the doctoral dissertation of Edward Shortliffe under the direction of Bruce Buchanan, Stanley N. Cohen and others. In this lab you created earlier Dendral expert system, but this time attention was focused on the use of decision rules with elements of uncertainty. MYCIN was designed to diagnose bacteria that cause severe infections such as bacteremia and meningitis, and to recommend the required number of antibiotics depending on the body mass of the patient. The system name comes from the suffix "-mitsin '" often found in the names of antibiotics. Also, Mycin was used to diagnose diseases of blood clotting.

Combining the existing technology of intelligent systems to create new systems for different application areas. Therefore, the prospect of the studied direction is to create a base of scientific research for new development of intelligent systems. [2]

III. Fuzzy Logic Concepts

Due to intrinsic uncertainty in decision making process of identifying the features of diseases, it is very difficult to apply quantitative approach and productive model to the measurement in many critical decision making tasks. For this reason, they are usually handled by a significant number of human experts. On the other hand, this causes unnecessary inconsistencies, together with the high fault risks and long duration of the processing. Consequently, we expect a significant demand on a framework for handling such uncertainties such as fuzzy logic.

Fuzzy logic may be viewed as an attempt at formalization and mechanization of two remarkable human capabilities: (1) to thinking and make decisions in an environment with imperfect information and (2) to perform a wide variety of intelligent tasks without precise measurements or intensive computation Experts in medical area may almost always utilize both of the capabilities in their work.

There are plenty of applications of fuzzy logic in risk management and decision making for matters related to health care systems. [3]

A new technique based on fuzzy logic is developed for prioritizing diseases and features of disease for decision making process by assessing risks of the disease discovering based on severity, recurrence of event, and detectability [3]. Trivially, fuzzy logic and its related methodology resolve basic issues in customary strategies for danger assessment construct entirely with respect to numerical techniques such that:

1) it permits the investigator to assess the danger connected with malady finding modes specifically utilizing the phonetic terms that are regular for human experts in making the decisions;

2) equivocal, subjective, and loose data, together with quantitative information, can be utilized as a part of the basic leadership procedure and they are taken care of in an integrative and predictable way; and

3) it gives a more adaptable structure for joining the seriousness, event, and perceptibility parameters.

In short, approaches based on fuzzy logic allow utilization of generalized, i.e. imprecise or incomplete information, whereas traditional numerical methods may only work well with precise numbers. Moreover, fuzzy logic based approaches are much more compatible with human reasoning, as quantities can be represented naturally in linguistic terms. This means that the inputs, as well as the outputs, to such systems can be partially or entirely linguistic so that human experts may likely interpret those naturally. When necessary,

consistent numerical representation can be generated as a result of defuzzification.

As can be seen, fuzzy logic was applied effectively but only in particular stages of decision making process, more precisely, in its risk identification and evaluation. Actually, assessment of the diseases risks requires a systematic approach to analyze various types of risks in a comprehensive manner. [4]

IV. Knowledge Base

Any expert system consists from the knowledge base, inference engine and user interface for interaction with human experts and ordinary users. [4]

The knowledge base — is a set of the facts and inference rules allowing a logical output and intelligent information processing. The events are provided by depth (fundamental) knowledge of medicine, such as the hierarchy of diseases and hierarchy of parameters for the description of symptoms. Inference rules are surface (managing directors, practical) knowledge, such as rules of diagnosis of illness on symptoms and differentiation of diseases. [5] To build a knowledge base we use two methods of elicitation knowledge from an expert. We use one of the active methods is Interview and the second one passive Document analysis. For the interview we create questions and we ask real expert to give for every answer (attribute) its weight. The example of one attribute values is shown on Table I.

TABLE I. Example of Heart Rate attribute values

Heart Rate	Low	normal	high	hypertony
>60	80%	20%	0%	0%
60-80	5%	75%	20%	0%
80-100	0%	10%	85%	5%
100-140	0%	0%	5%	95%
140<	0%	0%	0%	100%

Input fields (attributes) are chest pain type, blood pressure, cholesterol, resting blood sugar, resting maximum heart rate, gender and age. The output field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 4 (distinguish presence (values 1, 2, 3 and 4)); increasing value shows increasing heart disease risk. In this study we give for every input own parameters. Regarding the blood, we can use systolic blood pressure. In this dataset, fields are divided to some sections and each section has a separate value. For example, in this dataset, gender has two section (1=male and 0=female), chest pain has 4 parameters (1=typical angina, 2=atypical angina, 3=non-angina pain and 4=asymptomatic), resting blood sugar has 2 parameters (0=false and 1=true) and it is true and 2=Hypertrophy. The result of study depends on parameters of inputs and how we define the boundaries of the parameters.

V. Fuzzy Expert System Design

For finding solutions to issues of uncertain fuzzy system is the perfect tool. Accordingly, for analysis and determination of risk of diseases of fuzzy logic [6] is the appropriate tool. First design a fuzzy expert system is the definition of input and output variables. After need to determine membership function of all variables. Second, we introduce output variable with its membership functions. The following section, we show the rules of the system. The input variable are:

1) *Chest pain:* This input variable includes 4 types of chest pain. We determined the value in this system for each type chest pain, those values which we use to create the system. Each type of chest pain is a fuzzy set. Types of chest pain with their values showed the following.

1=typical angina
2=atypical angina
3=non-anginal pain
4=asymptomatic

2) *Blood Pressure:* Different values of blood pressure change the result easily. This variable input divided 4 fuzzy sets. Fuzzy sets "Low", "Medium", "High" and "Very high". The membership functions "Low" and "Very high" sets are trapezoidal and membership functions of "medium" and "high" sets are triangular. We defined fuzzy membership expressions for the input fields blood pressure These fuzzy sets are shown in Fig. 1. The structure of functions of the field of blood pressure are shown in Fig. 2.

INPUT FIELD	RANGE	FUZZY SETS
Systolic Blood Pressure	<134	Low
	127-153	Medium
	142-172	High
	154>	Very high

Fig. 1. Classification of Blood Pressure

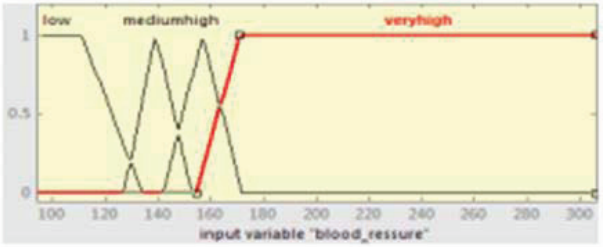


Fig. 2. Membership function of Blood pressure

3) *Cholesterol:* Cholesterol can influence to the result and can change. And for this input field, we use the value of low density lipoprotein (LDL) cholesterol. Cholesterol field has 4 fuzzy sets (Low, Medium, High and Very high). On Fig. 3 showed Classification of cholesterol and on Fig. 4 we can see Membership function.

INPUT FIELD	RANGE	FUZZY SETS
Cholesterol	<197	Low
	188-250	Medium
	217-307	High
	281>	Very high

Fig. 3. Classification of Cholesterol

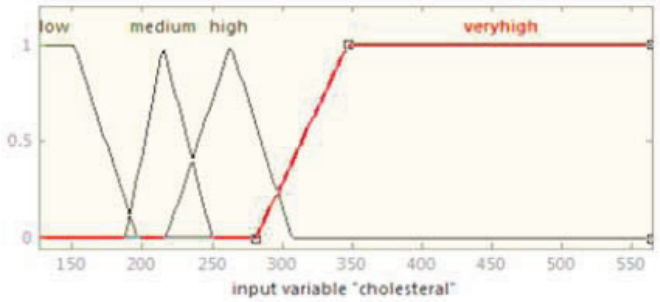


Fig. 4. Membership function of Cholestorole

4) *Blood Sugar (Diabetes):* Blood sugar field is one of the most important
factors in this system that changes the result. This input field has just one fuzzy
set. In this system, we have defined that if the mount value of blood sugar is higher
than 120 (>120) then man has blood sugar. Figure 5 shows the membership
function of blood sugar. Membership function of this fuzzy set is trapezoidal.

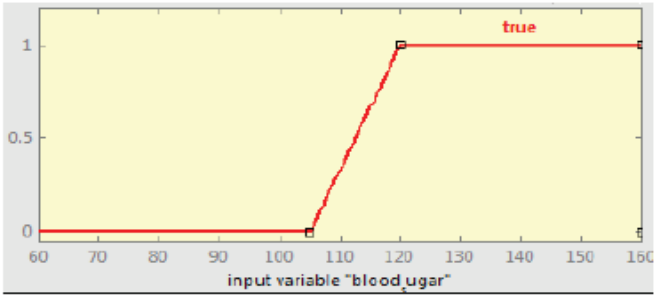


Fig. 5. Membership of Blood Sugar

5) Maximum Heart Rate: The Value of this input
field is a maximum heart rate of the person within 24 hours. With increasing age in humans, high heart
rate during 24 hours is reduced. In this area, we have 3 variables (fuzzy sets) (low, medium and high). In Fig 6 we defined these fuzzy sets. The membership functions "Low" and "High" fuzzy sets are trapezoidal and membership function of "Medium" fuzzy set is triangular that will be shown in Fig 7

INPUT FIELD	RANGE	FUZZY SETS
Maximum Heart Rate	<141	Low
	111-194	Medium
	152>	High

Fig. 6. Classification of Heart Rate

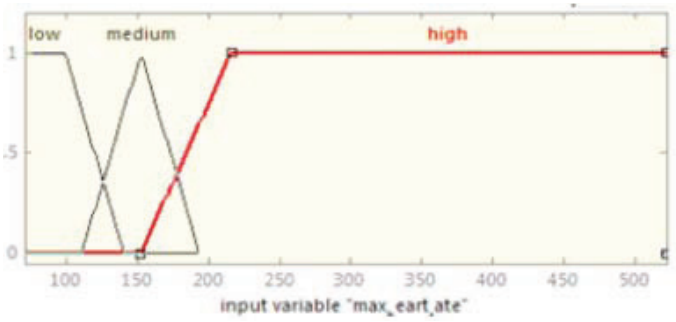


Fig. 7. Membership function of Heart Rate

6) *Gender:* This input field has only 2 values (0, 1) and sets (Female, Male). If value is 0, means that patient is male and if value is 1, means that patient is female.

7) *Age:* This input field divides into 4 fuzzy sets (Young, Mild, Old, Very old). These fuzzy sets with their ranges will be shown in Fig 8. Membership functions of "Young" and Very old" are trapezoidal and membership functions of "Mild" and "Old" are triangular. This function showed in Fig. 9

INPUT FIELD	RANGE	FUZZY SETS
Age	<38	Young
	33-45	Mild
	40-58	Old
	52>	Very old

Fig. 8. Classification of Age

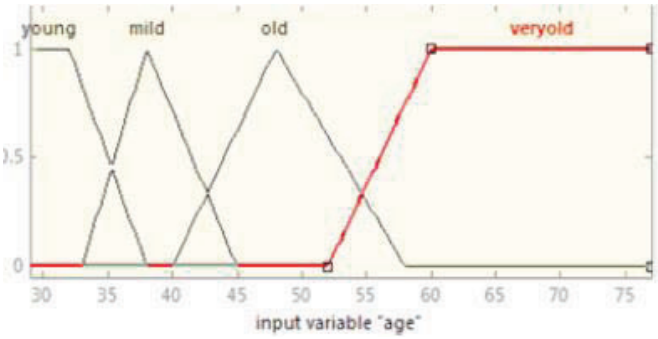


Fig. 9. Membership function of Age

Output Variable: The "goal" field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 4. By increasing of integer value, heart disease risk increases in patient. In this system, we have considered a different output variable, which divides to 5 fuzzy sets (Healthy, Sick (s1), Sick (s2), Sick (s3), sick (s4)). Fig. 10 shows these fuzzy sets with their ranges. Membership functions of "Healthy" & "Sick (s4)" fuzzy sets are trapezoidal and membership functions of "Sick (s1)", "Sick (s2)" and "Sick (s3)" are triangular. These membership functions will be shown in Fig. 11.

OUTPUT FIELD	RANGE	FUZZY SETS
Result	<1.78	Healthy
	1-2.51	Sick (s1)
	1.78-3.25	Sick (s2)
	1.5-4.5	Sick (s3)
	3.25>	Sick (s4)

Fig. 10. Classification of Results

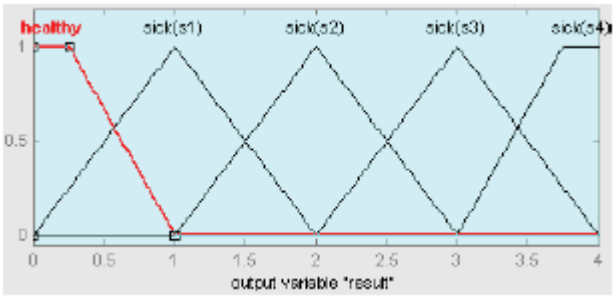


Fig. 11. Membership of Results

VI. Conclusion

On this paper we studied how to apply Fizzy logic concepts to one of the most popular problem in health care and solve this problem by creating the fuzzy expert systems. Fuzzy Expert System was design by knowledge of real expert (doctor), every input and input's parameter was created based on professional knowledge. This system can assist the physicians in making decision, reduce time for determine disease and can help to avoid the faults.

References

[1] World Health Organization 2012. //cardio - vascular disease
[2] L.Atymtayeva, A.Akzhatova, K.Kozhakhmel, L.Naizabayeva, Development of intelligent systems for information security auditing and management: Review and assumptions analysis, AICT 2011, Baku, Azerbaijan.
[3] M.Nikravesh Janusz Lotfi A.Zadeh, Foring New Frontier: Fuzzy Pioneers
[4] Kiselev Ye.N. Zaparozhe, Expert system in medicine, 2003.
[5] Herman O., Introduction to expert systems and knowledge processing
[6] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

Effect of Filter Selection On Classification of Extrasystole Heart Sounds Via Mobile Devices*

Huseyin Coskun
Computer Sciences Application and Research Center, Uşak University
Uşak, TURKEY
huseyin.coskun@usak.edu.tr

Tuncay Yigit
Computer Engineering Department, Engineering Faculty, S. D. U.
Isparta, TURKEY
tuncayyigit@sdu.edu.tr

Omer Deperlioglu
Vocational School of Afyon, Afyon Kocatepe University
Afyonkarahisar, TURKEY
deperlioglu@gmail.com

***Abstract*—Sound of the heart is the basic biomedical information utilized for diagnosis by medical doctors in heart diseases. These sounds show differences according to different pathological characteristics. Extra systole sounds, which mean extra heartbeat, can be perceived as throbbing by people. Occurrence of these sounds in certain age groups may be the indication of tachycardia. In this study, effect of Butterworth, Chebyshev and Elliptic filters on classification results for noise removal in extra systole specific sounds in heart sound database is analyzed. The filters chosen and other methods are paid attention to be faster because the application developed for this aim will be used on mobile devices. Db5 type wavelet transformation method has been used to gain less as feature set. Support vector machine has been used to classify. According to the results gained, the fastest filter for noise removal in extra systole specific heart sounds is Butterworth and the filter that gives the best classification results is Elliptic filter.**

Keywords— filter selection, phonocardiogram, extrasystole, classification, mobile, heart sound

I. INTRODUCTION

The heart is divided into four cubbies. The upper cubby is called atrium and lower cubby are called ventricle. The heart muscles squeeze blood from cubby to cubby. At each squeeze, the heart valves open to allow blood move backward. Thereby, the valves keep blood moving as adequately as possible through the heart and out to the body [1]. Acoustic vibrations occurred as a result of these mechanical movements of the heart forms heart sounds (HS). Listening to these sounds (Auscultation) is one of the basic medical knowledge used by medical doctors for diagnosis of heart diseases. The recorded forms of the sounds occurred as a result of these movements are called phonocardiogram (PCG) signals. Diagnosis of heart diseases are performed by analysis of the sound signs that belongs to the heart by expert doctors. Early diagnosis and fast recognition of heart diseases is of vital importance for treatment and early intervention. Therefore, besides heart diseases, it is very important to gain fast the general information (blood pressure, heartbeat speed, pulse, rhythm, etc.) of a patient having a heart disease in urgent delivery to the hospital. Because of this, biomedical devices processing heart sound should use fast signal processing methods [2]. Mobile devices in communication and information systems have additional hardware features like integrated camera, heat sensor, balance sensor, GPS, gravity sensor, integrated microphone, integrated speaker, integrated compass, fingerprint reader. Mobile computers have widespread usage more than they used to have, because they are small enough to be easily portable and they have features mentioned above. In spite of benefits mentioned above, mobile computers do not have high-speed processors, memory and graphic card as desktops or laptops have. Mobile devices mostly have low processor clock frequency, memory and graphic card capacities and limited batteries. The applications that will operate on mobile devices should be developed according to these characteristics. Thus, to classify heart sounds recorded by mobile devices effectively, least amount of data that can represent heart sounds and the fastest methods should be used. So, limited batteries of mobile devices will be used effectively with low memory and processor usage by utilizing less data and faster methods.

Before classification of heart sound signs, attributes belong to the sound should be attained [3]. The structure and gaining methods of these attributes affect classification result directly. Removal of noise, interference, respiration sound, movement sound, ambient sound, etc. besides heart sound from HS provides gaining of attributes belong to HS in higher accuracy. It also provides representation of sign by less data when the specific frequency information is removed from the sign, thus providing faster sign processing procedure.

These systems that can be used by both doctors and patients will be effective in gaining of data belong to heart both by doctors and patients in general. An average user will have the chance to get a pre-diagnosis for his/her disease maybe not going to hospital by sharing his/her heart sound that he/she listens by microphone of the mobile device. It will be also possible to make necessary preparations and pre-diagnosis for diagnosis and treatment in the hospital by transferring data of the heart of a patient having an urgent heart disease to the hospital database gained by mobile devices.

Among the recent studies with artificial intelligence methods through classification of heart diseases are noise removal with different methods from heart sounds with different pathological characteristics, gaining of attribute values and performing of frequency analysis [3, 4]. Fourier transforms and wavelet transforms are the mostly used methods among those methods respectively [5-10]. In studies aiming classification, methods like plexus in different models, different seed methods and support vector machines, Hilbert-Huang transform are used [2, 11-14].

II. MATERIALS

Heart sounds result from the vibrations of heart valves between auricle and ventricle. A normal heart sound is like the sounds “lub dub, lub dub” sounds. In medicine, “lub” is named as S1 sound while “dub” sound is called as the S2 sound. Every “lub dub” sound is one beat [15]. Normal heart can have 60 - 100 beats in a minute. In Fig. 1, normal heart sound, S1, S2 parts sound design are shown.

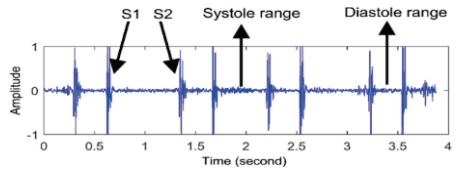


Fig. 1 Normal heart sound S1, S2 sections

Range between inception of S1 and S2 sounds are called as systole. Inception of S2 sound and starting of the S1 sound again is called as diastole. Systole sounds in a normal heart follow the diastole between every beat. This situation is shown in Fig. 1. Extra systole range, which is one of the heart sounds in an abnormal heart, is the event of hearing the S1 sound after the S1 sound or hearing S2 again instead of S1 after the S2. Extra systole heart sounds include sound design of “lub-lub dub” or “lub dub-dub” series. Heart records half beat in this way. This situation is shown in Fig. 2.

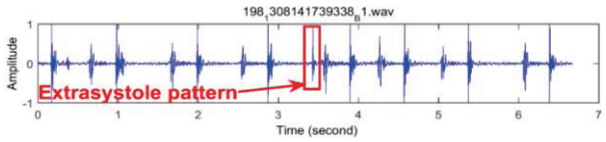


Fig. 2 Heart sound with extrasystole pattern

A. Extra systoles rate of incidence in Humans

Atrial extra systoles are frequently seen in 60% of healthy adults connected to a 24 hour Holter monitor. According to Framingham heart studies, 1-2 extra beat were observed in 33% of people who have no heart disease. In those with cardiovascular disease it raised to 54%. Ventricular extra systoles are very rare in little children. However, ventricular extra systoles were observed in 41% of healthy youth who are between the age of 14 and 16, in 50-60% of healthy adults, and finally 84% of people who are 70 [16]. Ventricular extra beats are very frequent in people with hypertension, hypertrophy depending on sports and other reason, in cardiomyopathy disease, mitral valve prolapse disease, and valvular heart disease, cardiovascular disease. Extrasystole which is identified as early or extra beat of heart is the most frequent incident seen in irregular beats resulting from ventricles and heart valves. Extrasystole heart sounds can be seen in a healthy person. Extrasystole heart sounds are associated with existence of symptom of a disease in cases that it often happens. If the frequency is above a definite value of age, it speeds up tachycardia heart disease [17]. Thus, extrasystole heart sounds must be taken seriously into consideration. In this study there are four types of heart sounds including 320 normal, 95 murmured, 46 with extrasystole pattern, and 195 unlabeled ones. These sounds were recorded by using Digiscope numbered stethoscope [18].

III. METHOD

Hearty sounds must be normalized before extraction of the noise. The noises were normalized with the statement in (1) [19].

$$x_{norm}[n] = \frac{x[n]}{\max(|x[n]|)} \quad (1)$$

A. Filtering

Filtering process in biomedical systems are usually performed by numerical filters. Numerical filtering is described as gaining required frequency values according to the separation characteristics of required filter to improve the sign for intended use [20, 21]. In this study, Butterworth, Chebyshev and Elliptic numerical filters which are widely used for filtering processes are used. Cardiac information in heart sounds includes low frequency ingredients. Generally, heart sounds have 10-500 Hz frequency range [26]. A normal S1 sound is in frequency of 70 -150 milliseconds and 25-45 hertz. A Normal S2 sound is nearly between 60 - 120 milliseconds and with 50 hertz [27]. During auscultation with an electronic stethoscope, sounds such as lung sound, rhonchus sound, body movement, and even environmental sounds mixed with heart sound make noise. If these noises aren’t extracted from heart sounds accuracy of features belonging to these sounds can be affected [28]. In this case, it is not expected to have correct results from this classification. It is noticed by Butterworth and Chebyshev are the best, fastest, and simple filters for extraction of noises from heart sounds [4, 29]. With the same level ad values, Butterworth, Chebyshev and Elliptic filters within MATLAB™ program were applied to some heart sounds having feature of extra systole in heart sounds data base. Uptimes are shown in Fig. 3.

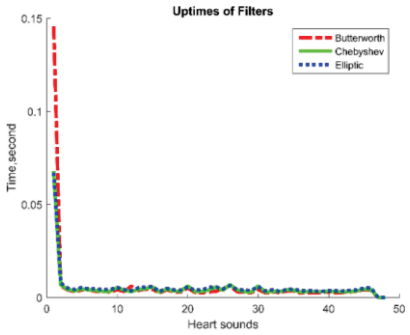


Fig. 3 Uptimes of the filters

Another feature in selection of filter for extraction is SNR (Signal-Noise-Ratio). This value is described in dB and its high level indicates the success of filter in noise extraction. When these filters were applied to the heart sounds in database, it was seen that the most approximate and the fastest filters for uptimes are Butterworth and Chebyshev. As seen in Fig. 4, the Elliptic filter has the best SNR.

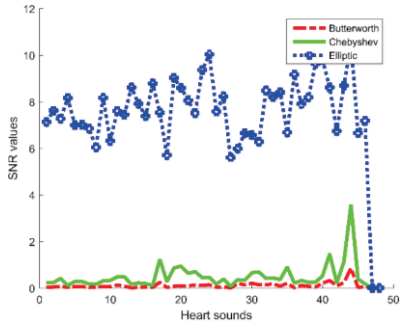


Fig. 4 SNR values of filters

B. Classification

1) Support Vector Machines: SVM can be defined as separation of data on the same plane belonging to different classes thanks to bound line formed by linear equation [30]. SVM is a core based learning approach giving fast and correct results in such issues; linear and nonlinear classification, regression analysis, identification of outlier, function and density estimation. It is especially used in classification problems when patterns between variances of data to be classified in data mining are not known [31]. In SVM, training data is in (2) when n dual dimensioned, and classed data set is considered as in (2).

$$(x_1, y_1), (x_2, y_2), (x_3, y_2), (x_4, y_1), .. (x_i, y_i)\ i = 1,2 \dots n \qquad (2)$$

In (2) $x \in R^d$ resembles d-dimensioned features vector, y vector symbolizes class labels. For dual dimensioned n numbered input data $x \in \mathcal{R}^2$ and $y \in \{+1, -1\}$. Support vectors, in condition that x point is $w = [w_1\ w_2\ w_3\ w_4\ \dots\dots.w_n]^T$ ve $w \in R^N$, identified in (3) as d differentiation or decision function. b in (3) is a constant number symbolizing bias value.

$$d(x,w,b) = w^T x + b = \sum_{i=1}^{n} w_i x_i + b \qquad (3)$$

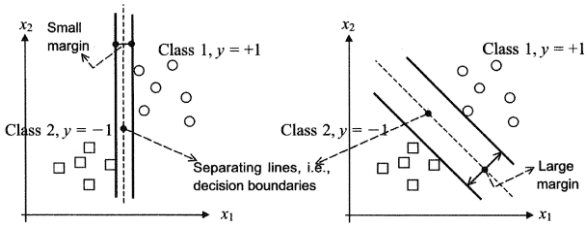


Fig. 5 Best separator plane selection

According to (3) separative striped line presented in Fig. 5 can be defined in (4).

$$w^T x + b = 0 \qquad (4)$$

Canonical upper hyper planes are described as;

$$D1 = w^T x + b = +1 \qquad (5)$$
$$D2 = w^T x + b = -1 \qquad (6)$$

Let w symbolizes weight vector in (5),(6), distance of x1 point to upper plane on D1 line d_1, and distance of x2 on D2 line to upper lane d_2 in turn symbolizes in class 1 and 2 distance of samples closest to upper plane to upper plane M (margin) = $d_1 + d_2$ is accepted as differentiation value [32].

M can be defined geometrically as in (7);

$$M = \frac{2}{\|w\|} \qquad (7)$$

SVM aims to find an upper plane to take differentiation value M on countless number of upper planes to the highest level [33]. It is called as Interval Maximization. In order to maximize M it is necessary to minimize $\|w\|$. When applying SVM to the problems that cannot be separated completely as linear -with the help of function of kernel- data is moved to a higher dimension and a space where it can be separated linearly. It is preferable that the classification is held in this new space. This high dimensional space is called the defining quality space. In order to leave out the features of classification in the original data, the function of mapping $\Phi(x_i)$ which is not linear, should be used. This function of mapping is a stable function that allows the transformation of xs to $\Phi(x)$ in data. This situation is shown in Fig. 6.

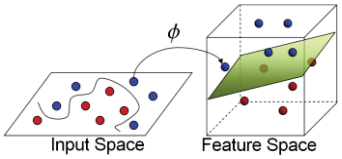


Fig. 6 Moving data which can not separate as linear to feature space by mapping function

The function of Φ that moves the samples to a different quality space is defined as;

$$\Phi : \Re^d \rightarrow H$$

While x vector, which forms the vector of input, consists of x_i value, and $\Phi(x)$ vector, which forms the quality space, consists of $\Phi_i(x)$ value. After inputs have moved to the defining quality space, the training of linear classification is done in order to find out the maximum range (margin). The training algorithm of SVM is only related to inner multiplying of data in H space. This multiplying is shown by K in (8) and called as the kernel function [34].

$$K(x_i, x_j) = \Phi(x_i).\Phi(x_j) \qquad (8)$$

In the mapping of quality space via positive defined kernel functions, there is a linear separator [35]. In this direction the values of classification derived from data in testing phase is calculated through the chosen function in (9) [21].

$$d(x,w,b) = w^T \Phi(x) + b = \sum_{i=1}^{n} w_i \Phi_i(x) + b \qquad (9)$$

IV. Results And Discussion

The fastest and the best methods should be chosen taking into account the limited hardware features of mobile devices after these methods are analyzed. Filters utilized are applied to heart sounds in the heart sound database [18] and the results gained for the filters at the same value and level are given in Fig. 7.

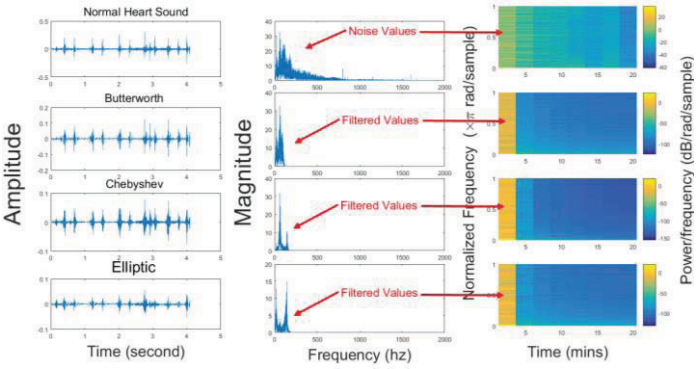


Fig. 7 Frequecny and spectrogram analysis of Noisy heart sound (left) and, noiseless heart sound (rigth)

Wavelet transform is applied to gain attribute that belongs to heart sounds gained after the usage of filters. db5 (daubechies 5. level) type fifth level wavelet approximation coefficients (a5) attained with the developed mobile application. Calculated wavelet coefficients are transformed into the data in Table-1 to be used as training data. In Table-1, first value in every line is mean of wavelet transform values of heart sounds while second value is variance and the last one is class label.

TABLE I. Train Data File Information (Specific Section)

Mean	Variance	Class
0.032719350	0.064847712	E
0.035215709	0.040964317	E
0.017195715	0.006889646	E
-0.026257453	0.114419844	E
0.031954579	0.036022337	E
-0.002094549	0.000732152	N
0.000400911	0.000487900	N
-0.015545786	0.043945478	N
0.002552997	0.003975180	N
E: Extrasystole N: Normal		

Uusing descriptions in Table-2, certainty, accuracy and precision values are calculated in compliance with the classification results.

TABLE II. Calculation of Performance Analysis

		Real Status	
		Positive	Negative
Predicted Status	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

If sample class label and class data was found positive after the calculation, it is defined as true positive (TP) and when sample class label is positive but class data is found as negative it is called as false- negative (FN). When sample class label is negative and class data was found negative it is called as true – negative. Finally, if when sample class label is negative but class data was found positive, it is then called as false positive (FP). Confusion matrix [36] according to classification results is given in Table-3. Performance analysis values calculated in this way with values are given in Table-3.

TABLE III. Confusion Matrix

Filter Type	Classified Heart Sound					
	Classes	E	N	A	B	C
Butterworth	E	6	9	15	**40**	9
	N	8	7	15	**53.33**	8
		14	16	30		**17**
Chebyshev	E	10	5	15	**66.66**	5
	N	4	11	15	**73.33**	4
		14	16	30		**9**
Elliptic	E	12	3	15	**80**	3
	N	4	11	15	**73.33**	4
		16	14	30		**7**
	E: Extrsystole		N: Normal			

A = Line Total, B = Classification accuracy of every class in percentage, C = Number of heart sounds which are in x class in reference data but detected that it is not in that class after classification. Necessary calculation of certainty (P(c)), sensitiveness (P(s)), accuracy (P(a)) values used in analysis of classification performance can be solved via (10), (11), (12) with the use of definitions provided in Table-2.

$$P(c) = TN / (TN + FP) \quad (10)$$

$$P(s) = TP / (TP + FN) \quad (11)$$

$$P(a) = (TP + TN)/N \quad (12)$$

Certainty value determines in which ratio normal heart sound data could be separated from extra systole pattern heart sound data. Precision value determines in which level of success the extra systole pattern heart sounds could be guessed.

TABLE IV. Values of Performance Analysis

Filter Type	Certainty	Sensitiveness	Accuracy
Butterworth	0,4375	0,4285	0,43
Chebyshev	0,6875	0,7142	0,70
Elliptic	0,7857	0,75	0,76

When classification success of the mobile application is calculated by use of confusion matrix in Table-3, the value is found as ((10+11)/30) * 100= 70% is found using Chebyshev filter and ((12+11)/30) * 100= 76.66% is found using Elliptic filter according to Table-3. In another study made on the same heart sound database [10], extra systole heart sounds were calculated as accuracy of 16.67%, while classification by the developed mobile application was calculated as accuracy of 40% at minimum and 80% at maximum. In Fig. 7, it is observed that Chebyshev and Elliptic filters are more successful in extraction of noise. Also in Fig. 7 noise extraction was shown with red arrow mark. In Table-3 and 4 it is

observed that Elliptic filter is the best filter type to classify of extra systole heart sound.

V. CONCLUSION

Noise removal step of sound is too important that beginning step of classification studies on heart sound with the purpose of early diagnosis of heart diseases. In this study, different filtering method is applied for same heart sound data set and classification method in an attempt to emphasise important of filtering step of heart sound, to offer an insight into studies after now in this area. For example, it is mentioned that the best, the fastest and the simplest filters to remove the noises in the heart sounds are Butterworth and Chebyshev [4, 29, 37]. In this study, to classify heart sounds on mobile devices with SVM it is observed that the fast filter is Butterworth and the filter having the best effect on classification result is elliptic filter. According to this result and if we take other studies into consideration, the effect of the filters which are used in classification process may show differences according to the data set that is applied to the filter.

ACKNOWLEDGMENT

This paper has been supported by Afyon Kocatepe University Scientific Research and Projects Unit with the Project number 16. KARİYER.42.

REFERENCES

[1] Brown, E.L.W.S., AP: 'Heart Sounds Made Easy' (Edinburgh: Churchill Livingstone, 2002. 2002)

[2] Guraksin, G.E; E., Ucman; Deperlioglu, Omer 'Classification of the Heart Sounds via Artificial Neural Network', International Journal of Reasoning-based Intelligent Systems (IJRIS), 2010, 2, (4)

[3] Singh, M., and Cheema, A.: 'Heart Sounds Classification using Feature Extraction of Phonocardiography Signal', International Journal of Computer Applications, 2013, (Volume 77– No.4), pp. 13-17

[4] Cherif, L.H., Mostafi, M., and Debbal, S.M.: 'Digitals filters in heart sound analysis', International Journal of Clinical Medicine Research, 2014, (1(3)), pp. 97-108

[5] Say, Ö., Dokur, Z., and Ölmez, T.: 'Classification of Heart Sounds By Using Wavelet Transform', Proceedings of the Second Joint EMBS/BMES Conference, 2002

[6] A. Djebarri, R.F.B.: 'Short-time Fourier transform analysis of the phonocardiogram signal', in Editor : 'Book Short-time Fourier transform analysis of the phonocardiogram signal' (The 7th IEEE International Conference on Electronics, Circuits and Systems 2000 (ICECS 2000), 2000, edn.), pp. 844–847

[7] Lee, J.J., Lee, S.M., Kim, I.Y., Min, H.K., and Hong, S.H.: 'Comparison between Short Time Fourier and Wavelet Transform for Feature Extraction of Heart Sound', IEEE TENCON, 1999, pp. 1547-1550

[8] Debbal, S.M., and Bereksi-Reguig, F.: 'Cardiac murmur analysis using the short time fourier transform', Journal of Mechanics in Medicine and Biology, 2005, 6

[9] P.S.Vikhe, S.T.H., N.S.Nehe: 'Wavelet Transform Based Abnormality Analysis of Heart Sound', in Editor : 'Book Wavelet Transform Based Abnormality Analysis of Heart Sound' (IEEE, 2009, edn.), pp. 367 - 371

[10] Jeyarani, A.D., and Singh, T.J.: 'Feature Extraction from Heart sound signal for Anomaly Detection', IJCSNS International Journal of Computer Science and Ne 132 twork Security, 2011, (VOL.11 No.9), pp. 132-140

[11] Ölmez, T.D., Zümray: 'Classification of heart sounds using an artificial neural network', Pattern Recognition Letters, 2003, 24, (1-3), pp. 617–629

[12] Grzegorz, R., Dawid, G., and Aleksander, P.: 'A System for Heart Sounds Classification', PLoS One, 2014, 9(11)

[13] Amit, G., Gavriely, N., and Intrator, N.: 'Cluster analysis and Classification of Heart Sounds', Elsevier, Biomedical Signal Processing and Control, 2009, 4, pp. 26-36

[14] Cota Navin, G.R., Palaniappanb; Sundaram, Swaminathana; Shankar, M. Krishnana;: 'Neural network classification of homomorphic segmented heart sounds', Applied Soft Computing, 2007, 7, (1), pp. 286–297

[15] Roger E. Phillips, J.: 'The Heart and the Circulatory System', in Editor: 'Book The Heart and the Circulatory System' (2014, edn.), pp.

[16] 'https://www.framinghamheartstudy.org/', in Editor: 'Book https://www.framinghamheartstudy.org/' (2015, edn.), pp.

[17] Israel, C.W.: 'Mechanisms of sudden cardiac death', Indian Heart Journal, 2014, (66), pp. 10-17

[18] Mannor, S., Bentley, P., Nordehn, G., and Coimbra, M.: 'The PASCAL Classifying Heart Sounds Challenge', in Editor : 'Book The PASCAL Classifying Heart Sounds Challenge' (2011, edn.), pp.

[19] Sung, P.-H., Thompson, W.R., Wang, J.-N., Wang, J.-F., and Jang, L.-S.: 'Computer-Assisted Auscultation: Patent Ductus Arteriosus Detection Based on Auditory Time–frequency Analysis', Journal of Medical and Biological Engineering, 2015, pp. 76-85

[20] Shenoi, B.A.: 'Introduction to Digital Signal Processing and Filter Design' (John Wiley & Sons Inc, 2005. 2005)

[21] Thede, L.: 'Analog & Digital Filter Design Using C' (Prentice Hall; 1 edition, 1995. 1995)

[22] Butterworth, S.: 'On The Theory of Filter Amplifiers', Experimental Wireless, 1930, pp. 536

[23] Hacısalihoğlu, Y.: 'Halka Rezonatör Yapılarının Kullanıldığı Tamamen Optik Butterworth Ve Chebyshev Filtrelerin Tasarımı', in Editor : 'Book Halka Rezonatör Yapılarının Kullanıldığı Tamamen Optik Butterworth Ve Chebyshev Filtrelerin Tasarımı' (Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2011, edn.), pp.

[24] Piskorowski, J.: 'A new concept of phase-compensated continuous-time Chebyshev filters', Signal Processing, 2008, 88, (3), pp. 437-447

[25] Ağoğlu, E.A.: 'Sayısal Süzgeç Tasarımı ve Uygulamaları', in Editor : 'Book Sayısal Süzgeç Tasarımı ve Uygulamaları' (Ankara: Başkent Üniversitesi Fen Bilimleri Enstitüsü, 2008, edn.), pp.

[26] Yazgan, E., and Korukürek, M.: 'Tıp Elektroniği' (İstanbul Teknik Üniversitesi Rektörlüğü, 1996. 1996)

[27] Koksoon, P., Jianfeng, C., Dat, T.H., and Shue, L.: 'Heart sound as a biometric', Pattern Recognition 41, 2008, pp. 906 – 919

[28] Zhidong, Z., Qinqin, S., and Fangqin, R.: 'Heart Sound Biometric System Based on Marginal Spectrum Analysis', Sensors, 2013, (13), pp. 2530-2551

[29] Smith, S.W.: 'The Scientist and Engineer's Guide to Digital Signal Processing' (Calofornia Technical Publishing, 1999. 1999)

[30] Cortes, C., and Vapnik, V.: 'Support-Vector Networks', Machine Learning,, 1995, 20, pp. 273-297

[31] Akşehirli, Ö.: 'Tıbbi Araştırmalarda Destek Vektör Makinelerinin Kullanımı', in Editor : 'Book Tıbbi Araştırmalarda Destek Vektör Makinelerinin Kullanımı' (Düzce Üniversitesi Sağlık Bilimleri Enstitüsü, 2012, edn.), pp.

[32] Kecman, V.: 'Learning and Soft Computing' (The MIT Press, 2001. 2001)

[33] Duman, F.: 'Destek Vektör Makineleri Yardımıyla İnvaziv EEG Verilerinden Epileptik Nöbet Tahmini', in Editor : 'Book Destek Vektör Makineleri Yardımıyla İnvaziv EEG Verilerinden Epileptik Nöbet Tahmini' (Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü, 2013, edn.), pp.

[34] Demirci, D.A.: 'Destek Vektör Makineleri İle Karakter Tanıma', in Editor: 'Book Destek Vektör Makineleri İle Karakter Tanıma' (Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü., 2007, edn.), pp.

[35] Bertsekas, D.P.: 'Nonlinear Programming, ' (Athena Scientific, 1999. 1999)

[36] R.Palaniappan;K. Sundaraj; S. Sundaraj: 'A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals', BMC Bioinformatics, 2014, 15

[37] COSKUN, H.: 'Developing Android Application With Support Vector Machines Intended For Classification Of Extra Systole', Master's Thesis, Suleyman Demirel University Graduate School Of Natural And Applied Sciences, 2016

Drag Design of Sulfanilamide based on the Technology of Immune Network Modeling and Ontological approach

Samigulina G.A./Chef of the laboratory
«Intellectual control systems and forecasting»
Institute of Information and Computing Technologies, IICT
Almaty, Kazakhstan
galinasamigulina@mail.ru

Samigulina Z.I./Senior lector of the department
«Automation and control»
Kazakh National Research Technical University after K.I. Satpaev, KazNRTU
Almaty, Kazakhstan
zarinasamigulina@mail.ru

***Abstract* - The article is devoted to computer molecular design of new drugs of sulfa group based on immune network modeling of the structure-activity relationship. There are considered sulfonamides with different durations of action. Developed a systematic approach to the treatment of chemical structural information based on artificial immune systems and ontological models. Ontological model are created in ontology editor Protégé.**

***Index Terms*—Drag design, sulfanilamide, immune network modeling, ontological approach, prediction of dependence "structure –property".**

I. Introduction

Nowadays is relevant the development of innovative approaches for computer molecular design of new drugs with desired properties using the latest computer technology, modern information technologies and different methods of artificial intelligence [1]. The high relevance of this problem to pharmacology is determined by huge financial costs and long duration of studies necessary for the identification of new chemical compounds with the desired pharmacological activity [2].

The number of structures of chemical compounds is about 60 million. [3, 4], the number of virtual computer generated molecules is almost one billion, the number of potential targets is around two million [5]. Obviously, the experimental testing of chemical compounds is not possible [6-8]. Therefore, the creation of effective methods of quantitative and qualitative analysis of structural relations for the study of properties of new chemical compounds and computer modeling of molecular structures with a given set of characteristics - are among the most important problems of bioinformatics and cheminformatics.

To describe the structure of chemical compounds there was used a descriptor approach. Molecular descriptors are parameters [9], characterizing physical, chemical and biological properties of the substance. There were studied more than 3000 descriptors and the most important task is to look for informative descriptors to describe the chemical structure of the test chemical substance.

II. Intelligent methods for predicting the dependence of the structure – activity relationship and ontological approach

There are actively developed intelligent approaches of QSAR (quantitative structure-activity relationship of chemical compounds) based on artificial neural networks, genetic algorithms, swarm intelligence algorithms, artificial immune systems.

There were well established neural networks. In work [10] the QSAR models and the QSPR models (Quantitative Structure-Property Relationships) are used to describe forecasting of a particular compound based on artificial neural networks. In work [11], there are given the researches in the field of quantitative relationships between molecular structures and activities of the sets CCR5 of derived inhibitors using two effective methods of non-linear regression and radial basis function of neural networks to evaluate the quantitative structure-property/activity relationships of test compounds. There was given an analysis of input data of the non-linear models by the method of principal components. Experimental and calculated molecular activity indicators for developing new derivatives of inhibitors with high activity and low side-effects, used during the model development stage, showed good correlation, whereby authors developed the PC-RBFNNs QSAR model to calculate the inhibitory activity of some of the compounds.

Together with neural networks there are widely used genetic algorithms as the optimization algorithm to create drugs. In work [12], there is presented a genetic algorithm for the preparation of medicaments using QSAR model based on Bayesian regularized genetic algorithm (Bayesian-regularized genetic neural networks, BRGNN) and genetic algorithm based on support vectors method.

Also, along with genetic algorithms and neural networks there is used a swarm intelligence for the studying "structure-property/activity" relationships of chemical compounds. Due to the fact that the selection of the most informative variables in modeling of QSAR/QSPR is an important task therefore the

work [13] presents a modified swarm optimization method based on multiple linear regression to select a small subset of descriptors.

For the development of new drug compounds there is applied a prospectively biological approach - artificial immune system (AIS). The term artificial immune system refers to information technology, using the theoretical immunology concepts for various applications [14]. Artificial immune systems have the following advantages: memory, learning, distribution, self-organization and a high degree of parallelism. Currently, there were formed three major directions in the AIS [15], based on the theory of adverse selection; clonal selection; theory of the immune network. In work [16] there is shown a detailed theoretical analysis of the main types of AIS algorithms, presented simple examples of the use of these algorithms.

With the use of the data of AIS approaches [17], which use behavioral mechanisms of the human immune system in the process of protecting the body from external factors (pathogens and antigens), there were developed a large number of different methods of multivariate data analysis, image recognition, optimization, etc. Article [18] is devoted to the development of an artificial immune network for multi-site optimization. In work [19] there is developed a multi-site multi-population artificial immune network for double clustering. The feature of this algorithm is to return several sets of equivalent solutions as opposed to a single set of solutions.

With the ability to learn, integrate and synthesize large amounts of disparate information, AIS can be used to predict the properties of unknown compounds when analytical view of the relationship between structure and properties of chemicals is unknown [20]. AIS approach allows modeling the relationship "structure - property/activity" taking into account their non-linear character.

The interdisciplinary nature of QSAR studies affects different areas of knowledge: pharmacology, bioinformatics, molecular biology, chemometrics, statistical analysis, cheminformatics, system analysis, pharmacodynamics, information theory, etc. Therefore, solving the problem of processing of chemical structural information it is prospectively to use the system analysis and ontological models to combine a variety of interdisciplinary approaches and for understanding of problems from different perspectives.

Recently, ontology - based development of intelligent prediction systems is an extremely urgent task. The use of ontologies in the processing of the chemical structural information and the creation of intellectual information system on their basis allows structuring the input and outputting data, to reduce the complexity of creating and maintaining of programs.

Researches [21] are devoted to the consideration of artificial immune systems, as key features and characteristics of biological immune system for solving a number of problems. There were described some approaches for modeling of biological and artificial immune systems. The article discusses the possibility of using the language OWL (Ontology Web Language) for modeling artificial immune systems in ontology editor Protégé. The work [22] deals with the modeling of the immune system and with the important role in understanding the computational aspects, because most of the existing models of artificial immune systems are developed for specific applications and have their advantages and disadvantages. This article presents a model of AIS based on ontology, which combines several models of AIS and algorithms. This model has the exact semantic, general knowledge understandable to the computer. Under the leadership of AIS the ontologies can be created and can dynamically develop more effective AIS models.

Difficulty of immune network modeling technology is that the implementation of the technological chain is possible by different ways, depending on the available data, the formulation of the problem and the available conditions for the realization of the task. Using multialgorithmic approach [23], in which different artificial intelligence approaches can be employed, the classical methods of multivariate data processing (factor analysis, support vector method, etc.), the modularity of the developed software - all of this require the use of the advantages of the ontologies during the creation a component - oriented software [24], implementing the intelligent forecasting technology based on "structure - property/activity" relationships to create new drugs based on the approach of artificial immune system [25].

III. Problem statement

Statement of the problem is formulated as follows: it is necessary to develop an integrated ontological model (OWL model) of immune network technology of chemical structural information data processing based on AIS for computer molecular design of new drugs sulfonamides.

Development of theoretical bases of immune network modeling using the ontological approach is an extremely important task, which will allow on the basis of in-depth analysis of the various ontological models of AIS to create a more efficient computational algorithms of AIS and will significantly simplify the process of creating a component - oriented software for the implementation of intelligent technology of chemical compounds activity forecasting.

As a tool for the development of an integrated model for the processing of chemical information based on immune network technology there was selected ontology editor Protégé.

IV. Immune network modeling technology to predict dependence of the structure-activity of sulfonamides

Sulfa drugs are drugs containing a sulfamide group and, in most cases, benzosulfamida derivatives [26]. These substances belong to the means of broad antibacterial spectrum of action. They were the first chemotherapeutic antibacterial agents and have been successfully used in a number of diseases (malaria, toxoplasmosis and others.). This group of antibiotics suppresses the vital activity of staphylococci (norsulfazol, sulfazol), streptococci (streptocid), meningococcus, Neisseria gonorrhoeae, they are used against intestinal and many other bacteria and viruses. Sulfonamides possess bacteriostatic

properties, that is, do not destroy but inhibit the multiplication of microorganisms. The advent of antibiotics, including penicillin, reduced the use of sulfonamides, but due to the complications of allergic reactions of patients to various antibiotics and getting used to them after repeated application there was actual of the development of new drug compounds of sulfa group. Sulfonamides are still successfully used in the treatment of many diseases, such as pyelonephritis, etc.

Sulfonamides can be classified by duration of action as follows: short-acting (less than 10 hours); the average duration of action (10-24 hours); long-acting (24-48 hours); extremely long duration of action (more than 48 hours). Currently, of particular interest are sulfonamides of long and extremely long action.

Let's consider immune network technology for predicting "structure - activity" relationship of sulfonamides drug compounds represented as an algorithm.

Algorithm:

1) Selecting chemical compounds for research.

2) Description of the structure of the compounds based on descriptors and their classification.

3) Preliminary data processing.

4) Construction of the optimal model (time-series - standards, consisting of descriptors that describe the structure of the selected chemicals with the known properties).

5) Training the artificial immune system and model analysis.

6) Formation of matrices - images.

7) Image recognition [14] based on the singular value decomposition of matrices (SVD).

8) Assessment of the energy errors.

9) Evaluation of the results of prediction of properties of unknown compounds.

10) Selection of compound candidates in the sulfa drug group for further research.

Intelligent technology based on immune network modeling has several advantages:

- The ability of a deep analysis of the hidden (latent) relationships between the descriptors and the underlying factors that affect them;
- recognition of chemical compounds present at the boundary of nonlinear split classes (with similar structures);
- the use of multialgorithmic approach (factor analysis, neural networks, etc.) in the construction of the optimal immune network model;
- reduction of time for training the immune network by constructing an optimal immune network model and reduction of descriptors carrying significant errors [27];
- modular structure of AIS and the ability to expand and reconstruction;
- parallelization of algorithms of service procedures, and work in real time;
- a systematic approach to the unification of best practices in biomedicine and pharmacology, computer technology, the latest achievements of artificial intelligence with the use of an ontological approach and with the use of modern data base of chemical information.

V. Development of Integrated OWL Model in Ontology Editor Protégé

Integrated ontological model AIS_Technology includes three different models:

- the ontological model of the pre-treatment data processing AIS_Preprocessing_Data.
- the ontological model of image recognition based on artificial immune systems AIS_Image_Recognition.
- assessment of the energy error based on homologous proteins AIS_Estimation_energy_error.

As an example, let's consider the ontological model AIS_Preprocessing_Data of chemical structural information pre-treatment.

There is used the Protégé ontology editor, which allows to design the ontology, expand the taxonomic structure of the abstract or concrete classes and slots. The ontology structure corresponds to the classical directory structure. Protégé can generate examples of classes and subclasses. A key objective in the design of the ontology in the editor Protégé is the distribution of classes in taxonomy. Basic concepts of chemical pretreatment of information according to immune network technology of data processing within AIS_Technology ontology will be the objects of ontology in the form of classes. Ontology development begins from the most general terms, and then proceeds to the following detalization and structurization.

In the software product Protégé the root of taxonomic tree is a class THING in which there is the whole set of concepts of the ontology. Figure 1 shows the enlarged hierarchical structure of the ontological model of the chemical information pretreatment, which consists of the following classes: DATA, INPUT_DATA, OUTPUT_DATA and subclasses: Chemical Data, Molecular_descriptors etc.

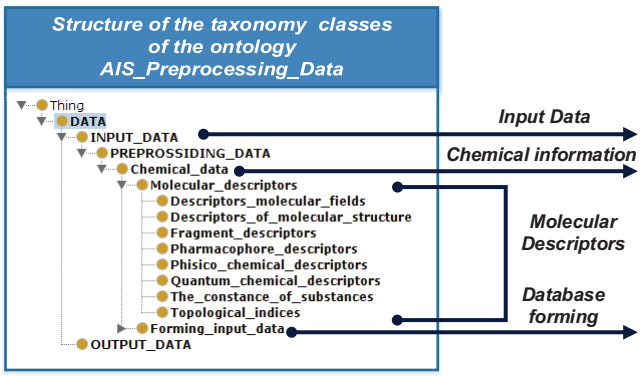


Fig.1. The enlarged hierarchical structure of ontology classes AIS_Preprocessing_Data

Preliminary data processing includes a description of the chemical information in the form of descriptors, quality control of the initial information, during which there are verified the completeness and accuracy of the descriptors, is carried out normalization, filling the missing data in different ways, as well as the selection of informative descriptors based on intellectual and statistical approaches (reduction of uninformative descriptors).

Figure 2 shows a fragment of the taxonomy of the ontology classes AIS_Preprocessing_Data.

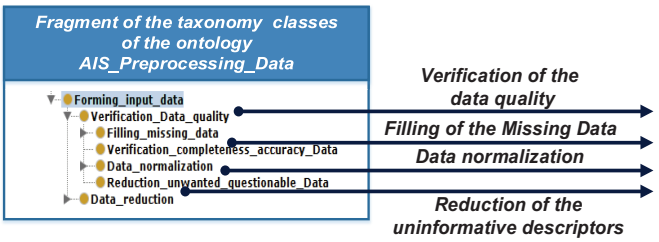


Fig. 2. A fragment of the taxonomy of the ontology classes AIS_Preprocessing_Data

Then there is a detalization of the subclasses and visualization of deployed taxonomy of classes.

On the basis of the developed ontological models there is built an integrated OWL model of AIS presented in Figure 3.

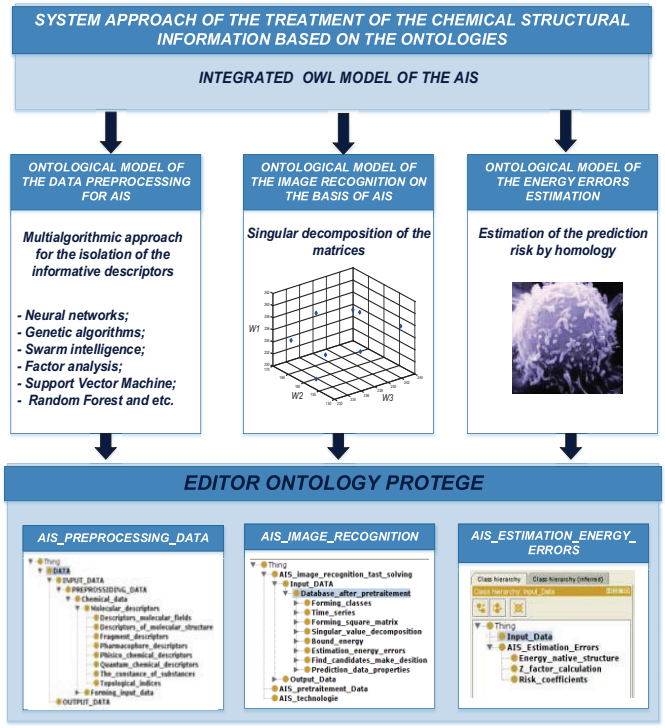


Fig. 3. An integrated OWL model of AIS

Development of component-oriented software using the integrated ontological model allows structuring of input and output data, to take into account the peculiarities of functioning and relationships, save time and computing resources.

VI. Conclusion

Thus, the creation and application of intelligent technology of immune network modeling based on a systematic approach and ontological models gives the ability for more targeted synthesis of medicinal compounds of sulfa group, having the necessary specified pharmacological activity, brings together different approaches to solving this problem, facilitates the creation of software and reduces the cost of new drugs development.

The researches are carried out by the grant of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan under the grant number GR0115RK00549 on the theme: Computer molecular design of drugs based on immune network modeling (2015-2017).

References

[1] K. Gerhard, “Drug design: methodology, concepts and Mode-of Action,” Springer, 850 p., 2013.

[2] S. J. Macalino, V. Gosu, V. Hong, S. Choi, “Role of computer-aided drug design in modern drug discovery,” Journal Archives of Pharmacal Research. Berlin: Springer, vol. 38, №9, pp. 1686-1701, 2015.

[3] L. Blum, J. Reymond, “970 million drug like small molecules for virtual screening in the chemical universe database GDB -13,” J. Am Chem Soc, vol. 131 (25), pp. 8732-8733, 2009.

[4] A. Nonell-Canals, J. Mestres, “In silico target profiling of one billion molecules”, Molecular Informatics, Weinheim, vol. 30(5), pp. 405-409, 2011. 2011.

[5] V. V. Poroikov, D. A. Filimonov, A.A. Lagutin, T. A. Gloriozova, A.V. Zaharov, D. S. Druzhilovsky, A. V. Rudik, A. V. Stepanchikov, “Computer prediction of the biological activity of chemical compounds”.

[6] I. Baskin, A. Varnek, “Fragment descriptors in SAR/QSAR/QSPR studies, molecular similarity analysis and in virtual screening”, Chemoinformatic Approaches to Virtual Screening, RCS Publishing, pp. 1-43, 2008.

[7] M. Waldman, R. Fraczkiewicz, R. Clark, “Tales from the war on error: the art and science of curating QSAR data,” Journal of computer - aided molecular design, Springer, №3, pp. 1-14, 2015.

[8] K. Gerhard, “Drug design: methodology, concepts and Mode-of-Action,” Springer, pp. 850, 2013.

[9] I. Baskin, A. Varnek, “Fragment Descriptors in SAR/QSAR/QSPR studies, molecular similarity analysis and in virtual screening,” Chemoinformatic Approaches to Virtual Screening, RCS Publishing, pp. 1-43, 2008.

[10] N. Montañez-Godínez, A. C. Martínez-Olguín, O. Deeb, R. Garduño-Juárez, G. Ramírez-Galicia, “QSAR/QSPR as an application of artificial neural networks,” Artificial Neural Networks, Springer, vol. 1260, pp. 319-333, 2014.

[11] M. Golani, I. I. Golani, “Neural network ensemble based QSAR model for the BBB challenge: a review,” Transactions on Engineering Technologies, Springer, pp. 55-68, 2015.

[12] M. Fernandez, J. Caballero, L. Fernandez, A. Sarai, “Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM),” Molecular Diversity, Springer, vol. 15, №1, pp. 269-289, 2011.

[13] Khajek, H. Modarres, H. Zeinoddini-Meymand, “Modified particle swarm optimization method for variable selection in QSAR/QSPR studies,” Structural Chemistry, Springer, vol. 24, №5, pp. 1401-1409, 2013.

[14] A. Tarakanov, G. Nicosia, “Foundations of immunocomputing,” Proc. 1st IEEE Symposium on Foundations of Computational Intelligence (FOCI’07), Honolulu, Hawaii, pp. 503–508, 2007.

[15] D. Dasgupta, “Artificial Immune Systems and Applications”, Ed. A. A. Romanyukha, Moscow: Fizmatlit, pp. 344, 2006.

[16] J. Timmis, A. Hone, T. Stibor, E. Clark, “Theoretical advances in artificial immune systems,” Theoretical computer Science, Elsevier, vol. 403, pp. 11-32, 2008.

[17] A. P. Karpenko, D. L. Shurov, "Hybrid method of global optimization based on artificial immune network," Science and education, Electronic Scientific-Technical Journal: MSTU named after N. Bauman, pp. 339-378, 2012.
[18] P. A. Castro, F. J. Zuben, "Mobais: a bayesian artificial immune system for multi – objective optimization," Proceedings of the 7-th Intern. Conference ICARIS, Springer, pp. 48-59, 2008.
[19] G. P. Coelho, F. O. Franca, F. J. Zuben, "A multi – objective multipopulation approach for biclustering," Artificial Immune Systems, Springer, vol. 5132, pp. 71-82, 2008.
[20] O. Ivanciuc, "Artificial immune systems in drug design: recognition of P – glycoprotein substrates with AIRS (Artificial Immune Recognition System)," Internet Electronic Journal of Molecular Design, vol. 5, №11, pp. 542-554, 2006.
[21] M. Husakova, "Artifisial Immune System Model Based on OWL Ontology," Proceedings of the IX conference «Znalosti», Praga, vol. 1, pp. 211-214, 2010.
[22] T. Li, Y. Liang, H. Yang, J. Chen, "An integrated artificial immune system model based on ontology," Proceedings of the Intelligent Computation Technology and Automation (ICICTA), vol. 11, pp. 44-55, 2008.
[23] G. A. Samigulina, Z. I. Samigulina, "Design of immune-network optimal model to predict the unknown properties of medicinal compounds based on multi algorithmic approach," Informatics Issues, Novosibirsk, № 2, pp. 21-29, 2013.
[24] G. A. Samigulina, "Immune network modeling technology for complex objects intellectual control and forecasting system," Monograph. USA: Science Book Publishing House, pp. 172, 2015.
[25] G. A. Samigulina, Z. I. Samigulina, W. Wójcik, U. V. Krak, "Structure-property dependence prediction of new organic compounds based on artificial immune systems," Control and Informatics Issues, Kiev, №2, pp. 81-88, 2015.
[26] V. G. Belikov, "Pharmaceutical chemistry", Pyatigorsk, pp. 720, 2003.
[27] G. A. Samigulina, Z. I. Samigulina, "Computational molecular design of sulfonamides based on immune network modeling," Twelve International Conference on Electronics Computer and Computation (ICECCO), pp. 1-5, 2015.

Generation of bioinspired search procedures for optimization problems

Vladimir Kureichik, Dmitry Zaporozhets, Daria Zaruba
Department of CAD
Southern Federal University
Rostov-on-Don, Russia
vkur@sfedu.ru, elpilasgsm@gmail.com, daria.zaruba@gmail.com

***Abstract* — The paper deals with generation of new hybrid algorithms based on the bioinspired search methods. Subsystem architecture involves a data management module and a module of automation of new search procedures generation. This module is found on genetic search and genetic programming mechanisms. The authors suggest a new approach for description of the data structure based on the XML. As a result it was developed an executable algorithm for optimization problems solving.**

***Index Terms* — bioinspired search, genetic algorithm, atomic search procedure, optimization, XML.**

I. Introduction

In the mid-1980s stochastic search methods based on nature phenomena were widely investigated to increase the effectiveness of optimization problems solution [1]. Bioinspired algorithms involve a parallel processing all alternative solutions and are an alternative to classical search methods that process only one solution. All bioinspired algorithms belong to the class of heuristic algorithms for which convergence to the global optimum is not proved theoretically but empirically established that a probability of optimal and quasioptimal solution is high.

II. General terms of bioinspired search theory

The authors suggest to use the term "agent" as a general name of population elements. In different bioinspired algorithms agents are represented by chromosomes, ants, bacteria, bees etc. A set of agents is a population.

To generate an initial population there are applied determine and random algorithms and agents cover all search space as far as possible.

One of the main advantages of bioinspired algorithms is its module structure. This approach allows to obtain a lot of new versions of the algorithm by the development of new and modification of existing rules for initialization and generation of new agents.

One of the key problems in developing bioinspired algorithms is a balance between algorithm convergence rate (intensification) and search diversification (diversity of agents). The high rate of convergence requires a decrease of agents diversity in a population, as well as diversification deals with more wide set of agents which cover all search space and has a great probability of finding a global solution [1, 2]. Diversification principle saves a diversity of agents in population during a great number of iteration of the bioinspired algorithm.

Adaptation solves the problem of balance between a convergence rate and search width. Adaptation methods implement a regular changing of algorithms setting parameters so that is realized a serial transfer from diversification at initial steps of the bioinspired algorithm to intensification at final iterations.

Today, a perspective means for bioinspired algorithm creation is a hybridization of bionic algorithms to obtain new structures increasing probability of global optimum finding by archiving of balance between intensification and diversification at different stage of hybrid search [1,3].

III. Generation of bioinspired search procedure

The general system structure of generation of bioinspired search procedure is shown on Fig. 1.

In the Block #1 a decision-making person inputs standardized descriptions of initial hierarchical search mechanisms and atomic search procedures (ASP). Information about ASP saves in ASP database (ASPDB) and standardized descriptions used for generation of hierarchical search procedures.

In the Block #2 on the basis of input data the system generates new hierarchical procedures and uses all ASP containing in ASPDB. At the third stage (Block #3) an obtained hierarchical search procedure is compiled.

Input data here is results obtained in the Block #2 and Block #5 represented as a standardized description. In the Block #5 a decision-making person may denote a function for generation of initial encoding solution and an objective function *calculate()* [1, 4]. These functions are implemented with taking into account a given problem. Generation of initial encoding solution contains two stages: search of solutions and encoding of solutions [5]. Compilation module modifies information in XML-file into finite sequence of actions which executes in the Block #4. As a result we obtain an executive algorithm.

The Block #4 is a module for estimation of obtained results.

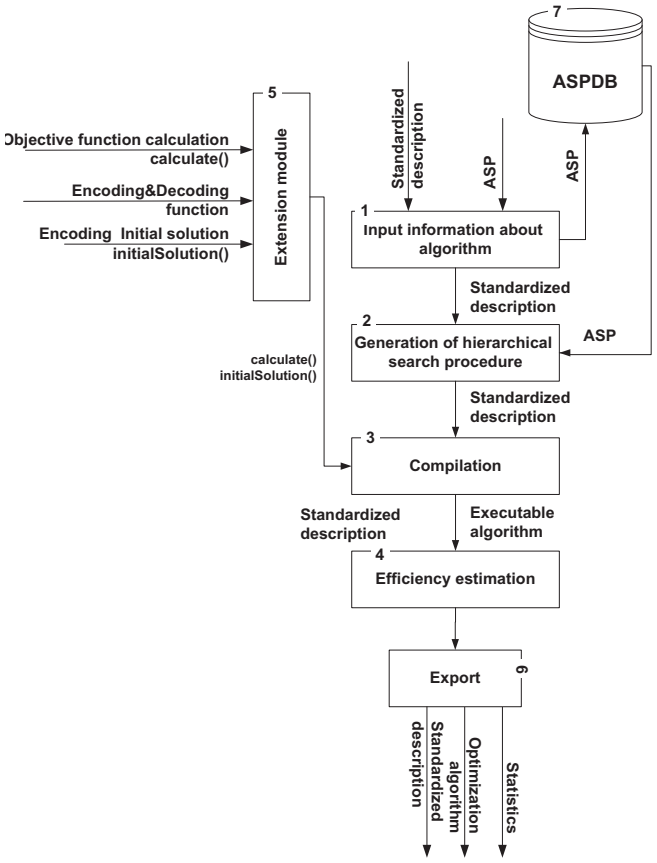


Fig. 1. General system structure of generation of bioinspired search procedure.

The Block #6 is responsible for output data. With the use of this module a decision making person may obtain information about generation of the algorithm, standardized description of obtained hierarchical mechanism, executive algorithm as a DLL and information about setting parameters which were used during the creation of new hierarchical mechanisms.

The authors carried out an analysis of derived architecture and make a conclusion that for decision making person the most labor-intensive stage is an input of initial information. At this stage a decision making person need to developed optimization hierarchy, calculate theoretical estimation of effectiveness and represent an algorithm according with a standardized description. This information input one time and can be used and modified later within the developing system.

The authors suggest a new optimization approach based on the hybridization of various ASP. Here ASP is an algorithm with unitized interface which returns one of several optimal or quasi-optimal solutions. The ASP interface (ASP_{intf}) can be written as

$$ASP_{intf} = \langle id, solution, objectiveFunction, attributesList \rangle,$$

where *id* is an algorithm identifier, *solution* is an initial encoded solution, *objectiveFunction* is a attribute list (for example, a number of iteration).

To obtain the effective hierarchical architecture based on the ASP a search hierarchy is describe by XML (2008) [6].

The authors suggest following instances in order to have an opportunity create sequential and parallel search procedures: *flow* – an instance describing global search procedure, *subflow* – an instance describing different processes within global search procedure. There are various interactions between instances. Parent interaction points at an instance which is a parent one for a current instance. A parent instance is an instance which initiates a current instance after the completion of it work. Correspondingly, child interaction points at an instance which is a child one for a current instance.

An instance *foreach* can initiate several processes dynamically.

Here is an example demonstrate the suggested description of hierarchical optimization procedures on the basis of a VLSI placement problem [6]. At the first stage it is used a bees algorithm (BA), then obtained quasi-optimal solution optimized by a genetic algorithm (GA). There are a function for generation of an initial encoded solution (*initialSolution()*) and function for objective function calculation (*calculate()*). Necessary parameters for the bees algorithm are a number of population (iteration = 100), a number of blocks (block = 10) and a number of bees (bees = 500). Necessary parameters for the genetic algorithm are a number of iteration (iteration = 50), a size of population (population = 100). Values of parameters are given as an example. Let us formalized algorithms according with the ASP interface and XML instances.

The BA is written as:

<search id=”beeAlgorithm”
solution=”initialSolution()”
objectiveFunction=”calculate()”
attributesList=”iteration=100,block=10,bees=500” />.

The GA is written as:

<search id=”geneticAlgorith”
solution=”result(beeAlgorith)”
objectiveFunction=”calculate()”
attributesList=”iteration=50,population=100”/>,

Here *result(beeAlgorithm)* is a result of the BA. The XML-description of the suggested algorithm, represented in Table 1, is as on Fig. 3.

At this case an instance with an attribute *parent="InitialPoint"* is an instance *<flowid="mainFlow" parent="InitialPoint" child = "beeAlgorithmFlow">* (lane 3, Table 1), where *mainFlow* is an identifier of global search space.

This instance starts the process with *beeAlgorithmFlow* identifier. Within this process the ASP *beeAlgorithm* is implemented with predetermined parameters. After that it is performed the GA *<subflow id = "geneticAlgorithmFlow" child = "geneticAlgorithm" parent = "beeAlgorithmFlow">* (line 19, Table 1) since corresponding process has the attribute *parent ="beeAlgorithmFlow".* The identifier *mainFlow* completes the GA and the *flow*. This event completes algorithms *<end id="finishPoint" parent="mainFlow"/>* (lane 23, Table 1) because mark end id="finishPoint" is linked in the parent instance *mainFlow(parent="mainFlow").*

TABLE 1. XML-DESCRIPTION

1	*<algorithm>*
2	*<start id="InitialPoint"/>*
3	*<flow id=" mainFlow" parent="InitialPoint" child="beeAlgorithmFlow">*
4	*<subflow id="beeAlgorithmFlow" child="beeAlgorithm">*
5	*<search id="beeAlgorithm" solution="initialSolution()" objectiveFunction="calculate()" attributesList="iteration=100,block=10,bees=500"/>*
6	*</subflow>*
7	*<subflow id="geneticAlgorithmFlow" child="geneticAlgorithm" parent="beeAlgorithmFlow">*
8	*<search id="geneticAlgorith" solution="result(beeAlgorith)" objectiveFunction="calculate()" attributesList="iteration=50,population=100"/>*
9	*</subflow>*
10	*</flow>*
11	*<end id="finishPoint" parent="mainFlow"/>*
12	*</algorithm>*

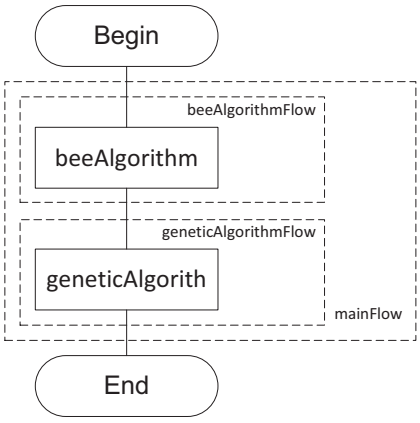


Fig. 2. Representation of XML-description

IV. GENERATION OF HIERARCHICAL SEARCH PROCEDURES

In general case automation of hierarchical search procedures generation contains following stages:

1. Definition of a set of ASP (a_i). At this stage a decision making person denote a set of ASP which includes in the finite optimization algorithm.

2. Definition of function of optimization algorithm assessment $\Theta(\psi(F))$, where F is a resulting optimization algorithm, ψ is an assessment function (criterion) with known global and quasi-optimal solutions.

$$\Delta = \psi_{best} - \psi_{result}$$

where ψ_{best} is a known global or quasi-solution solution, ψ_{result} is the best solution obtained by developed GA. Obviously, for minimization problem if $\Delta < 0$ a resulting algorithm is not effective in comparison with algorithm which obtain ψ_{best}.

3. Input of restrictions $C(c_i)$ (for example, time restrictions or specific restrictions).

4. A view of solutions *VIEW* for optimization algorithm. In other words, it is necessary to define data structure.

In the paper to simplify generation of hierarchical search procedure the authors suggest following assumptions:

1. As an assessment function $\psi(F)$ uses scalar mathematical functions such as Ackley's, Rastrigin's and Schwefel's functions etc. These functions are widely used in international scientific community for assessment quality of obtained optimization mechanisms.

2. As a search mechanism for obtaining new optimization algorithm it is used genetic algorithm with modified genetic operators. The authors select the genetic algorithm because description of hierarchical search procedures can be written as a chromosome; a number of ASP in the algorithm do not exceed 10-15 elements; GA time complexity is not high for small number of input parameters; as a result there are obtained a set of alternative solutions which give an opportunity for a decision making person to choose the best algorithm.

3. Solutions are represented as a numerical, binary and vector chromosome (*VIEW*) according to arguments of $\psi(F)$ function.

To create new solution the authors developed new mechanisms for generation of hierarchical search procedures. They allow to represent standard descriptions as alternative solutions (chromosome). Developing encoding and decoding structure should correspond to following requirements:

- encoding and decoding speed. Encoding/decoding algorithm time complexity should not exceed nlog(n) because the GA can process a lot of alternative solutions and each of which should be encoded at least one time for its estimation;
- a data integrity. The encoding and decoding algorithm should be realized so as after sequential applying of encoding and decoding methods an initial solution is identical to resulting one;
- a simplicity of encoded representation. Hierarchical search procedures should be encoded so as there is easy to make random changes (mutation operator) and unite two hierarchical search procedures (crossover operator).

To create effective encoding/decoding mechanism the authors create to use elements of genetic programming theory. [10, 11]. Traditionally, any algorithm can be represented as a tree. Each tree node contains a function, each leave contains an operand. An expression represented as a tree can be read recursively [11, 12].

In case of the given problem tree nodes are interpreted as a links between ASP and leaves – as ASP. To reduce a quantity of encoding information and simplification of genetic operators work the authors suggest a two-level encoding mechanism [13]. At the first level there is implemented a simplification procedure, but at the second level – encoding of simplify hierarchical search procedures.

The simplification procedure consists in the following. The XML standard enjoins to apply additional instances for structuring of hierarchical search procedures. Such structure as *algotithm, start, end*, *flow* and *subflow* are necessary for the

XML document but superfluous for the algorithm of generation. So, in this work there are deleted superfluous data in terms of encoding. Let us consider an example presenting in Table 1. After the simplification the description is written as shown in Table 2.

TABLE 2. XML-DESCRIPTION

	<search id="beeAlgorithm" solution="initialSolution()" objectiveFunction="calculate()" attributesList="iteration=100,block=10,bees=500"/>
	<search id="geneticAlgorith" solution="result(beeAlgorith)" objectiveFunction="calculate()" attributesList="iteration=50,population=100"/>

According to the results of Table 2, representation of hierarchical search procedures as a tree is shown in Fig. 3.

In Fig. 3 the vertex S points at relations between algorithms. Let introduce two types of relations: S – sequential and P – parallel. Note, relations are not commutative, hence, an order of operands is importance. As for the example, the *beeAlgoritm* is initial data for the *geneticAlgorithm*.

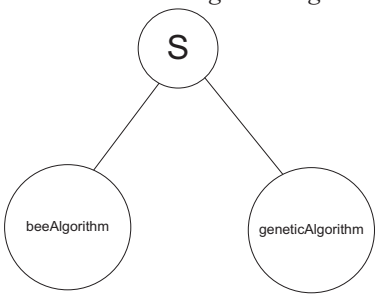


Fig. 3. Representation of simplify hierarchical search procedures as a tree

As a data structure in the chromosome it is suggested to use a Reverse Polish Notation (RPN). The RPN is a notation expression in which operands are located before operators. Here, an operand is ASP, an operator is a relation between ASP.

V. GENETIC ALGORITHM FOR GENERATION OF HIERARCHICAL SEARCH PROCEDURES.

To generate hierarchical optimization algorithm the authors suggest to use a modified genetic algorithm [5, 7]. As an input data it is used developed hierarchical structures. The main goal of automated generation is automated creation of new multilevel search mechanisms by crossing existing results, and estimation of its running time and solution quality thanks to new algorithms for finding a global optimum of scalar functions [3, 5, 9].

At the first stage a decision making person carries in initial data (search mechanisms) for the genetic algorithm. These procedures are represented as standardized description, considered above, and can be encoded and decoded in appropriate for GA way. The selection of search mechanisms is a significant step because definite features of initial data are passed on to descendants as a result of the GA work [5, 7].

At the next step a generation system creates a list of ASP and calculates configuration parameters in ASPDB. Then, a decision making person can specify values of these parameters and correct it. Such parameters contain probabilities of genetic operators, a number of iteration etc [8].

When system gets all necessary information there are calculated objective functions (estimations) for each hierarchical optimization algorithm. According to obtain values modified operators of crossover and mutation are implemented [9].

New solutions are decoded and estimated which are main information for a selection operator based on the “roulette wheel” [2, 7]. This type of selection allows ineffective solutions to get to the finite sample. Ineffective solutions have a good influence on population genotype and do not allow to fall into a local optimum. After that the process is continued iteratively until predetermined number of iteration is reached.

At last step there is a selection of the best solution according to the genetic algorithm results. The best solution is decoded and represented by decision making person as standardized description of the optimization algorithm. The result can be changed by decision making person or accepted in unmodified view.

VI. CONCLUSION

In the paper the authors considered main points of the bioinspired search theory as well as distinguished fundamental mechanisms for creation algorithms inspired by biological system simulation. Also it was developed the new system for generation of new search mechanisms. The general standardized data structure based on XML description was made. The developed data structure allows to manage information about hybrid search mechanisms. Also the authors provide examples of standardized description. To generate new hierarchical search procedures it was developed the modified genetic algorithm. Modified encoding and decoding algorithm and genetic operators were integrated to reduce algorithms running time.

ACKNOWLEDGMENT

This research is supported by grants of the Ministry of Education and Science of the Russian Federation, the project #8.823.2014.

REFERENCES

[1] Karpenko A.P. Modern search optimization algorithms. Algorithms inspired by nature– Moscow, Bauman MTSU, 2014. – 446 s.

[2] Kureichik, V.M., Malioukov, S.P., Kureichik, V.V., Malioukov, Genetic Algorithms for Applied CAD Problems. USA: Studies in Computational Intelligence, 2009.

[3] Zaporozhets, D., Zaruba, D.V., Kureichik, V.V. Hybrid bionic algorithms for solving problems of parametric optimization // World Applied Sciences Journal, 23 (8), pp. 1032-1036., 2013

[4] Bova, V.V., Kureichik, V.V., Lezhebokov, A.A. The integrated model of representation of problem-oriented knowledge in information systems (2014) 8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 - Conference Proceedings.

[5] Goldberg David E. Genetic Algorithm in Search, Optimization and Machine Learning. USA.: Addison-Wesley Publishing Company. Ind., 1989.

[6] Спецификация языка описания схем XML (XML Schema Definition – XSD) *http://www.w3.org/2001XMLShcema* .

[7] V.V. Kureichik and V.V. Kureichik Jr., "Bionic search in the design and management", Proceedings of Southern Federal University. Technical science, vol. 11 (136), pp. 178-183, 2012.

[8] V.V. Kureichik and D.Y. Zaporozhets, "Swarm algorithm for optimization problems", Proceedings of Southern Federal University. Technical science, vol. 7 (108), pp. 28-32, 2010.

[9] Zaporozhets, D.U., Zaruba, D.V., Kureichik, V.V. Representation of solutions in genetic VLSI placement algorithms // (2014) Proceedings of IEEE East-West Design and Test Symposium, EWDTS 2014.

[10] E.V. Kuliev, A.A. Lezhebokov and A.N. Dukkardt, "The approach to the study of neighborhoods in the swarm algorithms for solving optimization problems", Proceedings of Southern Federal University. Technical science, vol. 7 (156), pp. 15-26, 2014.

[11] E.V. Kuliev, D.Y. Zaporozhets and A.M. Ksalov, "Bioinspired search to solve the problem of accommodation of VLSI components", Proceedings of the Kabardino-Balkar Scientific Centre of Russian Academy of Sciences, vol. 6(62), pp. 58-65, 2014.

[12] E.V. Kuliev and A.A. Lezhebokov, "Investigation of the characteristics of the hybrid placement algorithm", Proceedings of Southern Federal University. Technical science, vol. 3, pp. 255-261, 2013.

[13] V.V. Kureichik and D.Y. Zaporozhets, "Modern problems when placing elements of VLSI", Proceedings of Southern Federal University. Technical science, vol. 7 (120), pp. 68-73, 2011.

Using Intelligent systems in Pharmacy

Sagdoldanova Aiym
Department of Information Systems Management
Kazakh-British Technical University
Almaty, Kazakhstan
aiym.sagdoldanova@gmail.com

Atymtayeva Lyazzat
Department of Information Systems Management
Kazakh-British Technical University
Almaty, Kazakhstan
l.atymtayeva@gmail.com

***Abstract*—This paper is devoted to the issues of the constructing the expert systems in the area of pharmacy. The use of expert systems in pharmacy allows reach a new level in the process of intelligent search the appropriate pharmacy products for the clients and pharmacists.**

***Index Terms*—Expert system, pharmacy market, intelligent search.**

I. Introduction

Nowadays, science and technology are developing with a high speed. As an example we can take an artificial intelligence, task is the understanding of human intelligence. In other words, artificial intelligence is a scientific field, which develops methods, that allows electronic computers to solve intellectual problems, if they are solved by man. Furthermore, artificial intelligence as a scientific discipline consists of several major trends. It should be noted, that nowadays as a progressive direction of artificial intelligence have become expert systems. Purpose of expert system is forming and outputting the recommendations depending on the current situation, which describes a collection of data and also user input data interactively. Moreover, recommendations issued by the computer must conform to the recommendations of highly qualified specialist.

Today on the pharmaceutical market we can observe a lot of different systems that work with storing and extracting the data. For example, in Kazakhstan there are some web-oriented systems like vidal.kz, i-teka.kz and eurapharma.kz, which allow searching drugs, sorting the search results by price and give some special information about this drug. These types of systems use only data management and processing. But in the most cases the pharmacists and clients need more intelligent results and recommendations about the proposed drugs. For instance, the relevance certain type of drugs for people with some contraindications by health. Also as everyone knows, that pharmacist's life like feeling tired or lacking of information that may lead to make an error in drug selecting or dispending. That is why the creation of expert systems in Kazakhstan's pharmacy market will be useful for pharmacists, minimizing their efforts and time. In other words, system will greatly facilitate the decision making process of the pharmacists, especially in unusual situations when the selection needs to be more accurate and complete.

II. Expert System Structure

To develop such type of expert system we have considered the following components like user interface, inference engine and knowledge base.

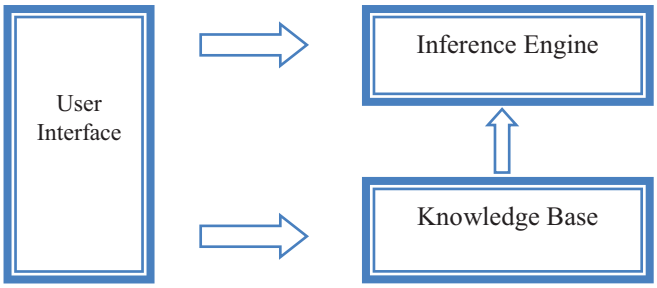


Figure 1. Major parts of an expert system

User Interface is a tool, which shows relationship between the user and system. In other words, user interface enables users to interact with our system. Interface of this system simple and easy to use.

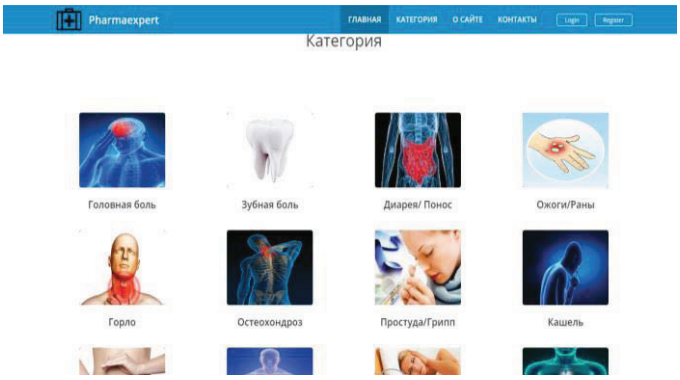


Figure 2. Category of diseases

In the main page of this system shows the list of all disease's categories (Figure 2). After selecting one disease, appears another page which shows a list of contraindications only selected category drugs (Figure 3). In this web-page a pharmacist can choose one or more contraindications, also at the client's discretion, he can choose other options, such as a form of release and price of drugs.

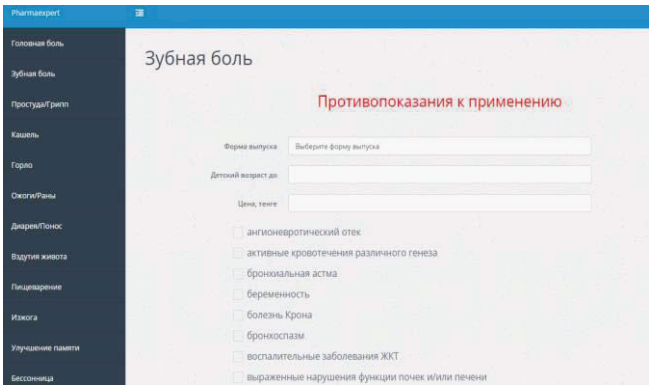


Figure 3. Page of contraindications

After clicking one or more type of contraindications, appears recommendation page (Figure 4), which shows one suitable drug with full instruction including type of drug, price, indications, dosage, adverse effects and contraindications.

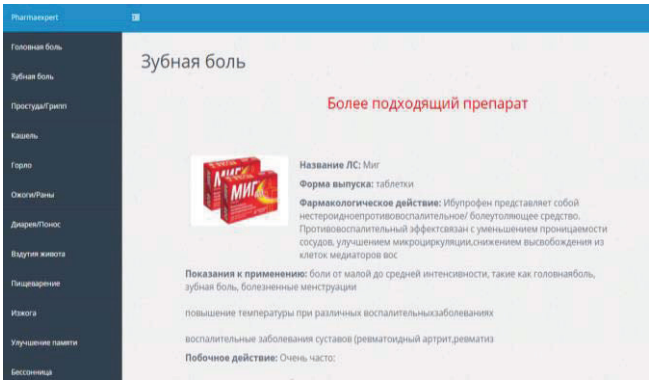


Figure 4. Recommendation page

Also, as shown in table (Figure 5) pharmacist can recommend other drugs if client not satisfied with prices or other instructions of drug. It means system shows other drugs as sub-drugs which have same number of symptoms, contraindications with essential. Furthermore, it should be noted, that in table presents only drugs from chosen category.

Figure 5. Table of sub-drugs

Furthermore, in our system we considered:

- ✓ only frequent diseases, in which people don't need the doctor's prescription;
- ✓ only such drugs which are available without doctor's prescription.

Likewise, how this user interface works and how they will interact with each other depends on inference engine. Because inference engine is a main brain of our system, also it plays main role to get all appropriate information from database.

III. Inference Engine

Inference Engine is a mechanism that's manages the expert systems and decide if the problem has reached an acceptable solution or not [3]. In our case we use the classification by categorizing all the drugs. Also, in order to give recommendation system will use if-then rule by counting the numbers of contraindications and fuzzy logic rule. For instance, if the number of contraindications in one drug more than others in selected category, then system will remove this type of drugs and recommends only one suitable drug with minimum number of contraindications by using classification. Furthermore, we considered the Dempster-Shafer's theory of believe and decided to improve this system by using this theory.

IV. Knowledge Base Forming

The core of the proposed expert system is Database. It is very important to know how data are stored in database and by which logic or rules they interact with the interface. If inference engine is a universal thinking machine, then knowledge base is the over what this machine will think [2].

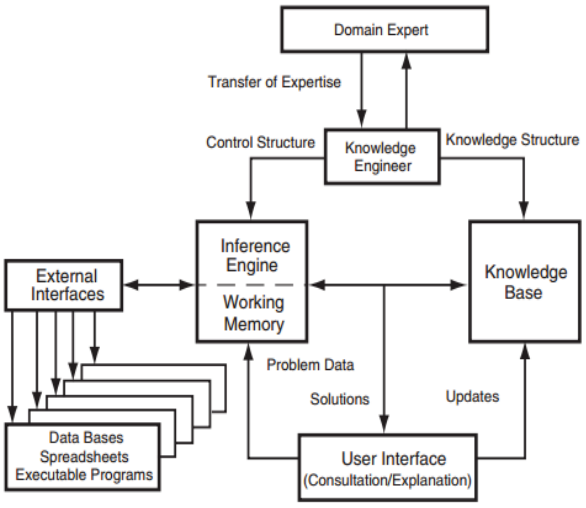


Figure 6. Integration of ES components

That is why the success and effectiveness of our expert system mostly depends on the completeness and the quality of the information stored in the system's knowledge base. In our knowledge base part, we have a table with column names like:

- name of drugs;
- type of drugs;
- price;
- pharmacological properties;
- indications;

- adverse effects;
- dosage;
- contraindications.

V. STEPS IN THE SYSTEM DEVELOPMENT PROCESS

All tasks implemented step by step in terms of time.

1) Problem selection:

At this stage:

- the problem area is defined;
- specialists - experts are selected;
- the group of developers are selected;
- the detailed plan of development prepares.

As we noted above, in some cases pharmacists can make an error in drug selecting or dispending. Also when we come to the pharmacy and ask them an instruction of one drug, in most cases pharmacists become to search this drug and start reading instruction of drug. It takes more time, especially when we are in hurry. In our point of view, it is one of the problems on pharmaceutical market in our country.

2) Development of ES prototype (Prototyping):

The prototype system is a reduced version of ES and correctness representation of facts, designed for check, communications and strategy of experts [1]. Development of ES prototype is divided into six stages:

- identification of a problem;
- extraction of knowledge;
- conceptualization (structuring) of knowledge;
- formalization;
- realization of a prototype;
- testing.

a) Identification of a problem – acquaintance and training the group members of a developers, and also creation of informal formulations of a problem. At this stage our tasks are specified and planned the course of development ES prototype. Moreover, in this stage are defined:

- resources (time, people, etc.);
- sources of knowledge (book, additional experts);
- the available similar ES;
- classes of solvable tasks, etc.

During the implementation of our project we considered all available and similar ES in the sphere pharmacy as DoseChecker, PharmADE, Microbiology, etc. Also we have considered all sources and have consulted with opinions of experts who work in pharmacies and of course have consulted with programmers to know which programming languages or platforms are better to realize ES.

b) Extracting the knowledge - getting the all possible knowledge concepts from engineer about the certain type of subject area. To extract the knowledge engineer uses various methods: analysis of texts, conversations, lectures, discussions, interview, observation, and others.

c) Conceptualization (or structuring) knowledge - development of non-formal description of domain knowledge in the form of a graph, table, or chart, which reflects the basic concepts and the relationship between the concepts of the subject area. At this stage, are determined: the terminology, the list of basic concepts and their attributes, relationships between concepts, the structure of the input and output of information, decision-making strategy, etc.

Firstly, we collected all dates about the drugs and by means of an experts' knowledge have divided them into categories.

	Заболевания	Название ЛС	Форма выпуска	Цена	Фармакологическое действие	Показания к применению
2	Головная боль	Анальгин	таблетки 500 мг:10шт	30 тг.	Анальгетик-антипиретик. Является производным	боли различного происхождения (головная
3		Аспирин	таблетки 500 мг: 20 шт	40 тг.	Ацетилсалициловая кислота относится к группе НПВС,	– для снижения риска заболеваемости и смертности
4		Парацетамол	таблетки 200 мг: 10шт таблетки 500 мг: 10шт	30 тг.	Парацетамол – анальгетик-антипиретик. Оказывает	– болевой синдром: при головной боли, мигрени,
5		Ибупрофен	таблетки, покр. оболочкой	45 тг.	Оказывает противовоспалительное,	Воспалительно-дегенеративные заболевания
6		Цитрамон П	таблетки, 10 шт	30 тг.	Цитрамон комбинированный препарат,	–болевой синдром слабой и умеренной интенсивности
7		Темпалгин	таблетки, 500 мг: 20шт	370 тг.	Анальгетик-антипиретик комбинированного состава.	– болевой синдром, особенно у больных с
8		Нурофен	таб. шипучие 200 мг: 10 шт.	810 тг.	НПВС, производное фенилпропионовой кислоты.	Воспалительно-дегенеративные заболевания
9		Фервекс	порошок д/пригот.р-ра д/приема внутрь: 8	1450 тг.	Эффект препарата обусловлен антигистаминным действием,	Для лечения простуды, ринита, ринофарингита и

Figure 7. Category of drugs

Also, it should be noted that we have considered only frequent diseases and drugs which are available without doctor's prescription in our country. In figure 7 shown the table of categories of all drugs, which are used in database part.

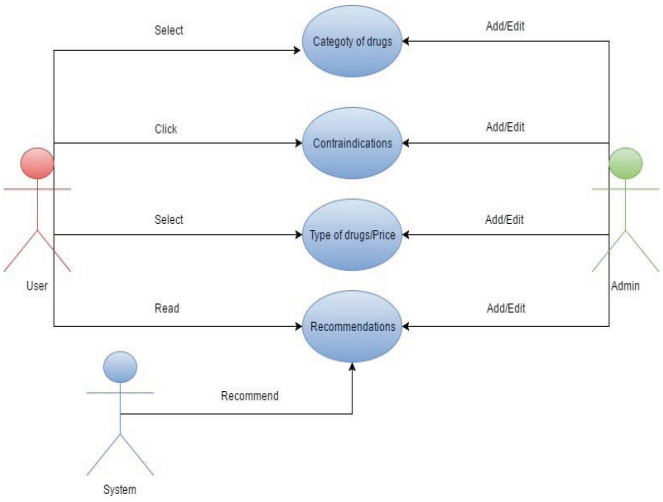


Figure 8. Class diagram of system

In figure 8 shown class diagram of project, where admin can add and edit all dates and user can see it. Also system can recommend only one suitable drug, using dates which are added by admin.

d) Formalization of knowledge is a development of knowledge base in language of knowledge representation. In this stage are used: logical methods, productional models, semantic models, frames, the object focused languages.

As a logical method we have decided to use if-then rule and fuzzy logic. Because for the beginning developers of expert systems, this logic rules are more suitable.

e) Realization of a prototype – development of complex program showing the viability of approach in general. At this stage ES prototype (including the knowledge base, other

program modules) is created with the help of: programming languages (traditional, specialized), tools of development of ES, "empty" covers of ES.

In order to store all dates in database and to create a web-site we stopped to use MySQL and PHP. Because, today an expert system developers have an opportunity to develop, realize their applications via the web also. Since all data are stored in database, this language is suitable to connect with database and to solve complex problems.

f) Testing – is a process of identification mistakes in approach and realization of a prototype. The prototype is checked on: convenience and adequacy of the interface of input-output, quality of test examples, completeness and consistency of rules in the knowledge base.

During the testing we have detected some mistakes. For example, our system not considered the type of drugs after clicking contraindication part. Also, in contraindications page (Figure 3) appeared all contraindication of all added drugs. It was some kind of mistake because one contraindication shouldn't have repeated twice in selected category. In this stage we have corrected all our mistakes which have appeared in a testing time.

3) Development of a prototype to industrial ES:

The main work at this stage consists in extension of the knowledge base. After establishment of the main structure of ES knowledge the knowledge engineer starts development and adaptation of interfaces by means of which system will communicate with the user and with the expert. The system should give to the user an opportunity to specify the unclear moments, to stop work and etc.

In this project as experts pharmacists will add categories, add new drugs to the existing castegories, also edit dates.

Users can choose one category from list of them, then click contraindications by chosen category and if they need, can choose type of drugs and price. After that system will recommend them most suitable drug by couting number of contraindications.

4) Stage joining of system:

The ES connection with other programs is carried out in the environment at which it will work. For confirmation of system usefulness, it is important to give to each of users an opportunity to give some real tasks to ES to check how it will carry out them. Joining includes ensuring communication of ES with the existing databases and other systems at the enterprise.

In this stage, we connected all works that we have done. For example, database part we connected with user interface part by using logical rules as if-then and fuzzy logic.

5) Support of system:

Complete systems in order to improve its performance and increase portability can transcode into another language (such as C), but this will reduce its flexibility. This can be done with the systems, which are designed for problem areas where knowledge is not changed. If the subject area for a system is changed, then it is necessary to maintain the tool in that environment in which it was created.

VI. Using Software Tools And System Effectiveness

Today, expert systems developers have the opportunity to distribute their applications via Web. Since all data are stored in the database, we should choose and connect it with one web based programming language. The development of the most web based expert systems embodies a number of benefits. For instance, the use of an internet based database was effective in storing large amount of facts and data for web based expert systems, also using PHP or HTML makes it easier to enhance the expert system user interface [7]. After analyzing and testing various types of Web oriented programming languages, we stopped at the PHP, because this language is simple to use and allows complex conditions and complex conclusions. Additionally, to PHP we used programming languages like Laravel (framework of PHP), JavaScript, HTML, CSS, MySQL, bootstrap, OOP, MVC. By using this languages and production model of the expert system we can achieve approximately 70% of the efficiency of the proposed recommendations. In order to increase effectiveness of the system we plan to expand database in terms of diseases and recommendations for them.

VII. Conclusion

The development of such types of expert systems relates to the different issues of the data store structure, its relationships representations, and relevance of the appropriate rules models [5] that serve as a base for intelligent search and filtering. The main idea is to find the relevant links between the data stored, special conditions (rules) and proposed solutions (recommendations). Thus, further researches in this area may discover some new options for selection process and facilitate the process of choosing the essential drugs.

During the performance of this work all the tasks have been achieved set of objectives. As a result of this work created expert system as web application(pharmaexpert.kz) in the sphere pharmacy, which can greatly facilitate the decision making process.

It is supposed that this expert system will be actual for pharmacists, minimizing their efforts, time and likelihood of making errors during the selecting and dispending drugs to clients.

In the course of performance of this work have been studied the ES development environments, logical rules and their application, also general principles of creation knowledge base.

Furthermore, ES for pharmacy is developed with the purpose to facilitate and to automate activity of experts in order to avoid errors due to the human factors. Unlike the person, our system has a smaller probability of mistakes. Also, in the future we planned develop it as a mobile application and expand dates about diseases and symptoms.

EXPERT'S OPINION

Expert #1: Development such type of system in the field of pharmacy is very good. But as we all know medicine is not staying in one place, it means, every day have been created new drugs and are included to the pharmaceutical market. I would like to, that this system has become a self-learning and working as experts in the future. And of course, people should entrust their health to this system, how they are entrusted to the doctors during the inspection.

Expert #2: In my opinion, system should include not only such diseases in which people don't need a doctor's prescription, but also it should be able to connect with hospitals, in order to get doctor's diagnosis for other diseases.

Expert#3: You should to develop it with the help of experts in the field of medicine and pharmacy, because they know better than others. And of course in your system not enough drugs, symptoms and contraindications. I think if you develop this system with experts it will be great and people start to use and believe to your system.

Expert #4: I think that you should consider that one tablet has a similar instruction with other tablets. So you have to deal with such situations, for example in case when person by his health reasons cannot use some kind of drugs then he cannot use the tablets that have the same properties.

REFERENCES

[1] Abraham, A. "Rule-Based Expert Systems. Handbook of measuring system design", 2005.
[2] Ajith Abraham "Rule –based Expert Systems"
[3] Bayan Bakyr Elyas "Expert pharmacy for drug interactions and diagnoses", 2015.
[4] Expert pharmacy for drug interactions and diagnoses [Electronic resource]- https://prezi.com/_7zujyc8wkyz/expert-pharmacy-for-drug-interactions-and-diagnoses/
[5] Ioannis M Dokas "Developing web sites for web based Expert Systems: a web engineering approach"
[6] Sheriyev M., Atymtayeva L. Automation of HCI Engineering processes: System Architecture and Knowledge Representation // Int.Journal "Advanced Engineering Technology and Application (AETA)", Natural Science Publishing, Vol.4, N2 (May 2015), ISSN 2090-9535, pp. 41-46.
[7] Y Duan, J S Edwards, M X Xu "Web-based Expert Systems: benefits and challenges", 2003

Utilizing Fuzzy Colored Petri-nets to Monitor Cardiac Pacemaker Behavior

Negar Majma
PhD. Candidate, Department of Computer, University of Kashan, Kashan, Iran
Instructor, Naghshejahan Higher Education Institute, Isfahan, Iran
Majma@grad.kashanu.ac.ir

Seyed Morteza Babamir
Associate Prof., Department of Computer, University of Kashan, Kashan, Iran
Babamir@kashanu.ac.ir

Amirhassan Monadjemi
Associate Prof. Department of Artificial Intelligence, Faculty of Computer Engineering University of Isfahan, Isfahan, 81746, Iran
Monadjemi@eng.ui.ac.ir

***Abstract*—Connected health devices are used to in vivo cure and prevent abnormal conditions without manual intervention. This paper presents an intelligent agent-based method for runtime verification of cardiac pacemaker runtime behavior. Construct the agent's knowledge base based on the Fuzzy Colored Petri Net (FCPN). Based on our previous experiences, the FCPN will be reduced the scale of our network in comparison to the Petri-net(PN). Compared to a simple inference engine, the FCPN can cover the concurrent states and in addition intelligent agent can ensure the accuracy of the runtime verification operation of the cardiac pacemaker in vital and unexpected situation.**

Index Terms — ***Runtime Verification, Cardiac Pacemaker, Fuzzy Colored Petri-net, Software agent***

I. INTRODUCTION

A cardiac pacemaker is a device helping regulation of the heart beat by connecting to the cardiac vessels. It is used to cure arrhythmia (irregularity in the normal heart rate) and implanted under the skin by a physician. It is critical to make sure that the cardiac pacemaker functions accurately since it is set only once when it is implanted. Afterward, the device is assumed to be able to work properly for five to ten years without any physical access to it. 2,294 fault cases and 1,154,451 side effects cases were reported by Food and Drug Administration (FDA) between 2006 and 2011[1]. 92,600 cases of these reports include injuries, and 4,590 cases are fatal. Modern medical devices have software applications controlling their behavior and activities. 1,210 (22.9%) of these faults were blamed on the medical device software applications[1].

The above mentioned problems of a medical device can be dealt with in two ways: (1) static verification (2) runtime verification. In the first way, the correctness of limitations and needs of the users are considered. However, the software may have an error while running due to the possibility of error in its implementation. Proving the correctness of the software is complicated and not easily possible due to the high number of the running states of the software. In the second way, the software runtime behaviors are examined. The runtime verification studies, develops and implements the verification approaches which allow examining the user needs, goals, and limitations. Regarding the above mentioned items, runtime verification can be a solution to verify cardiac pacemaker runtime behavior.

In this paper, runtime verification is accomplished by a software agent, which momentarily controls accuracy of the target software running, regarding the software inputs' situations in order to prevent any error and unsafe state. This software agent uses a knowledgebase which has the accurate process of the medical device behavior. One of the knowledge representation techniques is Petri-nets (PN). Regarding the advantages of PN in software modeling[2]–[6], PN is used for represent the knowledgebase of the software agent in this paper. Regarding the experiences of using: 1) Fuzzy Petri-nets (FPN) to verify insulin pumps [7], 2) Colored Petri-nets (CPN) to verify the cardiac pacemaker [8] and 3) Software agent with PN knowledgebase to verify the cardiac pacemaker [9], which we gained during the previous studies, the runtime behavior of the cardiac pacemaker software is examined by a software agent with FCPN knowledgebase in this paper. This is considered a developing step compared to the earlier work.

The organization of this paper is as follows. In Section II an overview of backgrounds is presented. Next in Section III, the proposed method is discussed. Finally the two remaining sections contain the related work and conclusion.

II. BACKGROUNDS

A. Fuzzy Expert System

The cardiac pacemaker is a fuzzy expert system as it makes decisions based on the input values and the knowledge existing in its knowledgebase and acts accordingly. In the real world, the variables of the cardiac pacemaker are fuzzy. The membership functions of input and output parameters of the cardiac pacemaker are described in Section III. Fuzzy knowledge base consists fuzzy rules in form of IF-THEN statements. This part has been shown as a FCPN in this paper.

B. Fuzzy Colored Petri-nets

PN theory provides a graphical language for describing software needs and limitations. It is easy to combine PN approach with other techniques and theories including programming, fuzzy theory, and neural networks [10]. A FPN is formally defined as 8-tuple $FPN = (P, T, D, I, O, f, \alpha, \beta)$. FCPN is a special kind of FPN. In FCPN, tokens, places, and transitions can be colored [11].

In this paper, FCPN is used for presentation of the cardiac pacemaker knowledgebase. PN is used in a fuzzy form because of uncertain quantities measured by input sensors. Therefore, the

uncertain rules exist in decision making. The number of input places are decreased and decision making are made simple by using CPN, compared to the non-colored PNs. In this paper, FCPN is employed to represent the knowledgebase of the software agent, which will be discussed in Section III.

C. Agents

By definition, an agent is a computational process including software program or an actual robot. The agent connects with the environment, controls its behavior, and reacts automatically to the environment. The main features of the agent – e.g. autonomy and intelligence - make them easy for medical usages. An intelligent agent is capable to recognize if the cardiac pacemaker functions properly by collecting data from sensors and comparing them with its knowledgebase. The agent may put the pacemaker in safe mode in the case of risky and unexpected situations. Here, the agent is used in the form of runtime verification and its knowledgebase is developed by FCPN, which determines the exact state of cardiac pacemaker behavior

III. PROPOSED METHOD

A software agent was used to verify accurate function of the cardiac pacemaker and to decide about the data received from the sensors. The decision making process is based on FCPN, which indicates the exact state of the device. Fig. 1 illustrates diagram of the proposed method. Clearly, the received data from the sensors is fed to the agent and the cardiac pacemaker simultaneously and the software agent checks output of the pacemaker before it is applied to the heart. Runtime verification is featured with an agent and a knowledgebase that checks if the heart rate is accurate. The knowledgebase is developed based on the FCPN. The parameters that influence heart rate are discussed in the following subsections and FCPN is presented in subsection B. Fig. 2 pictures the pseudo-code.

A. Input and output parameters of cardiac pacemaker

The cardiac pacemaker does a complicated function, which can be studied in[12]. Five parameters of cardiac pacemaker are considered as the input parameters in this paper. Each of them is divided into a few diverse subsets for more accurate fuzzy estimation. In Table 1, the proposed categorization of activity level can be seen. The subset of the other parameters are not shown due to the limitations of the paper.

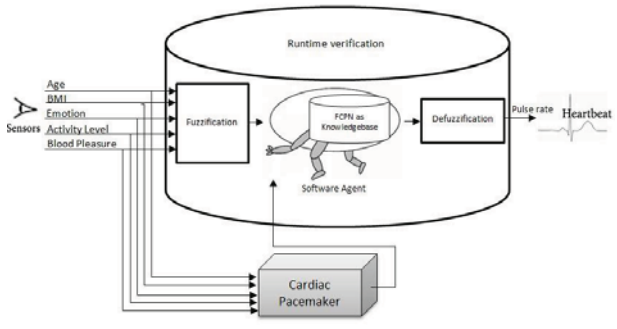


Figure 1. Diagram of the proposed method

```
RuntimeVerification
{
    Array SensorData[5]
    SensorData[1]= ReadSensorAge
    SensorData[2]= ReadSensorBMI
    SensorData[3]= ReadSensorEmotion
    SensorData[4]= ReadSensorActivityLevel
    SensorData[5]= ReadSensorBloodPresure
    PacemakerOutput=Pacemaker(SensorData)
    Fuzzification (SensorData)
    AgentOutput=Agent(SensorData, PacemakerOutput)
    If  AgentOutput=true then
                Apply_heart_rate_to_the_heart()
    Else
                Decide_for_the_correct_heart_rate()
                Apply_correct_heart_rate_to_the_heart()
    End if
}

Agent(SensorData, PacemakerOutput)
{
    PulseRate=Run_FCPN(SensorDate)
    Defuzzification(PulseRate)
    If PulseRate = PacemakerOutput then
       Return true
    Else
      Return false
    End if
}
```

Figure 2. Pseudocode of proposed method

TABLE 1. THE ACTIVITY LEVEL FOR PACEMAKER

Title	Support values
Level 0 (Normal zone)	0-15
Level 1 (weight management zone)	10-25
Level 2 (Healthy hearth zone)	23-50
Level 3 (Aerobic zone)	45-75
Level 4 (Anaerobic Zone)	70-95
Level 5 (Red zone)	90-100

- Age

Age is categorized into five ranges, each transforms to a membership function set. The allowable age range is from 0 to 100.

- Body Mass Index (BMI)

The second parameter that determines the human heart rate is the body mass index. This index is obtained using height and weight. This index is divided into 4 levels: (1) Light-weight (below), (2) Normal-weight (Normal), (3) Over-weight, and (4) Obese.

- Activity Level

Activity level shows the patient activity level and is between 0 and 100, using an activity sensor referred to as Accelerometer [13] which is the best way to identify the activity level of a body.

- Emotion

Emotions are considered as effective factors in human heart rate. Feelings are categorized into four states.

- Blood Pressure

Blood pressure is one of the most effective factors in the heart rate.

B. FCNP for Pacemaker

In this Section, knowledgebase representation with FCPN is discussed. A FCPN is an 8-tuple,

$FCPN= (P, T, C, D, I, O, \alpha, \beta)$ Where

$P= \{P_1, P_2, \ldots, P_k\}$ is a finite set of places;

$T= \{T_1, T_2, \ldots, T_L\}$ is a finite set of transition;

$P \cup T \neq \emptyset$;

$C=\{C_1,C_2,\ldots,C_i\}$ is a finite set of linguistic variable where I is the number of linguistic variables where $C_j=\{C_{j1},C_{j2},\ldots,C_{jk}\}$, $C_j \in C$ and j is a one of the linguistic variable of C.

$D_{Pi}=\{(C_{xy}, \alpha(C_{xy}))\}$ is a finite set of token is place P_i that has color where $D_{Pi} \in P$ and $C_{xy} \in C_j$

$I:P \rightarrow T$ is an input function for mapping places to transitions.

$O:T \rightarrow P$ is an output function for mapping from transitions to places.

$\alpha:C_{ij} \rightarrow [0,1]$ is an association function for mapping from a linguistic variable to fuzzy truth values; between zero and one.

$\beta:T \rightarrow [0,1]$ is an association function for mapping from transitions to fuzzy truth values; between zero and one.

In a FCPN, enabled and firing condition are important and are defined as:

Enabled condition: transition *t* is enabled if $D_{I(P \rightarrow t)} \neq \emptyset$ (i.e., data tokens are available in P_1), then transition *t* is enabled.

Firing condition: An enabled transition fires when its condition is met. When transition *t* fires, then a new color token $D=\{(C_y, \gamma(C_y))\}$ appears in $P_{O(t \rightarrow P)}$ where

$$\gamma(C_y) = min\left\{C_{xy_{I1(p \rightarrow t)}}. C_{xy_{I2(p \rightarrow t)}}. \ldots . C_{xy_{Iz(p \rightarrow t)}}\right\} * \beta(t)$$

C. Mapping Fuzzy rule into FCPN

The mapping of the below fuzzy production rules into the FCPN is shown in Fig. 3.

R_1) *IF X is* C_1 (α_1) *THEN Y is* C_2 (α_2). ($CF= \beta_1$)

C_1 represents antecedent part which is consisted of one or more propositions connected either by *"AND"* or *"OR."*. C_2 denotes consequent proposition; while C_1 and C_2 are the terms including fuzzy variables (e.g. high and low). α_1 and α_2 are the validity scale antecedent and consequent term between zero and one. The β_1 determines the certainty factor of the rule.

For example, consider the following rule:

Rule1: IF Age is Teenager (0.7) AND Bmi is Normal (0.5) AND Activity_level is level4 (0.4) THEN Pulse is Normal (CF=0.9).

FCPN of the Rule1 is defined as follows:

FCPN= (P, T, C, D, I, O, α, β)

Where

$P= \{P_{Age}, P_{Bmi}, P_{Activity\text{-}level}, P_{Puls}\}$, $T=\{T_{Rule1}\}$,

C= {Age, BMI, AC},

C_{Age}= {'Child', 'Teenager', 'Young', 'Middle-age', 'Aged'},

C_{BMI}= {'Below', 'Normal', 'Over', 'Obese'},

C_{AC}= {'Level0', 'Level1', 'Level2', 'Level3', 'Level4', 'Level5'},

C_{Pulse}= {'Low', 'Medium', 'Normal', 'High'},

I: $P_{Age} \rightarrow T_{Rule1}$, $P_{Bmi} \rightarrow T_{Rule1}$, $P_{Activity\text{-}level} \rightarrow T_{Rule1}$

O: $T_{Rule1} \rightarrow P_{Pulse}$

Figure 3. FCPN of the rule R_1

α ('Teenager') = 0.7, α('Normal') = 0.5, α('Level4') = 0.4

$D_{P_{Age}} = \{('Teenager'. 0.7)\}$

$D_{P_{BMI}} = \{('Normal'. 0.5)\}$

$D_{P_{AC}} = \{('Level4'. 0.4)\}$

$\beta(T_{Rule1}) = 0.9$

Transition T_{Rule1} is enable since $D_{P_{Age}} \neq \emptyset$, $D_{P_{Bmi}} \neq \emptyset$ and $D_{P_{Activity\_level}} \neq \emptyset$. If transition T_{Rule1} fires, a token with color $D=\{'Normal', \gamma('Normal')\}$ appears in P_{Pulse} where

$$\gamma('Noraml') = \min\{\alpha('Teenager'). \alpha('Below'). \alpha('Level0')\} * \beta(T_{Rule1}) = 0.36 \quad (1)$$

Each of the three propositions in premises (*Age is Teenager (0.7) AND Bmi is Normal (0.5) AND Activity_level is Level4 (0.4))* of the *Rule1* is shown by a place in Fig. 4. Variables are shown as color tokens with two parts in any places. The first part is linguistic variable and the second part is the Certainty Factor (CF). In that example there are three tokens in places *Age, BMI and Activity_Level(AC)* with values *Teenager*, *Normal* and *Level4* respectively (Fig 4(a)). Transition T_{Rule1} can be fired if the condition of the transition is true and also required tokens are in its input places. If the transition T_{Rule1} fires, a token will move to the destination place. The value of that token in the destination place is *Normal* and its CF is *0.36*. CF is calculated as shown in (1). There are 1440 rules regarding the different zone of each fuzzy variable, however, some fuzzy variables are irrelevant in some rules, so this number has been reduced to 277. Fig. 5 represents part of these rules and part of the FCPN of cardiac pacemaker represented in Fig.6. To use the software agent for runtime verification, it is needed that the sensor values are fuzzificated before entrance. Also the output values are defuzzificated in order that the required crisp value is given to the heart. The following scenario is designed to evaluate the proposed method. We implemented the proposed method in C# and connected it with CPN-tools software.

Scenario: Consider a *sixty-two* years old patient with body mass *thirty*, and activity level *normal*. He/she is *anxious* with blood pressure *high*. For this situation there is no path in FCPN, then intelligent agent can select nearest path from knowledge base and produce the output heart rate 124 with below rule, while if we used a rule based inference engine there is no output and then the patient will be in vital situation.

IV. Relater Work

In our previous research the FPN was used to verify the insulin pump behavior[7]. The insulin pump is a device used for diabetes patients to regulate their blood sugar. This device measures the blood sugar and other factors such as the insulin produced by the

human body and four other parameters. Then it calculates the insulin which needs to be injected.

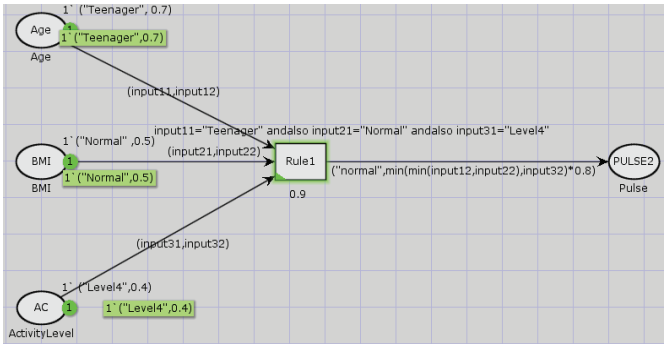


Figure 4(a). Rule1 before firing

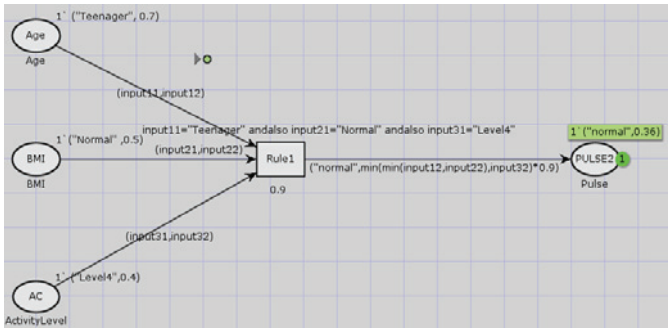


Figure 4(b) Rule1 after Firing

Figure 4. Showing the rule1 in FCPN

FPN was employed to examine the correctness of runtime function since the behavior of the insulin pump is uncertain. In this study, FPN and CPN will be combined, to verify the pacemaker behavior. A software agent with PN is also presented by the authors previously in[9] that only had got three parameters. In this paper, these parameters increased to five, and the structure changed to FCPN. In [8] the pacemaker verification has been done using the CPN, in this paper the previous method has been improved and the verification has got a more accurate structure than the previous study. PN is used to control the system behavior of the insulin pumps in [14] and [15]. The accurate behavior of the insulin pump is modeled and its inaccurate behavior is identified if the system safety is not assured by employing PN. In [16] a method presents to examine the medical equipment behavior using PN to track system operational and security needs. The hardware implementation corresponding to the PN is presented in [17] which provides a fast implementation of the PN. In [18] designing of a rule-based system to make decision about Arteriovenous Shunt Stenosis diagnosis is presented. Also the identification and control systems in nuclear power plants are automatically monitored by an FCPN [11]. In [19] fault identification modeling using an FCPN is applied. Cardiac pacemakers are considered as real time systems using timed automata in [20]. Software testing using the agents is done in [21], and [22]. In [21] utilizes the advantages of agent-based systems like modularity, independency, and parallel activation and establishes a multi-agent system for software testing. In [22] describes a technique which simplifies the software testing using agents. In [23] presents a methods of test data generation employing an agent based method applied on the UML diagrams.

100. If (Age is Middle-age) and (BMI is Normal) and (Activity-Level is Level0) and (Emotion is Ex-Sad-Happy) and (Blood-Pressure is Low) then (Pulse is Medium) (1)
101. If (Age is Middle-age) and (BMI is Normal) and (Activity-Level is Level0) and (Emotion is Ex-Sad-Happy) and (Blood-Pressure is Normal) then (Pulse is High) (1)
102. If (Age is Middle-age) and (BMI is Normal) and (Activity-Level is Level0) and (Emotion is Ex-Sad-Happy) and (Blood-Pressure is High) then (Pulse is High) (1)
103. If (Age is Middle-age) and (BMI is Normal) and (Activity-Level is Level0) and (Emotion is Anxious) and (Blood-Pressure is Low) then (Pulse is High) (1)
104. If (Age is Middle-age) and (BMI is Normal) and (Activity-Level is Level0) and (Emotion is Anxious) and (Blood-Pressure is Normal) then (Pulse is High) (1)

Figure 5. Part of pacemaker rules

Figure 6. Part of FCPN of the pacemaker

In [24], and [25] web-based software testing describe using agents. application of agents is a suitable testing method due to the disruption and complexity of web-based software. In [26] agents are used to test distributed systems. Agents are automatically trained and are located in the distributed system and test each member. In [27] PN is used to test the analysis and design phases of a multi-agent system. Medical devices are considered as an agent in the multi agent system in [28]. This analogy is due to the ability of measurement, monitoring, detection (identification), response, and communication characterizing these devices. An algorithm is implemented using fuzzy neural networks which led to increased accuracy and efficiency in [29]. In [30], a hardware to control the adaptive rate of generated pulse by the pacemaker based on fuzzy logic is presented.

IV. Conclusion And Future Work

A practical approach was proposed to monitor behavior of medical devices. The best approach to do this is runtime verification through which behavior of software is checked and any case of divergence from the safe mode is adjusted. With regard to the implanted medical devices that are assumed to function properly without even the slightest changes, the presence of an intelligent software to lead decision making by the devise is critical. Finally, by applying the proposed method, we can verify the runtime verification of medical device such as cardiac pacemaker using an FCPN. Compared to a simple inference engine, the FCPN can cover the concurrent states. This case is not achievable in one short running of the inference engine and the inference engine has to be run twice or acts in a parallel form in two processors for examining the concurrent states. The experimental studies of this paper also show the superiority of this

method. In this paper, five effective criteria in measuring the heart rate are considered. It should be noted that other parameters can be considered in the future works.

REFERENCES

[1] H. Alemzadeh, R. K. Iyer, Z. Kalbarczyk, and J. Raman, "Analysis of Safety-Critical Computer Failures in Medical Devices," *IEEE Security and Privacy*, vol. 11, pp. 14–26, 2013.

[2] S. J. H. Yang, J. J. P. Tsai, and C.-C. Chen, "Fuzzy Rule Base Systems Verification Using High-Level Petri-nets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 2, pp. 457–473, 2003.

[3] X. Li and F. Lara-Rosano, "Adaptive Fuzzy Petri nets for Dynamic Knowledge Representation and Inference," *Expert Systems with Applications*, vol. 19, no. 3, pp. 235–241, 2000.

[4] S.M. C. S.-M. Chen, J.-S. KE, and J.-F. CHANG, "Knowledge Representation Using Fuzzy Petri Nets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 2, no. 3, pp. 311–319, 1990.

[5] H.C. Liu, Q.-L. Lin, L.-X. Mao, and Z.-Y. Zhang, "Dynamic Adaptive Fuzzy Petri Nets for Knowledge Representation and Reasoning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 6, pp. 1399–1410, 2013.

[6] X. L. X. Li, W. Y. W. Yu, and F. Lara-Rosano, "Dynamic Knowledge Inference and Learning Under Adaptive Fuzzy Petri-net Framework," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 30, no. 4, pp. 442–450, 2000.

[7] N. Majma and S. M. Babamir, "Specification and Verification of Medical Monitoring System Using Petri-nets.," *Journal of medical signals and sensors*, vol. 4, no. 3, pp. 181–193, 2014.

[8] N. Majma, S. M. Babamir, and A. Monadjemi, "Runtime Verification of Pacemaker Using Fuzzy logic and Colored Petri-nets," *4th Iranian Joint Congress on Fuzzy and Intelligent Systems*, 2015.

[9] N. Majma and S. M. Babamir, "Medical Software Runtime Checking Using Petri-nets & Software Agents," *4th International Conference on Computer and Knowledge Engineering*, pp. 449–454, 2014.

[10] K.-Q. Zhou, A. M. Zain, and L.-P. Mo, "A Decomposition Algorithm of Fuzzy Petri net Using an Index Function and Incidence Matrix," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3980–3990, 2015.

[11] S. J. Lee and P. H. Seong, "Development of Automated Operating Procedure System Using Fuzzy Colored Petri nets For Nuclear Power Plants," *Annals of Nuclear Energy*, vol. 31, no. 8, pp. 849–869, 2004.

[12] N. K. Singh, M. Lawford, T. S. E. Maibaum, and A. Wassyng, "Formal Evaluation of the Cardiac Pacemaker Resynchronization Therapy, " *Centre for Software Certification McMaster Centre for Software Certification,* no. October, 2014.

[13] W. V. Shi and M. Zhou, "Body Sensors Applied in Pacemakers: A Survey," *IEEE Sensors Journal*, vol. 12, no. 6, pp. 1817–1827, 2012.

[14] S. M. Babamir, "A Framework for Specifying Safe Behavior of the CIIP Medical System.," *Advances in experimental medicine and biology*, vol. 696, pp. 637–644, 2011.

[15] S. M. Babamir, "Constructing a Model-Based Software Monitor for the Insulin Pump Behavior.," *Journal of medical systems*, vol. 36, no. 2, pp. 829–840, 2012.

[16] P. E. S. Barbosa, M. Morais, K. Galdino, M. A. Andrade, L. Gomes, F. Moutinho, and J. C. a. de Figueiredo, "Towards Medical Device Behavioural Validation Using Petri Nets," *26th International Symposium on Computer-Based Medical Systems*, pp. 4 – 10, 2013.

[17] C. Chen, C. Lin, and H. Tsai, "A Rule-Based Expert System With Colored Petri Net Models for Distribution System Service Restoration," *IEEE Transactions on Power Systems,*vol. 17, no. 4, pp. 1073–1080, 2002.

[18] W. L. Chen, C. D. Kan, C. H. Lin, and T. Chen, "A Rule-based Decision-making Diagnosis System to Evaluate Arteriovenous Shunt Stenosis for Hemodialysis Treatment of Patients Using Fuzzy Petri nets," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 703–713, 2014.

[19] M.-S. Ouali, D. Ait-Kadi, and N. Rezg, "Fault Diagnosis Model Based on Petri net with Fuzzy Colors," *Computers & Industrial Engineering*, vol. 37, no. 1–2, pp. 173–176, 1999.

[20] E. J. E. Jee, S. W. S. Wang, J. K. K. J. K. Kim, J. L. J. Lee, O. Sokolsky, and I. L. I. Lee, "A Safety-Assured Development Approach for Real-Time Software," *16th International Conference on Embedded and Real-Time Computing Systems and Applications*, 2010.

[21] D. Ponnurangam and U. Anbarasan, "Evolution of Agent-Oriented Distributed Model for Software testing," *the International Arab Journal of Information Technology*, vol. 3, 2006.

[22] K. Karnavel, V. Divya, and P. Karthika, "Agent Based Software Testing Framework (ABSTF) for Application Maintenance," *International Journal of Research in Engineering & Advanced Technology,* vol. 1, no. 1, pp. 1–4, 2013.

[23] S. Yu and J. Ai, "Software Test Data Generation Based On Multi-Agent," *International Journal of Software Engineering and Its Applications,* vol. 4, no. 1, pp. 67–74, 2010.

[24] Y. Qi, D. Kung, and E. Wong, "An Agent-based Testing Approach for Web Applications Department of Computer Science," *Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05)*, 2005.

[25] S. Greenwood, "A Multi-agent Software Engineering Environment for Testing Web-based Applications," in *27th Annual International Computer Software and Applications Conference.*, pp. 210–215, 2003.

[26] S. Enyedi, L. Miclea, and I. Ştefan, "Agent-based Testing and Repair of Heterogeneous Distributed Systems," *International Conference on Automation, Quality and Testing, Robotics*, vol. 1, pp. 104–108, 2008.

[27] B. Athamena and Z. Houhamdi, "A Petri net Based Multi-Agent System Behavioral Testing," *Modern Applied Science*, vol. 6, no. 3, pp. 46–57, 2012.

[28] T. Drew and M. Gini, "Implantable Medical Devices as Agents and Part of Multiagent Systems," *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems - AAMAS '06*, p. 1534, 2006.

[29] J. W. Shin, J. H. Yoon, and Y. R. Yoon, "Rate-adaptive Pacemaker Controlled by Motion and Respiratory Rate Using Neuro-fuzzy Algorithm," *Medical and Biological Engineering and Computing*, vol. 39, no. 6, pp. 694–699, 2001.

[30] A. Wojtasik, Z. Jaworski, W. Kuźmicz, A. Wielgus, A. Wałkanis, and D. Sarna, "Fuzzy Logic Controller for Rate-Adaptive Heart Pacemaker," *Applied Soft Computing Journal*, vol. 4, pp. 259–270, 2004.

The Mesothelioma Disease Diagnosis with Artificial Intelligence Methods

Hamza Osman Ilhan
Department of Computer Engineering
Yildiz Technical University
Istanbul, Turkey
hoilhan@yildiz.edu.tr

Enes Celik
Department of Computer Programming
Kirklareli University
Kirklareli, Turkey
enes.celik@klu.edu.tr

***Abstract* — Asbestos is a carcinogenic substance, and threatens human health. Malignant Mesothelioma disease is one of the most dangerous kind of cancer caused by asbestos mineral. The most common symptom of the disease, progressive shortness of breath and constant pain. Early treatment and diagnosis are necessary. Otherwise, the disease can lead people to die in a short period of time. In this paper, different types of artificial intelligence methods are compared for effective Malignant Mesothelioma's diseases classification. Support Vector Machine, Neural Network and Decision Tree methods are selected in terms of regular machine learning concept. Additionally, Bagging and Adaboost re-sampling within ensemble learning terminology is also adapted. Totally 324 Malignant Mesothelioma data which consists of 34 features is used in this study. K-fold cross-validation technique is performed to compute the performance of the algorithms with different K values. 100% classification accuracies are obtained from three tested methods; Support Vector Machine, Decision Tree and Bagging. Additionally, the process time of methods are measured in case of using method in lots of data. In this sense, methods are evaluated based on accuracy and time complexity. The results of this paper are also compared with previous studies using same Malignant Mesothelioma's dataset.**

Keywords— Malignant Mesothelioma, Support Vector Machine, Decision Tree, Neural Network, Ensemble Learning

I. INTRODUCTION

Malignant Mesothelioma (MM) is one of the cancer type. It appears on the thin layer of tissue and rapidly affects to various internal organs [1]. Lining parts of lungs and the chest wall is the most infected parts and organs in cases [2] [3]. Different symptoms such as difficulties in breathing, affliction in chest wall, cough, bloated abdomen, exhausted morality, extremely loss in weight etc. can be seen. Disease advances rapidly while the symptoms appear slowly [4].

The asbestos mineral plays important role on mesothelioma disease. According to medical report, 80% of disease is caused by the mineral [3]. More exposure to mineral increase the risk of developing the disease. In this sense, people living in industrialized countries encounters more than small towns. More specifically, disease is mostly seen in miners and produces who deals with the asbestos mineral. Normally, incubation stage of the disease is around 40 years for [3]. The late awareness of Malignant Mesothelioma disease has made it impossible to diagnosis.

The diagnosis is performed by observation of the X-ray images of chest and the scan findings of computed tomography. In both techniques, doctors mainly examine the fluid produced by the cancer in results or the tissue obtained by biopsy [4].

Addition to regular techniques, computerized methods are also utilized in few studies. Currently, computer based diagnosis systems, which named as Computer Assisted Systems (CAS) become more popular due to high accurate, consistent and efficient results [5]. CAS mainly employs the artificial intelligence methods such as Support Vector Machine (SVM), Decision Tree (DT), Neural Networks (NN) etc. on the stored numerical data. Similar to various medical application, MM disease diagnose is, basically, also a significant classification problem. Methods might conclude different results according to arranged data [6]. In this sense, in order to define the useful method for the corresponding data, several artificial algorithms need to be tested.

In the study, the classification of the data for the Malignant Mesothelioma disease is performed and test results is compared. This study also provides a decision support system, which contributes to the doctors in their diagnosis decisions. Paper is organized as follow; current studies over MM disease diagnose are presented in Section 2. Methods used in testing are briefly explained in Section 3 with data information. Results and explanations are given in Section 4. Paper is concluded with future works and final decisions as last chapter.

II. LITERATURE REVIEW

Visual investigation technique on the diagnosis of medical images is a time-consuming and subjective procedure. Experiences of doctors play effective roles on decision step. In this sense, using the image processing algorithms and artificial intelligent methods prevent diagnoses from different decisions of doctors such as in computed tomography analyses. Computer based technique presented in [7] easily identifies the pleural contours and detects pleural thickenings with two steps. Firstly, they detect the thorax and then remove the air and trachea. In both steps, they implemented 3D morphological operations. According to paper, image retrieval system over MM diagnose is a promising method to detect the disease.

Another study published by Chen et al. [8] explains the implementation of the random walk-based segmentation [9] method. They used mesothelioma computed tomography image datasets and aimed to establish an automatic segmentation. They observed the progression of the disease by volumetric assessments to decide the treatments. Similar to this approach, Onama et al. used 3D version of random walk-based segmentation method on PET images [10]. They aimed to increase success rates for the detection of Lung Tumor.

Er et al. used numerical dataset instead of images. They adapted probabilistic neural networks (PNN) for using in the diagnosis of MM disease. They compared the results to multilayer and learning vector quantization neural networks. They reported in [11] that PNN is evaluated as best classifier with 96.30% accuracy.

A different approach to MM disease diagnose is presented in [12] by K. Chaisaowong et al. They observed the contours of the pleura form in healthy and patient cases. According to comparison of tracing, they detected the thickenings. In this meaning, they formed a tissue-specific segmentation by implementation of the 3D Gibbs-Markov random field (GMRF) [13]. It is adopted to distinguish thickenings from thoracic tissue. Then, morphometric analyses and volumetric assessments are performed to 3D modeling. According to results of the paper, authors assure that the automated approach can help physicians to diagnose pleural mesothelioma in its early stage.

In another study, Principal Component Analysis (PCA) and Brain Emotional Learning (BEL) network is employed to classification of gene-expression microarray data [14]. The proposed combination of methods are utilized for the small round blue cell tumors (SRBCTs), high grade gliomas (HGG), lung, colon and breast cancer classification. Malignant pleural mesothelioma records are located in the lung cancer dataset. According to the results, the PCA–BEL classified the data with an average accuracy of 100%, 96%, 98.32%, 87.40% and 88% respectively [14].

III. METHODOLOGY

Currently, several machine learning algorithms are already utilized for mesothelioma dataset. However, classification results might be increased with other methods. Hence, in this study, different machine learning methods tested on mesothelioma dataset. Methods are selected due to not applied on dataset before. Hereby, in case of more accurate results, method can be used for advanced diagnosis.

Five fundamental classification methods are tested in this study. Methods are categorized into two titles: a.) Machine-learning and b.) Ensemble-learning methods. The brief descriptions of the used methods and parameter arrangement are separately explained in following sub-sections.

A. Machine Learning Methods

A great deal of machine learning algorithms and their variation with differently selected parameters are stated in literature by means of classification. Majority of them are highly modified for biomedical datasets. Accurate results provide more informative and meaningful diagnosis. In that meaning, three fundamental methods of machine learning are adapted for mesothelioma dataset.

a) Support Vector Machine (SVM)

SVM is one of the prominent classification algorithms that can be used large-scale datasets and provides more accurate results. It can be achieved by even small size trainsets with the help of well-fitted cost function in kernel space [15].

SVM uses the core idea of kernel based learning. It aims to separate data in high dimensional feature space with a kernel function. SVM creates a decision surface between the samples of different classes over optimal hyperplane.

SVM provides binary classification of two-class datasets. "One against one" or "one against all" are the most popular strategies in literature. Each strategy has own advantages and disadvantages mentioned in [16]. In our study, “one against one” strategy is used owing to 2 classes’ presence in datasets.

In order to define well-fitted settings of SVM for mesothelioma dataset, different kernels, penalty and kernel parameters are tested at the initial part of study. Table 2 indicates the all parameter test results.

b) Decision Tree (DT)

Decision Tree is known as rule based machine-learning method [17]. Principally, it works based on tree terminology. The path from root to leaf presents classification rules. The roots represent the most informative features and the leaves indicate the labels. Information gain (IG) is the rule defining criteria. The most widely used algorithms are entropy, twoing, and Gini to calculate the IG.

Decision Tree is easy to implement. Additionally, interpretation of the classification is much easier than other methods. It is useful for some regression problems. However, DT results low performance on large scale datasets with few training samples compare to SVM [18]. Pruning process is another obstacle point to avoid over-fitting. According the results of preliminary studies on parameter settings, DT model is modified with pruning functionality and Gini Diversity Index for IG.

c) Neural Networks - Multi Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is the advanced version of NN [19]. Minimum two layers connected with two functions should be utilized. Different parameters and functions are tested at initial studies. According to results, MLP network is arranged as the weight and bias are fixed with 0.8 and 1, respectively.

B. Ensemble Learning Methods

Ensemble learning is emerged from the principles of machine-learning concepts. The key point behind the ensemble is the proper combination of several machine-learning algorithms. Not only one learner as in regular methods, multi learners gather in decision step for ensemble methods, therefore it gives more success. Machine-learning classifiers such as DT, KNN etc. is named as base learner. Mainly two ensemble models having the same base learner (Decision Tree - DT) combinations but different sample

selection strategies are evaluated in this study. Majority voting is used to define final decision of base learners.

a) Bagging with DTs

Bagging, in other words bootstrap aggregation, is a way for improving the classification by the aid of well-formed train samples. It is also cited as re-sampling process in literature [20]. The idea of bagging is to distort the dataset by re-sampling, and to train weak learners using re-sampled trainsets. The distortion of the samples is made by a voting process of weight parameters. The weights of the samples are fixed equally; therefore, trainsets are randomly selected. Consequently, different samples are used in trainset iteratively. It provides more diversity in the samples' distribution. The average of each decision of base learners determines the final decision. More information can be found in [20].

b) Adaptive Boosting (Adaboost) with DTs

Boosting is another technique in re-sampling process similar to bootstrap. The difference is that bootstrap ignores the weight values of the samples and re-samples randomly, however boosting technique defines different weights for each samples after first iteration. Then, the probabilities of misclassified samples are boosted for the second step, and subsequent classifiers are trained. Likewise, other steps are sustained with different weight parameters. Readers are referred to an essential guide [21] for boosting theorem.

Adaptive boosting is mainly outperforming other regular boosting techniques and more robust for over-fitting problem. However, it is still easily affected by noise and outliers owing to iteratively arranging process for weights.

C. Dataset

Dataset is obtained from UCI dataset repository [22]. It includes the patient's records obtained from Dicle University, Faculty of Medicine. 324 MM patient data were recorded and tested by aforementioned AI methods. These data were also investigated by Orhan Er et al. in terms of PNN as mentioned in Section 2 [12].

In the dataset, 324 samples individually have 34 features with multivariate variables. There is no "unidentified" or "missing value" presence in dataset. Details of data and features can be found in [12]. Decision labels provided by doctors as sick and healthy (2 classes).

IV. Test Results

Classification of mesothelioma dataset is performed by three regular machine learning and two ensemble learning methods. DT, SVM and NN methods are selected within the regular machine learning concept. On the other hand, Bagging and Adaboost with same weak learners (DT) is performed as ensemble idea. Accuracy and computational time are considered as the evolution metrics. Computational time is recorded to estimate efficiency of method for big data problems due to so many patients suffering from MM disease. In case of future studies with more patient record, time complexity become more important factor according to including 34 features besides plenty of patients.

Only 10 Fold Cross validation tests are measured in terms of computational time. Less computational time and high accuracy rate are preferred to indicate the best algorithm. Overall results are presented in Table 1.

TABLE 1. Overall Results of Methods

	DT	SVM (Linear)	MLP	Bagging	Adaboost
10-Fold	100	100	96,87	100	70,54
5-Fold	100	100	95,82	100	65,35
2-Fold	100	100	94,44	100	68,82
Time	0,019	0,095	13,89	17,52	0,25

According to Table 1, simple DT and SVM as regular machine learning idea and DT with Bagging in terms of ensemble method outperform over other methods with common 100% accuracy rates. Differently formed train sets (2, 5 and 10 Fold) has no effect in general. However, another ensemble method, Adaboost using same form of DT as base learner but different sample selection strategy as weighed re-sampling, stay far behind over all methods. In this sense, randomly selection of train samples is more effective strategy in the detection of mesothelioma. Selection of sample with weight parameter is useless due to lots of features (34 features) using in classification. However, Bagging needs more computational time because of irregular sample selection process. In that meaning, Bagging is not preferred method when compare to DT and SVM because of the same accuracy rates.

One of the prominent Kernel based method, SVM, is tested with different kernels and parameters. Obtained the highest results of each kernel with different parameters are individually registered in Table 2. Linear kernel gives the best result with 100% in all *K* values. RBF (radial basis function) outcome is depended on training size. It resulted 100% accuracy rate with more training samples, but success is decreased when train set reduced. Besides the inconsistent results of RBF, it includes exponentially operations, thus, needs more time to classify big data. To avoid that time consuming process, Linear SVM might be utilized in practice owing to simplicity of algorithm and less time complexity. Polynomial, quadratic and MLP (Multi-Layer Perceptron) kernels generally concluded with 97%, 88% and 90% respectively. These kernels are also directly related with training sample size. Addition to low accuracies, computational time analyses of kernels are not too far ahead from linear kernel. Therefore, SVM should be utilized with Linear Kernel to classify mesothelioma dataset. Results emphasize that it might give better results with big data over other methods.

TABLE 2. SVM RESULTS ACCORDING TO USING KERNEL

Kernels	Polynomial	Quadratic	MLP	Linear	RBF
10-Fold	97,72	88,98	90,93	100	100
5-Fold	97,18	88,75	89,21	100	99,84
2-Fold	92,40	84,01	86,11	100	99,07
Time	0,186	0,385	0,089	0,095	0,286

As a final method, MLP in Neural Network terminology is adapted. Normally, MLP gives higher accuracies on non-linear classification problems, but deals with all samples in dataset. In that meaning, algorithm success might be decreased easily by outliers and needs more computational time as it is emphasized in Table 1. Dataset has 34 features over 324 observations which means 34-dimension data. In that case, MLP is resulted with 97% accuracy rate owing to complexity of data set. On the other hand, SVM focus on the samples near support vectors. Therefore, SVM surpasses MLP due to less complexity and using pre-arranged data.

V. CONCLUSION

In this study, different machine and ensemble learning methods are tested on the detection of mesothelioma disease. In that meaning, a prevalent dataset provided by Orhan Er et al. [8] is utilized to measure the methods.

Orhan Er et al. published a study about the classification of their dataset with PNN before. They reported 96% success with 3-Fold cross validation. In this study, we also perform a MLP network having 0.8 weight and 1 bias parameters and obtained same results. This indicates the testing methodology is similar and analogous. In that meaning, other obtained results express consistent output.

DT and SVM as regular machine learning, and Bagging as ensemble learning are highly compatible algorithms for mesothelioma dataset considering to Table 1. Methods successfully provide 100% accuracy rate in classification. However, linear kernel SVM and DT are simpler algorithm and require less computational time compare to Bagging. In this sense, Bagging is not preferably. Rule Based algorithm, DT, fails on big data problem according to report [14]. Therefore, it is also useless in practice owing to numerous patient suffering from Mesothelioma disease. In order to generalize the results, more record is necessary. In that condition, DT might give misleading diagnose. As a result, Linear SVM might be better to utilize in practice due to abovementioned results and reasons.

As future works, abovementioned methods will be tested on more obtained data in classification. Then, more generic diagnose system can be improved.

REFERENCES

[1] Malignant Mesothelioma, Retrieved 3 May 2016, Available: http://www.cancer.gov/types/mesothelioma

[2] General Information About Malignant Mesothelioma, Retrieved 3 May 2016, Available: http://www.cancer.gov/types /mesothelioma/patient/mesothelioma-treatment-pdq

[3] B.M. Robinson, "Malignant pleural mesothelioma: an epidemiological perspective", Annals of cardiothoracic surgery vol. 1 (4), 2012.

[4] S. Kondola, D. Manners, A.K. Nowak, "Malignant pleural mesothelioma: an update on diagnosis and treatment options", Therapeutic Advances in Respiratory Disease, 2016.

[5] Delp, S. L., Loan, J. P., Robinson, C. B., Wong, A. Y., & Stulberg, S. D. (1997). U.S. Patent No. 5,682,886. Washington, DC: U.S. Patent and Trademark Office.

[6] H. Kadoz, S. Ozsen, A. Arslan, and S. Gunes, "Medical application of information gain based artificial immune recognition system (AIRS): diagnosis of thyroid disease", Expert Syst Appl vol. 36 (2), 2008.

[7] J. Lerdsinmongkol, K. Chaisaowong, S. Roongruangsorakarn, T. Kraus, and T. Aach, "Efficient Application of 3D Morphological Operations in the Framework of a Computer-Assisted Diagnosis System". 9th International Conference on Signal Processing, pp. 857-860, 2008.

[8] M. Chen, E. Helm, N. Joshi, S.M. Brady, "Random walk-based automated segmentation for the prognosis of malignant pleural mesothelioma", IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1978-1981, 2011.

[9] Grady, L. (2006). Random walks for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 28(11), 1768-1783.

[10] Onoma, D. P., Ruan, S., Gardin, I., Monnehan, G. A., Modzelewski, R., & Vera, P. (2012, May). 3D random walk based segmentation for lung tumor delineation in PET imaging. In 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 1260-1263). IEEE.

[11] O. Er, A.C. Tanrikulu, A. Abakay, and F. Temurtas, "An approach based on probabilistic neural network for diagnosis of Mesothelioma's disease", Computers & Electrical Engineering, vol. 38(1), pp. 75-81, 2012.

[12] K. Chaisaowong, C. Akkawutvanich, C. Wilkmann, and T. Kraus, "A fully automatic probabilistic 3D approach for the detection and assessment of pleural thickenings from CT data", Computational Intelligence in Medical Imaging (CIMI), IEEE Fourth International Workshop on, pp. 14-21, 2013.

[13] Schroder, M., Rehrauer, H., Seidel, K., & Datcu, M. (1998). Spatial information retrieval from remote-sensing images. II. Gibbs-Markov random fields. IEEE Transactions on geoscience and remote sensing, 36(5), 1446-1455.

[14] E. Lotfi, A. Keshavarz, "Gene expression microarray classification using PCA–BEL", Computers in Biology and Medicine, vol. 54, pp. 180–187, 2014.

[15] B. Scholkopf, A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT press, 2001.

[16] J. Milgram, M. Cheriet, R. Sabourin, one against one or one against all: Which one is better for handwriting recognition with svms in: Tenth International Workshop on Frontiers in Handwriting Recognition, Suvisoft.

[17] Quinlan, J. Ross. "Induction of decision trees." Machine learning 1.1 (1986): 81-106.

[18] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and Regression Trees, Wadsworth International Group, Belmont, CA, 1984.

[19] Hagan, Martin T., Howard B. Demuth, and Mark H. Beale. Neural network design. Boston: Pws Pub., 1996.

[20] L. Breiman, Random forests, Machine learning 45 (2001) 5– 32.

[21] R. Rojas, Adaboost and the super bowl of classifiers a tutorial introduction to adaptive boosting (2009).

[22] UCI Machine Learning Repository, Mesothelioma Disease Data Set, Retrieved 3 May 2016, Available: http://archive.ics.uci.edu/ml/machine-learning-databases/00351/

Experience of using of machine learning methods to identify the left ventricle region in echocardiographic records

A.O. Bobkova, V,V, Zyuzin, S.V. Porshnev
Institute of Radioelectronics and Information Technologies
Ural Federal University
Yekaterinburg, Russia
iconismo@gmail.com

V.V. Bobkov
Institute of Information Technologies
Ural State University of Economics
Yekaterinburg, Russia
btow@yandex.ru

***Abstract*—The possibility of an ultrasound study of the heart (Echocardiography, EchoCG) is widely used in modern cardiology. This non-invasive technology allows studying cardiac activity of the patient by determining the global contractility of the heart muscle. The methods, which used in echocardiography, require performing manual operations from specially trained medical professionals. A number of researchers are working on the problem of automation of this medical technology. The article shows the way of solving the problem of the left ventricle area identification in echocardiography records with machine learning techniques. The task of the left ventricle delineation is reduced to the problem of pixels classification on video frames. A pixel can belong to one of two classes (the background region or the region of the left ventricle). The possibility of solving the task was tested with the following classifiers: decision tree, gaussian naive Bayes, linear discriminant analysis, quadratic discriminant analysis, adaboost classifier, random forest classifier. The assessment of classification results was performed using ROC curves. The best performance was obtained for both algorithms: decision tree classifier (AUC 0.93) and random forest classifier (AUC 0.93).**

***Index Terms*—Echocardiography, left ventricle, edge detection, ultrasound images, apical four-chamber view.**

I. Introduction

Ultrasonography is the non-invasive method widely used in various fields of medicine. In cardiology it called echocardiography [1] and is applied in the study of global and local contractility of the left ventricle (LV) of the heart.

The apical four-chamber view is a widespread projection of the heart (Fig. 1). In this projection, the expert indicates the contour of the LV. This usually occurs manually. This procedure is very time consuming and requires specialists of a certain experience and qualifications. Experts find it difficult to explain clearly and objectively why they build the contour of the LV in a certain way. In some cases, they rely on intuition. Therefore, it is difficult to directly transfer their expertise in contouring algorithms.

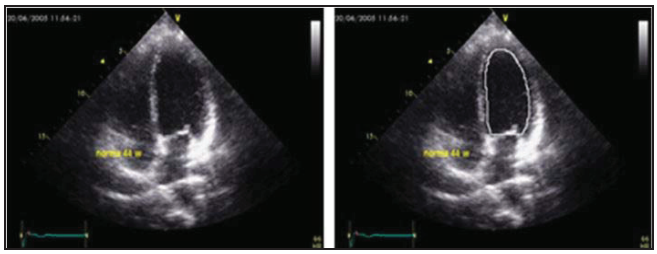


Fig. 1. The ultrasound image and the contour marked by expert

In this regard, numerous attempts to develop automatic algorithms for LV borders identification in echocardiographic images have been made [2, 11, 12]. Today there are many different ultrasound scanners, each of which is equipped with one or another tool for LV contouring (Philips, Aloka, Hitachi, Toshiba, Siemens, General Electric and others). Analysis of the commercial proposals of this companies showed that there is no echocardiography devices with built-in programs for LV contouring in fully automatic mode. In this context, the development of automated algorithms to delineate LV on echocardiographic images is important.

II. Problem statement

From the Fig. 1 we can see that the image on the echo frame is a set of pixels having different brightness in the grayscale. Analysis of the medical experts experience have shown that the cardiologists use both a set of low-level features of each single frame and dynamics of their change in constructing the LV contour. In addition, they are guided by their own ideas about the right shape of LV. In the result, the expert can build the LV contour despite the presence of noise and artifacts on the border of the LV for each ultrasound frame. Even in the absence of clearly visible boundaries in some parts of the frame, the expert is able to identify the contour of the LV.

The task of LV contouring can be formulated as the problem of pixel classification. It is clear that in this case the number of classes is equal to two — pixel can belong to either

region of the LV or the background region. Consequently, the task of the LV region identification is to classify pixels to the region of the LV or the rest of the region ultrasound frame depending on their coordinates in feature space. It means to be a task of binary pixel classification. Thus, the problem of the LV area identification on the image is formulated as following. Let X be a set of pixel description, Y — a set of pixel classes. There is an unknown target dependency:

$$y^* = X \to Y, \quad (1)$$

which values are known only on the objects of the finite training sample:

$$X^m = \{(x_1, y_1), ..., (x_m, y_m)\}. \quad (2)$$

It requires to develop the algorithm:

$$a : \mathrm{X} \to \mathrm{Y}, \quad (3)$$

which will be able to classify a random object:

$$x \in X. \quad (4)$$

III. Description of the Used Machine Learning Methods

Traditionally, machine learning methods are used for solving the task of binary classification. Brief description of the applied methods is presented below.

A. Decision tree classifier

Decision trees [3-5] are a non-parametric supervised learning method used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. We uses an optimized version of the CART algorithm.

B. Gaussian naive Bayes

Naive Bayes methods [6] are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of independence between every pair of features. In spite of their apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many real-world situations, famously document classification and spam filtering. They require a small amount of training data to estimate the necessary parameters. Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods. We used Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian.

C. Linear discriminant analysis and quadratic discriminant analysis

A classifier with a linear decision boundary is generated by fitting class conditional densities to the data and using Bayes' rule. The model fits a Gaussian density to each class, assuming that all classes share the same covariance matrix [7]. A classifier with a quadratic decision boundary is generated by fitting class conditional densities to the data and using Bayes' rule. The model fits a Gaussian density to each class [7]. These classifiers are attractive because they have closed form solutions that can be easily computed, are inherently multiclass, have proven to work well in practice and have no hyperparameters to tune.

D. AdaBoost classifier

An AdaBoost [8] classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases. We used implementation of the algorithm known as AdaBoostSAMME [9].

E. Random forest classifier

In random forests [10], each tree in the ensemble is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is nolonger the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. As a result of this randomness, the bias of the forest usually slightly increases (with respect to the bias of a single nonrandom tree) but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model. In contrast to the original publication [10] we used the scikit-learn implementation, that combines classifiers by averaging their probabilistic prediction, instead of letting each classifier vote for a single class.

IV. Experimental Details

A. Dataset description

A collection of echo records, which included a grayscale EchoCG images of the heart (Fig. 2, series A) and corresponding binary images of the LV, marked by the expert (Fig. 2, series B), were used for training and testing.

The collection was received as a result of echocardiography studies for adult patients without pathology on the ultrasound machine Philips. Each video sequence contains one cardiac cycle. Expert segmentation was conducted manually by doctors. The collection contained ultrasound video of 26 patients. The number of frames used in the videos were different (minimum - 20, maximum - 50). The total number of frames was 724. Reduction of video frames was carried out to reduce the required computing resources. The size of each image is reduced to 60×80 pixels using bicubic interpolation. Reducing frame size also reduces the influence of noise on the results of the delineation. This is due to the fact that the small bright region inside the LV also decrease or disappear completely when you decrease the frame.

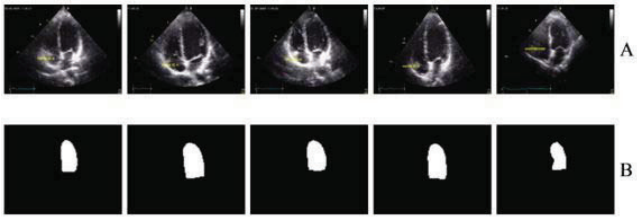


Fig. 2. Source data: A - ultrasound images, B – LV regions identified by experts

B. Classification features

The analysis of the results of the contouring performed by experts showed that the expert uses a set of low-level features of each single frame and their dynamic changes. Therefore, we have calculated multiple low-level features to solve the problem of classification of pixels. The features that characterize image-pixels is presented in table 1 . Features from 4 to 42 were converted into vector from matrix representation.

Thus, 42 vectors were received after the calculation all features on a frame:

$$f_i^1, f_i^2, \ldots, f_i^n, \quad (5)$$

where $n = 42$ – the number of features, $i = \overline{1, p}$ – number of pixel, p – the total number of pixels.

TABLE I. FEATURES OF PIXELS

№	Description
1	The row number of the pixel
2	The column number of the pixel
3	The ratio of the number of the frame to the total number of frames
4	The absolute difference between the first and current frame of the video sequence
5	Summation of the first and current frame of the video sequence
6	Division the first and current frame of the video sequence
7	Multiplication of the first and the current frame sequence
8	Matrix of corner metric (with adjusted contrast)
9-10	Horizontal and vertical lines (applied dilation operation and filling holes)
11-12	Operation of extended maximum (extended-maxima transform) values, H=150 and H=200 (4 connectivity)
13	Histogram equalization (contrast-limited adaptive histogram equalization)
14	Decorrelation stretching
15-18	Contrast enhancement using histogram equalization with 20 intervals, 10, 5 and 3
19	Smoothing of intensity values
20	Two-dimensional median filter
21	Filtering based on statistics of second order (2-D order-statistic filtering)
22	Smoothing of intensity values
23	Adaptive Wiener filtering
24	Filter approximating the linear motion of the camera
25	Rounded averaging filter
26	Averaging filter
27	Gaussian low pass filter
28	Approximating the two-dimensional Laplace operator
29	Laplacian of Gaussian filter
30	Improving contrast filter
31	Two-dimensional cosine transform
32	Top-hat filter
33	Bottom-hat filter
34	Addition the current matrix of the image and the result of this matrix after top-hat filtering, subtraction of the bottom-hat filtering, applied median filtering.
35	Suppress light structures that are adjacent to the edges of the image and apply median filtering
36-39	Morphological closing operations, dilation, erosion, opening
40-42	The calculation of local entropy, amplitude and standard deviation

Accordingly, each vector of features f_i^n contained p elements. Then all the features were combined into a single matrix of features:

$$X^{\langle 1 \rangle} = f_i^1, X^{\langle 2 \rangle} = f_i^2, \ldots, X^{\langle n \rangle} = f_i^n. \quad (6)$$

After that the vector Y_i containing class labels (0 or 1) for each i pixel was formed. In this case, the pixels belonging to the background area had 0, the pixels belonging to the LV region had 1.

Thus, the total number of pixels, which were used for classification, was equal to 912 964.

C. Validation scheme

684 723 (75%) pixels were used as training set. Evaluation of classification quality was conducted on the remaining 228 241 (25%) pixels. AUC criteria was used for assessment the quality of classification pixels. AUC is the area enclosed by the ROC curve and axis of the proportion of false positive classifications. Note that the higher value of AUC, the better classifier. A value of AUC = 0.5 indicates the unsuitability of the chosen method of classification (corresponds to random divination).

D. Testing machine learning methods

We checked the applicability of the following methods:

- Decision Tree Classifier (DTC);
- Gaussian Naive Bayes (GNB);
- Linear Discriminant Analysis (LDA);
- Quadratic Discriminant Analysis (QDA);
- AdaBoost Classifier (ABC);
- Random Forest Classifier (RFC).

Setting of methods are shown in table 2.

TABLE II. SETTINGS

№	Designation	Setting
1	DTC	The Gini criteria is used for the split quality measurement. The minimum number of samples required to split an internal node is equal to 2. The minimum number of samples required to be at a leaf node is equal 1.
2	GNB	The default values from the sklearn library [13] were used.
3	LDA	
4	QDA	
5	ABC	Boosted ensemble is built from decision tree classifier. The maximum number of is equal 50.
6	RFC	The number of trees in the forest is equal to 20. The Gini criteria is used for the split quality measurement. The minimum number of samples required to split an internal node is equal to 2.

V. RESULT AND DISCUSSION

Validation result are presented in table 3. The graphical representation of the validation results is presented in the Fig. 3. The table 3 shows that the assumption about the possibility of the successful application of machine learning

methods to solve the contouring problem is confirmed. All tested algorithms showed the AUC more than 0.5.

TABLE III. QUALITY ASSESSMENT OF PIXELS CLASSIFICATIONS

Number	Designation	AUC
1	DTC	0.93
2	GNB	0.73
3	LDA	0.62
4	QDA	0.86
5	ABC	0.82
6	RFC	0.93

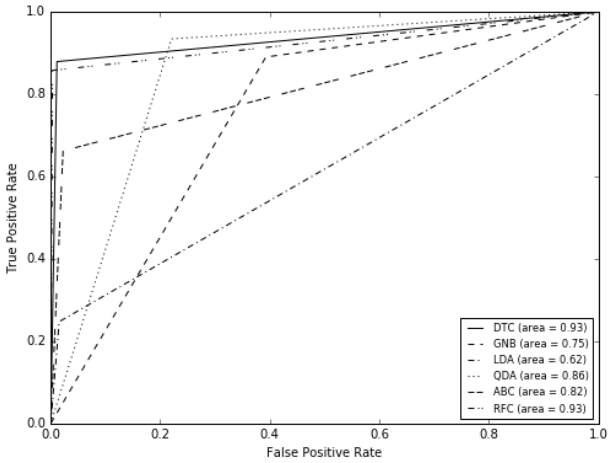


Fig. 3. Validation results

This indicates that machine learning methods can detect poorly formalized patterns when constructing the LV contours. The greatest attention is attracted by decision tree classifier and random forest classifier. These methods demonstrate the high accuracy of pixel classification.

However, the required accuracy level of the algorithm for automatic LV contouring has not been determined. Therefore, it is necessary to conduct additional research to identify the expert range, which includes the contours of the LV constructed by different experts in the ultrasound record frames. It is also possible to continue the search for methods that will demonstrate higher classification accuracy pixels.

VI. CONCLUSION

The results of applying decision tree classifier and random forest classifier in the task of automatic LV contouring in echocardiography records confirm their applicability. They have the same area under the ROC curve (it is equal to 0.93). Therefore, the study results show that this methods may be used for solving the problem of automatic LV contouring on the frame of the ultrasound films with a high level of confidence.

REFERENCES

[1] Flower, M. A., [Webb's physics of medical imaging], CRC Press (2012).

[2] Porshnev, S., Bobkova, A., Zyuzin, V., and Bobkov, V., \Method of semi-automatic delineation of the left ventricle of the human heart on echographic images," Fundamental research (8), 44-48 (2013).

[3] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

[4] J.R. Quinlan. C4. 5: programs for machine learning. Morgan Kaufmann, 1993.

[5] T. Hastie, R. Tibshirani and J. Friedman. Elements of Statistical Learning, Springer, 2009.

[6] H. Zhang. The optimality of Naive Bayes. Proc. FLAIRS, 2004.

[7] Hastie T., Tibshirani R., Friedman J., The Elements of Statistical Learning", Section 4.3, p.106-119, 2008.

[8] Y. Freund, R. Schapire, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", 1995.

[9] Zhu, H. Zou, S. Rosset, T. Hastie, "Multi-class AdaBoost", 2009.

[10] Breiman, L., "Random forests," Machine Learning (45), 5-32 (2001).

[11] Hamarneh G., Gustavsson T. Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images //Computers in Cardiology 2000. – IEEE, 2000. – C. 115-118.

[12] Lacerda S. G. et al. Left ventricle segmentation in echocardiography using a radial-search-based image processing algorithm //Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. – IEEE, 2008. – C. 222-225.

[13] Scikit-learn library (Machine Learning in Python). http://scikit-learn.org/

Classification Of Breast Cancer By Using Pattern Recognition Tools

Laven Wahhab
Department of Computer Engineering, Fatih University, Istanbul, TURKEY
lavinwahab@gmail.com

Kadir Tufan
Department of Computer Engineering, Fatih University, Istanbul, TURKEY
ktufan33@gmail.com

Shahriar Shamiluulu
Department of Computer Sciences Mathematics and Natural Sciences, Kaskelen, KAZAKHSTAN
shahriar.shamiluulu@sdu.edu.kz

Abstract— **Breast cancer is one of the most deadly type's cancers worldwide. Diagnosing breast cancer always is invasive, costly, time-consuming and often risky. That is why many patients reach late stages of breast cancer without treatment, especially in those countries where resources are limited. Therefore, the early diagnosis remains vital, early detection strategy, particularly in developing countries, where the disease is diagnosed in late stages. Finding a solution for above-mentioned problems and riding out from disadvantages became a strong motive to conduct this study.**

To conduct this study, in this paper, we propose automatic classification algorithm for breast cancer diagnosing basis on clinical history, physical examinations, and laboratory tests, which are noninvasive, cheap and save. The feature vector is created by using above mentioned information. The effects of using features to classify breast cancer with pattern recognition techniques are investigated.

Index Terms—**Breast Cancer, support vector machine, Diagnosis, Biomedical.**

I. INTRODUCTION

Cancer is a disease in which cells grow out of control. The growth of these cells is called tumors which classified to (non-cancerous (benign), pre-cancerous and cancerous (malignant)). Malignant type differentiates from other types of the tumor by its capacity to invade nearby tissues and they can spread to other organs by blood vessels and lymphatic channels [1].

Breast Cancer considered as a malignant type of tumor that starts in breast cells, it is one of the most common leading death factor in developed countries. It usually affects the women and it's rare among men (1%) [2]. Diagnosing breast cancer usually starts with clinical data, lab tests, imaging studies (mammogram, sonogram, and MRI of the breast) and finally biopsy (removing a piece of a breast mass and searching for malignant cells by microscopy). Although biopsy is the standard diagnosing test, it has many disadvantages, such as being invasive, costly, time-consuming and sometimes risky. For example; when a biopsy is performed, the patient may face infection, the scare of surgery and misdiagnosis. Imaging studies (mammogram, sonogram, and MRI of the breast) has been used for many years to detect breast cancer. But using them has some limitation; more expressly is exposure effects of radiation. Besides being risky, the data provided by imaging is insufficient to diagnose breast cancer [3].

The automated diagnosis of breast cancer has attracted many researchers. There are several research studies has been conducted so far [4, 5]. Comparison of the methods and accuracy of previous studies and this study are summarized herein. The one study showed the performance of the model which is evaluated by using nine attributes which represent nine cytological characteristics of breast fine-needle aspirates (FNAs). The medical data for this study was taken from Wisconsin breast cancer database (WBCD) [6]. In another study, the researchers developed a system for diagnosis, prognosis and prediction of breast cancer by using Artificial Neural Network (ANN) models like MLP, RBF, LVQ, and Competitive learning network. These models are useful for performing the classification of breast cancer diagnosis by using 10 attributes. For the study they also obtained data from Wisconsin Breast Cancer Data (WBCD) [7]. In one more research study, the researchers proposed an approach for distinguishing between different classes of breast cancer. During the study the Differential evolution algorithm (DE) has been used, in order to determine the optimal value or near optimal value for ANN parameters [8]. In another study researchers used SVMs [9, 10] as a classification model for detecting and diagnosing malignant and benign based on MRI features[11], ultrasound feature [12, 13] and mammographic features[13].

In this study, by using the pattern recognition techniques we have used cheap, simple and noninvasive tests that can be performed easily. Data were taken directly from patients therefore, we could interview the 206 subjects. They are divided into two groups 100 of them have features of breast cancer, the rest denied to have any feature of breast cancer, and every patient is expressed by 19 features which are divided into three parts clinical history, physical examination and lab tests.

By this strategy, we hope to produce "downstaging" (increasing in the proportion of breast cancers detected at an early stage) of the disease to stages that are more amenable to curative treatment.

In order to improve the accuracy of breast cancer classification as benign and malignant, the performance of Support Vector Machine (SVM) was evaluated. The SVM is a flexible classifier algorithm that has been proposed as an effective statistical learning method for pattern recognition [14], which is based on finding optimal hyperplane to separate different classes mapping input data into higher-dimensional feature space. The SVM has been used for many applications such as object recognition and face detection [15].

II. Materials And Methods

A. Data Collection

The right diagnosis of breast cancer is one of the major issues in the medical field. The participants in this research have been classified into two groups. The first group included patients who had all features of the breast cancer and second group is opposite to the first one. In total 206 cases were collected. A 106 of the cases were complaining from bad effects of breast cancer and the rest were not. The original dataset contains 19 features in the dataset that are used in this application. The attributes represent our data resources as history, examination, laboratory tests. According to the properties of the attributes, the breast cancer was classified into benign (expressed by "Positive") and malignant (expressed by "Negative"). All data has been taken from Sheen International Hospital (Oncology Department) which is one of the major hospitals in Erbil, Iraq.

B. Support Vector Machine (SVM).

The SVM is a supervised learning algorithm that is used for data classification and regression [14]. It searches for a best hyperplane which separate between classes. The best hyperplane is considered the one that leaves the maximum margin between the two classes. The margin is defined as the width of the hyperplane from the closest point of the two classes. Bounds between data sets and hyperplane are called support vectors [15]. The hyperplane can be found by:

$$g(x) = w^T x + b \quad (1)$$

- " x " : refers data points
- " w ": is a coefficient vector
- " b " : is offset from the origin

In the case of linear SVM $g(x) \geq 0$ for the closest point on the one of the class, $g(x) < 0$ for the closest point belongs to another class. The margin between support vectors is defined by:

$$d = \frac{2}{||w||} \quad (2)$$

For better separation, the margin d should be maximized. The data points which are called support vectors are the only ones which are relevant for the solution of the problem. All remaining data points could be deleted from the data set and the same solution can be obtained.

C. Performance Measures

Evaluation of the classifier to measure the quality is commonly evaluated using the data in the confusion matrix. Several standard measures have been defined for correct and incorrect classification results of the matrix. The most common practical measure to evaluate the performance is accuracy, which is defined as the proportion of the total number of instances that were classified correctly.

Sensitivity is the mean proportion of actual positives which are correctly identified. Specificity is the mean proportion of negatives which are correctly identified.

These performance metrics are calculated according to the data in the confusion matrix which are mentioned in [16].

D. Simulated Program WEKA

The Waikato Environment for Knowledge Analysis (WEKA), which is a Java-based machine learning tool that contains a large number of algorithms for data preprocessing, feature selection, classification, clustering, and finding the associative rules [17]. WEKA uses a common data representation format, making comparisons easy. It has three operation modes i.e., GUI, Command Line, and Java API.

III. Result And Discussion

In this study, the 19 original features of the breast cancer data are used for classification. The accuracy, sensitivity, specificity of 19 features has been performed using 10-fold cross-validation. To construct the SVM classifier proper kernel function and its parameters has been chosen. Generally Sequential minimal optimization (SMO) is used for training SVMs, which is implemented by the popular application to find proper hyperplane for SVMs [18]. Such implementation replaces all missing values and transforms nominal features into binary ones.

In the present study, the accuracy evaluation of SVM has been computed for linear, where the complexity parameter is set to 1.0.

SVM have been tested and trained to find out maximum accuracy adjusting their parameter. The performance measures such as accuracy, sensitivity, specificity of the classifiers are compared to each other. The parameters of the classifiers which provide maximum accuracy are selected to be compared to the other classifiers.

10-fold cross validation with 19 features used to compare the performances of the classifier. In input data of the classifier, and the test data are compared to the original class label to find out TP, TN, FP, and FN values. These values for classifiers are provided in the form of confusion matrix in Table 1.

Sensitivity refers successfully identified malignant samples in cancer classification. Thus, higher sensitivity means the higher diagnostic capability of malignant tumors and it can be used to help physicians to diagnose cancerous mass more correctly. The accuracy, sensitivity and specificity measures in this study are given in Table 2 to compare the effect of the feature using pattern recognition tools. It should also be noted that, medical dataset used in this research was obtained from the personal resources; where in the other studies provided in literature review section obtained from different sources like UCI, WBC and Indian Pima.

TABLE 1. THE CONFUSION MATRICES OF THE CLASSIFIER

Actual Value	Expected Value	
	Positive	**Negative**
Positive	TP = 99	FN = 1
Negative	FP = 6	TN = 100

TABLE 2. SUPPORT VECTOR MACHINE WITH LINEAR KERNEL

Total Number of Folds	Accuracy (%)	Sensitivity (%)	Specificity (%)	Complexity Parameter (C)
2	96.602	96.600	96.700	1.0
3	95.631	95.600	95.700	1.0
4	96.602	96.600	96.700	1.0
5	96.602	96.600	96.700	1.0
6	96.117	96.100	96.200	1.0
7	96.117	96.100	96.200	1.0
8	95.631	95.600	95.700	1.0
9	96.117	96.100	96.200	1.0
10	96.602	96.600	96.700	1.0

IV. CONCLUSION

The diagnosis of the breast cancer is a cumbersome problem. In the war zone the diagnosis process by using lab tests, imaging studies and biopsy might be a time-consuming, invasive, and costly. In this study, we propose automatic classification algorithm for breast cancer diagnosing based on clinical history, physical examinations, and laboratory tests, which are noninvasive, cheap and save. The medical data was taken from personal resources. The performance measures of SVM classifier with linear kernels have been evaluated in order to find sensitivity, specificity, and accuracy metrics. Experimental results showed that properly implemented classifier can reach the overall performance of 96.602%. The sensitivity value of SVM classifier with a linear kernel is 96.600%. In this study, we have validated the obtained results with biopsy output, in order to be sure about the result. In the future more compact and autonomous tool can be developed that could be used as a screening and also diagnosing mechanism as compare to other mentioned aggressive methods.

ACKNOWLEDGMENT

The authors thanks, Dr. Abdulsalam Zrar for providing the medical data.

REFERENCES

[1] A. Mert, N. Kılıç, E. Bilgili, and A. Akan, "Breast cancer detection with reduced feature set," Computational and mathematical methods in medicine, vol. 2015, 2015.
[2] A. Tartar, N. Kilic, and A. Akan, "Classification of pulmonary nodules by using hybrid features," Computational and mathematical methods in medicine, vol. 2013, 2013.
[3] N. Brasic, D. J. Wisner, and B. N. Joe, "Breast MR imaging for extent of disease assessment in patients with newly diagnosed breast cancer," Magnetic resonance imaging clinics of North America, vol. 21, pp. 519-532, 2013.
[4] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast cancer detection using rank nearest neighbor classification rules," Pattern recognition, vol. 36, pp. 25-34, 2003.
[5] K. U. Rani, "Parallel approach for diagnosis of breast cancer using neural network technique," International Journal of Computer Applications, vol. 10, pp. 1-5, 2010.
[6] F. Paulin and A. Santhakumaran, "Classification of breast cancer by comparing back propagation training algorithms," International Journal on Computer Science and Engineering, vol. 3, pp. 327-332, 2011.
[7] R. Janghel, A. Shukla, R. Tiwari, and R. Kala, "Breast cancer diagnosis using artificial neural network models," in Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, 2010, pp. 89-94.
[8] H. T. T. Thein and K. M. M. Tun, "An Approach for Breast Cancer Diagnosis Classification using Neural Network," Advanced Computing, vol. 6, p. 1, 2015.
[9] M. M. R. Krishnan, S. Banerjee, C. Chakraborty, C. Chakraborty, and A. K. Ray, "Statistical analysis of mammographic features and its classification using support vector machine," Expert Systems with Applications, vol. 37, pp. 470-478, 2010.
[10] N. H. Sweilam, A. Tharwat, and N. A. Moniem, "Support vector machine for diagnosis cancer disease: A comparative study," Egyptian Informatics Journal, vol. 11, pp. 81-92, 2010.
[11] J. Levman, T. Leung, P. Causer, D. Plewes, and A. L. Martel, "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines," Medical Imaging, IEEE Transactions on, vol. 27, pp. 688-696, 2008.
[12] R. Liao, T. Wan, and Z. Qin, "Classification of benign and malignant breast tumors in ultrasound images based on multiple sonographic and textural features," in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on, 2011, pp. 71-74.
[13] A. M. Arymurthy, "A system for computer aided diagnosis of breast cancer based on mass analysis," in Robotics, Biomimetics, and Intelligent Computational Systems (ROBIONETICS), 2013 IEEE International Conference on, 2013, pp. 247-253.
[14] V. Vapnik, The nature of statistical learning theory: Springer Science & Business Media, 2013.
[15] E. Gumus, N. Kilic, A. Sertbas, and O. N. Ucan, "Evaluation of face recognition techniques using PCA, wavelets and SVM," Expert Systems with Applications, vol. 37, pp. 6404-6408, 2010.
[16] K. Tufan, "Noninvasive diagnosis of atherosclerosis by using empirical mode decom-position, singular spectral analysis, and support vector machines," Biomedical Research, vol. 24, pp. 303-313, 2013.
[17] M. Hall, E. Frank, and G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD explorations newsletter, vol. 11, pp. 10-18, 2009.
[18] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, p. 27, 2011.

Sentiment Analysis for Agglutinative Languages

Eshref Adalı
Faculty of Computer Eng. and Informatics
Istanbul Technical University
Istanbul, Turkey
adali@itu.edu.tr

Assoc.Prof. Abzetdin Z. Adamov
Applied Research Center for Data Analytics and Web Insights (CeDAWI), Qafqaz University
Baku, Azerbaijan
aadamov@qu.edu.az

***Abstract*—Other people judgment, sentiment, opinion and think about actions or product or speech are important for person or company or institutions for decision making processes. Today people declare their opinions, feeling and judgment about a product or a service or a film or a speech on social media platforms. Social media are easily accessible platforms and provide remarkable sources for analysis and evaluation. Sentiment analysis find out the feeling of people on an object. The feeling covers attitudes, emotion and opinions. Sentiment is subjective impression; not facts. Sentiment analysis is basically a text classifying problem. Therefore, use Natural Language Processing (NLP), statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text.**

***Index Terms*—Sentiment Analysis, NLP, Parity, Semantic Orientation, Machine Learning, Sentient Lexical.**

I. Introduction

Everybody are curious about other people judgment, sentiment, opinion and think about actions or product or speech. These kind of information are important during decision making processes. Companies want to know opinion of customer about their product. Politician wants to know the effect of his or her speech. Customer service should know the satisfaction of customer. Customer wants to know opinion of other people about the product and service before buy it.

In order to learn satisfaction of customers and performances of product, companies need a market researcher. These kind of research are very costly and time consuming; need comprehensive survey. Some firms collect customers evaluation forms and analyze them by computer or manually.

Today people declare their opinions, feeling and judgment about a product or a service or a film or a speech on social media platforms. Social media are easily accessible platforms and provide remarkable sources for analysis and evaluation. The data of social media can be captured and process by computer. Therefore, today many companies working on collection of data from social media and process them by using of sentiment analysis tools and techniques.

Sentiment analysis find out the feeling of people on an object. The feeling covers attitudes, emotion and opinions. Sentiment is subjective impression; not facts. Generally, a binary opposition in opinions is assumed; good or bad, like or dislike or for or against, etc. These are called as "polarity" or "semantic orientation". Sentiment analysis is the process of determining whether a text includes sentiment or not, and classifying the sentiment into positive, negative and neutral classes. Sentiment analysis is basically a text classifying problem. Therefore, use Natural Language Processing (NLP), statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text. Sentiment analysis sometimes referred to as opinion mining. In other word, sentiment analysis and opinion mining aims to extract the embedded polarity from textual data.

In order to extract the embedded polarity from a text, some features of the selected language (Turkish, English, and Korean) or language family (agglutinative or fusional) and some feature of domain (travel, hotel, movies) should be known. A word can get many suffixes in agglutinative language but one or two suffixes in fusional language. Therefore, to get the meaning of word in fusional language is easier than agglutinative language. Knowing the characteristics of the language are essential for NLP and sentiment analysis because different languages require different preprocessing techniques.

II. Level of Sentiment Analysis

The sentiment analysis may be done at different level which are explained as blow;

Word level: In this level, assigning a sentiment polarity to a word or Multi-words. The polarity of a word may be different in different domain or even in the same domain. (Big car, big house has positive polarity; big mobile phone has negative polarity)

Phrase level: One sentence may have two phrases and each phrases has different polarity. (I liked the movie, but she did not.) So there is no unique polarity of this sentence.

Sentence level: One sentences may consist of several polarities. In this case the overall polarity of a sentence is assigned to the sentence.

Aspect level: One sentence in the same domain may consists of several aspects. The polarity of each aspect may be different then polarity of sentence. (The actor was good, but I did not like the scenario.) In this example, the domain of two aspects are the same but the polarity of each aspect are different the overall polarity of sentence.

Document level: Document level sentiment analysis extract overall polarity of a document. It is complicated work to find out the polarity of a document which consists of many sentences.

Word level sentiment analysis is assumed to be fine and document level is assumed to be coarse analysis. Therefore, many research focused on word level or lexicon base sentiment analysis.

III. CLASSIFICATION

As mentioned earlier, sentiment analysis is basically a text classifying problem. Therefore, use NLP to extract, identify, or otherwise characterize the sentiment content of a text. There are many possibilities for what we might want to classify:

- Users
- Texts
- Sentences (paragraphs, phrases, chunks of text)
- Predetermined descriptive phrases (<ADJ N>, <N N>, <ADV ADJ>, etc)
- Words (POS tag)
- Multi-words (POS tag)

Before classification, one text has to be parsed into sentences; then sentences into phrase, if there are; then into multi-words and finally words.

IV. SENTIMENT LEXICON

The lexicon based sentiment analysis approach use a pre-determined Sentiment Lexicon (SL). While sentiment analysis is done, the algorithm uses the SL to find out sentiment of a word. Sentiment analysis does not require a detailed pre-processing [1] work before classification for fusional languages like English. But it is necessary and important for agglutinative language like Turkish in which it is possible to add many suffixes to stem. These derivational and inflectional suffixes can change the POS tag and sentimentally polarity of the word.

A sentiment analysis system must understand some linguistic markers such as negations, intensifications, and conditional constructions. This is necessary for better classifications. The linguistic markers are language-dependent. In order to extract these marker, language specific morphological analyzers and parsers required. Before given linguistic markers, we need to develop a sentiment lexicon with Semantic Orientation (SO). One sample are given in Table-1 and Table-2. [2]

TABLE I. EXAMPLES OF WORDS IN THE NOUN & VERB LEXICON FOR MOVIE

Word		SO Value
monstrosity	ucube	−5
hate (noun and verb)	iğrenç	−4
disgust	tiksinti	−3
sham	yapay	−3
fabricate	uyduruk	−2
delay (noun and verb)	aksak	−1
determination	Kararlılık	1
Inspire	ilham verici	1
inspiration	esinlenme	2
endear	çekici	3
relish (verb)	zevk vermek	4
masterpiece	Başyapıt	5

TABLE II. EXAMPLES FROM THE ADVERB DICTIONARY

Adverb		SO Value
excruciatingly	azap verici	−5
Inexcusably	affedilemez	−3
Foolishly	aptalca	−2
satisfactorily	yeterli	1
purposefully	amaçlı	2
hilariously	neşeli	4

Negation: Negation markers change the polarity of a predication or main verb. To negation is simply to reverse the polarity of the lexical item next to a negator, for example changing good (+3) into not good (−3). Some negators are: not, none, nobody, never, nothing, without.

- *Nobody gives a good performance in this movie. (nobody negates good)*
- *Out of every one of the fourteen tracks, none of them approach being weak and are all stellar. (none negates weak)*
- *Just a V-5 engine, nothing spectacular. (nothing negates spectacular)*
- *Beğendim diyen olmadı (ma negates beğenmek)*

Intensification: Intensification markers change the level of polarity like stronger or weaker. Some of intensification are given in in Table-3

TABLE III. INTENSIFIERS AND THEIR MODIFICATION FACTORS

Intensifier		Modifier %
Slightly	biraz	−50
Somewhat	biraz	−30
Pretty	oldukça	−10
Really	kesinlikle	15
Very	çok	25
Extraordinarily	Sıra dışı	50
(the) most	Ençok	100

Conditional: Conditional marker make an initial condition to polarity of sentence.

If it was not too long, I would have given it a five stars.

Rhetorical questions: The answer were clear but this answer wants to be accepted by another person.

Can one not like this movie?

Sarcastic: Sarcastic markers are usually adjective used to describe something that means the opposite tot what is literally implied.

it was a great movie!

Idiomatic: An expression whose meanings cannot be inferred from the meanings of the words that make it up. Some idiom composes of several words.

Ball is in your court (it is up to you to make the next decision or step)

Kafayı yemek (kafa: head; yemek: to eat; go crazy)
Göz boyamak (göz: eye, boyamak: to paint; sugar the pill)

V. Sentiment Analysis Processes

Several methods are used for sentiment analysis. Performance of some methods are sufficient but some are not. Some methods are time consuming some works fast. We will explain lexicon based sentiment analysis method. The steps of this method are given as follows:

- Cleaning
- Morphological Analysis
- Morphological ambiguity
- POS (Part-of-Speech)
- Multi-words expression
- Machine Learning
- n-Gram Modelling

Cleaning Processes: Text data may have orthographic and grammatical problems. Especially Tweets contain abbreviations, missing vocals letters and many spelling errors. Another problem is related with alphabet. Some people do not use original alphabet of the language. For example, they use English alphabet for Turkish message. Therefore, they need character correction, devocalization and ungrammatical constructs both due to the character limitation of Twitter and mobile devices with limited text entry capabilities.

Example:

Dusunuyorum *düşünüyorum*
Gnc *genç*
Slm *selam*
Okumucam *okumayacağım*

Morphological Analysis: Agglutinative language like Turkish has a very productive inflectional and observational morphology.

Morphological Disambiguation: Morphological analyzer proposes several ambiguous results; one of which will be the correct solution. In order to find out the correct solution a disambiguation tool must be used.

Multi-words expression: Multi-words have different meaning and will be interpreted together. [5]

Example:

Kafayı yemek *go crazy*
Ayağına kurşun sıkmak *shoot oneself in the foot*

The flow of algorithm is shown in Figure-1.

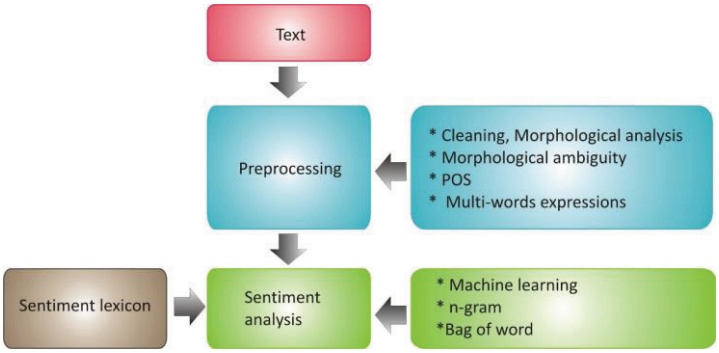


Fig. 1: The algorithm of lexicon based sentiment analysis

References

[1] Annett, M. ve Kondrak, G., *A comparison of sentiment analysis techniques: Polarizing movie blogs*, Advances in artificial intelligence, Springer, s.25–35. 2008

[2] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede; *Based Methods for Sentiment Analysis*, 2011 Association for Computational Linguistics.

[3] R. Dehkharghani, B. Yanikoglu, Y. Saygın, K. Oflazer, *Sentiment Analysis in Turkish: Towards a Complete Framework*, Natural Language Engineering 1, 2015

[4] U. Eroğlu, *Sentiment Analysis in Turkish*, Master thesis METU, 2009

[5] C. Türkmenoğlu, *Türkçe Metinlerde Duygu Analizi*, Master thesis İTÜ, 2015

[6] G.G. İşgüder, H. R. Zafer, E. Adalı, *Polarity Detection of Turkish Comments on Technology Companies,* ialp2004. Colips, Malasia