Big Data analitika əsasında informasiya təhlükəsizliyi obyektində anomaliyaların aşkarlanması modeli

Ramiz Alıquliyev[1], Məkrufə Hacırəhimova[2]

[1,2]AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

[1]r.aliguliyev@gmail.com, [2]makrufa@science.az

***Xülasə*— Anomaliyaların aşkarlanması verilənlərin analizində əsas məsələlərdəndir və şəbəkə təhdidlərinin aşkarlanmasında geniş istifadə olunur. Məqalədə Big data analitika əsasında şəbəkə trafikində anomaliyaların aşkarlanması üçün daha dəqiq və sadə multi-klassifikator modeli təklif olunmuşdur. Eksperimentlər NSL-KDD verilənlər dəsti üzrə WEKA proqram təminatında aparılmışdır. Anomaliyaların aşkarlanmasının dəqiqliyi baxımından təklif olunan model yaxşı nəticələr göstərmişdir.**

Açar sözlər— Anomaliya, Big data, Big data analitika, informasiya təhlükəsizliyi, klassifikatorlar ansamblı, IDS, NSL-KDD, WEKA.

I. GİRİŞ

Bəşəriyyətin mövcudluğundan təhlükəsizlik problemləri insanları daim düşündürmüş. Big data erasında isə təhlükəsizlik məsələlərinə maraq daha da artmış və siyasi, iqtisadi, sosial, demoqrafik, hərbi, ekoloji və s. kimi aspektlərdə çox ciddi elmi-tədqiqat istiqamətinə çevrilmişdir. [1]-də də göstərildiyi kimi Big data və informasiya təhlükəsizliyi məsələlərinə iki müxtəlif aspektdən baxmaq olar: informasiya təhlükəsizliyində Big data analitikanın tətbiqi və Big data texnologiyalarında informasiya təhlükəsizliyi problemləri.

Şəbəkə xidmətlərindən, veb əlavələrdən geniş istifadə müəssisə və təşkilatlarda şəbəkə və kompüter təhdidlərinə qarşı təhlükəsizlik tədbirlərinin həyata keçirilməsini ön plana çəkmişdir. Belə ki, son zamanlar kiber-terror, kiber-müharibələr və s. kimi kiber-məqsədli hücumların (*məsələn, kiber-terror, kiber-müharibələr, APT- Advanced Persistent Threat və s.*) sayı sürətlə artmaqdadır [2]. Korporativ kompüter şəbəkələrin təhlükəsizliyinə böyük təhdidlər yaradanbu hücumların aşkarlanması üçün şəbəkə, host, təhlükəsizlik qurğuları və s. mənbələrdən çox böyük həcmdə verilənlərin toplanması, verilənlərdə anomaliyaların aşkarlanması məsələsini aktuallaşdırır və məsələnin həllində daha effektiv analiz metod və alqoritmlərinin yaradılmasını tələb edir. Çünki kompüter şəbəkələrində təhlükəli trafiklərin erkən aşkarlanması, log faylların analizi şəbəkə təhlükəsizliyinin təmin edilməsində əsas şərtdir. Verilənlərin analizində anomaliyaların aşkarlanması isə əsas məsələlərdəndir. Anomaliyaların aşkarlanması qeyri-müntəzəm davranışlara müdaxilə etmək imkanı verir. Hücumların qarşısının qabaqcadan alınması (*0-cı gündən*), yəni əvvəllər rast gəlinməyən və bədniyyətli verilənlərin filtirlənməsi xüsusilə vacibdir.

Anomaliya verilənlərdə müəyyən olunmuş normal davranışa uyğun olmayan qanunauyğunluq kimi başa düşülür və ya verilənlərin göstəricisi kimi təyin olunur. Tədqiqatlarda anomaliya sözünə sinonim kimi “outliers”, “exceptions”, “peculiarities”, “surprise” terminləri də istifadə olunur [3]. Yəni anomaliyaların aşkarlanması verilənlərdə ehtimal olunan davranışlara uyğun olmayan şablonların tapılması problemidir. Bu problem özünü Big data kontekstində daha qabarıq göstərir. Anomaliyaların aşkarlanmasındakı ənənəvi metodlar böyük həcm, müxtəliflik, yüksək sürət kimi kimi xüsusiyyətlərlə təyin olunan böyük verilənlərdə yaxşı nəticə göstərmirlər [4-7].

Anomaliyaların düzgün aşkarlanmaması və ya emalı əldə olunan biliyin etibarlılığına birbaşa təsir göstərir. Ona görə də anomalyaların düzgün identifikasiyası vacib məsələdir, həm də sadə məsələ deyildir.

Hesablama buludları, MapReduce, Hadoop və s kimi texnologiyalar geniş miqyaslı verilənlərin emalında kifayət qədər hesablama gücünə malikdirlər. Bu texnologiyaların köməyi ilə çoxsaylı şəbəkə verilənlərini inteqrasiya və analiz etmək mümkün olmuşdur. Nəticə etibarı ilə böyük verilənlərin analizinə əsaslanan təhlükəsizlik analitikası texnologiyası meydana gəlmişdir [1,5,8].

Təqdim olunan işdə də məqsəd Big data analitika əsasında kompüter şəbəkələrində anomaliyaları daha düzgün, dəqiq analiz etməkdir. Big data analitika optimal (ən yaxşı) qərarlar qəbul etmək üçün böyük həcmli verilənlərdə gizli qanunauyğunluqların, məlum olmayan korrelyasiyaların və digər faydalı informasiyanın aşkarlanması prosesidir. Bu kontekstdə anomaliyaların aşkarlanması çox ciddi problem olaraq həm elmi tədqiqat, həm də tətbiqi sahələrdə diqqət mərkəzindədir [1,8-10]. Müxtəlif informasiya təhlükəsizliyi obyektlərində böyük həcmli verilənlərdə anamaliyaların aşkarlanması tətbiqi kontekstlərdən biridir. Bu məqsədlə məqalədə Big data analitika əsasında şəbəkə hücumlarının aşkarlanması üçün model təklif olunur. Təklif olunan modeldə tədqiqatın obyekti Big data mənbələrindən hesab olunan şəbəkə trafikidir (*analiz üçün serverdə saxlanılan log fayllar və manitoriq verilənləri və s.*).

DOI: 10.25045/NCInfoSec.2017.22

II. ANOMALIYALARIN AŞKARLANMASI METODLARI

Ümumiyyətlə, şəbəkə təhdidlərinin aşkarlanması sistemləri (*Intrusion Detection System - IDS*)informasiya təhlükəsizliyi sahəsində verilənlərin intellektual analizinə əsaslanan ilk sistemlərdir [11,12]. IDS-lərin məqsədi giriş və çıxış tarfiklərə nəzarət etməklə zərərli trafikləri – anomaliyaları aşkarlamaqdır. Bu sistemlər adətən qaydalara və anomaliyalara əsaslanırlar. Qaydalara əsaslanan ənənəvi təhdidlərin aşkarlanması sistemləriməlum şablonları aşkar edə bilir.Səhvlərin çoxluğu (*trfikdə zərərli axını hüquqi kimi təsnifatlandırmaq və ya tərsinə*) və yüksək ötürücülük qabiliyyətinə malik şəbəkələrdə istifadəsinin qeyri-mümkünlüyü, yeni hücumları müəyyən edə bilməməsi isə onun əsas çatışmayan tərəfidir. Anomaliyalara əsaslanan yanaşmalar isə normal davranışdan fərqli şablonları aşkar etməyə qadirdir. Bu da onun üstün cəhətidir.

Verilənlərdə anomaliyaların aşkarlanması hələ XIX əsrdə öyrənilməyə başlanmışdır [3]. Zaman keçdikcə bir çox tətbiq sahələrindəanomaliyaların aşkarlanmasıüçün xüsusi metodlar işlənmişdir. Bəzi metodlar isə daha ümumi nəzəri yanaşmalardan ibarətdir. [3,12,13]-də anomaliyaların aşkarlanması metodlarının geniş icmalı verilir. Bu tədqiqatlarda anomaliyaların aşkarlanması problemlərinin müxtlif aspektləri (*verilənlərin təbiəti; anomaliyaların tipləri–kontekst anomaliya, kollektiv anomaliya; verilənlərin nişanı*), metodikalar əsasında anomaliyaların aşkarlanmasının təsnifatı verilmişdir.

Metodoloji baxımdam şəbəkə anomaliyalarının aşkarlanmsı metodları iki qrupa bölünür: *stoxostik və determinik*. Stoxostik metodlarda verilənlər ehtimala əsasən modelləşdirilir. *Stokastik* metodlar verilənləri ehtimal olunan bir modelə uyğunlaşdırır və yeni trafikin uyğunluğunu bu modelə nəzərən qiymətləndirir. Qiymətləndirmə statistik hipotezə əsasən aparılır. *Determinik* metodlar isə funksiyanı iki hissəyə bölür: “normal” və “anormal”. Sərhədlər isə klaster analiz və SVM metodları və s. ilə təyin olunur. Verilənlər baxımından isə anomaliyaların aşkarlanması metodları axın əsaslı, paket əsaslı və window əsaslı olur [13].

Araşdırmalar göstərir ki, şəbəkə anomaliyalarının aşkarlanması metodları əsasən iki tipə bölünür: nəzarət olunan və nəzarət olunmayan. Nəzarət olunan aşkarlanma metodlarında sistemin və ya şəbəkənin normal davranış modeli təlim verilənləri əsasında yaradılır. Nəzarət olunmayan aşkarlama metodlarında isə heç bir təlim verilənləri istifadə olunmur [3,12,13].

Çox sayda tədqiqatlarda anomaliyaların aşkarlanması üçün statistik yanaşmalar, ölçünün kiçildilməsi, maşın təliminə əsaslanan, neyron şəbəkələr, bayes şəbəkəsi, entropiya, qaydalara əsaslanan, SVM əsaslı və s. model və alqoritmlər təklif olunmuşdur [7,9,14-18].

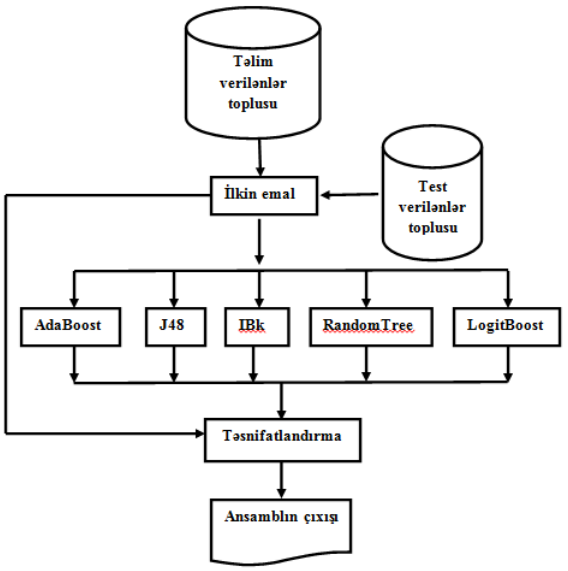
Bir çox tədqiqatlarda anomaliyanın aşkarlanmasında klassifikasiyanın dəqiqliyini artırmaq məqsədi ilə hibrid və ya çox səviyyəli klassifikasiya modelləri təklif olunmuşdur [7, 9, 15, 19, 20].

III. TƏKLIF OLUNAN MODEL

A. Məsələnin qoyuluşu

Məqalədə kömpüter şəbəkəsi trafikində anomaliyaları aşkarlamağa imkan verən multi-klassifikasiya yanaşması təklif edilir (şək.1).

Fərz edək ki, şəbəkə trafikini əks etdirən $D = \{x_1, x_2, \dots, x_n\}$ nöqtələr çoxluğu verilmişdir. $M = \{m_1, m_2, \dots, m_k\}$ sayda klassifikasiya alqoritmləri seçilmişdir. IDS-lərin işinin səmərəliliyinin artırılması üçün şəbəkədə zərərli trafiklərin yüksək dəqiqliklə aşkarlanması tələb olunur.



Şəkil 1. Təklif olunan modelin ümumi sxemi

Təklif edilmişmodelin işlənməsi aşağıdakı addimlardan ibarətdir:

*Addım 1.*Təlim və test verilənlər dəstinin seçilməsi—NSLKDD. Bu dataset haqqında sonrakı bölmələrdə məlumat veriləcək;

*Addım 2.*İlkin emal mərhələsi. Ancaq faydalı informasiyanın saxlanması məqsədi ilə küy təşkil edən verilənlərin təmizlənməsi və yaemal prosesini sadələşdirmək məqsədi ilə normallaşdırma və ya korreksiya metodların tətbiqi;

*Addım 3.*J48, LogitBoost, IBk, AdaBoost, RandomTree klassifikatorlarından ibarət hibrid modelin qurulması ;

Addım 4. Verilənlər üzərində klassifikatorların test edilməsi;

Addım 5. Klassifikator ansamblının yaradılmasında metodun seçilməsi (*məsələn, Stacking*).

B. Dataset – NSL KDD

Şəbəkə anomaliyalarının aşkarlanmasında tədqiqatçıların qarşılşdığı əsas problemlərdən biri də açıq verilənlər dəstinin olmamasıdır. Problemin həlli məqsədiilə 1990-cı illərdə IDS sistemlərin test edilməsi üçün etalon test verilənlərinin vacibliyi etiraf edildi və DARPA verilənləri topluları meydana çıxdı [20].

Qeyd etmək lazımdır ki, hazırda IDS sistemlərinin test edilməsi üçün tədqiqatçılara bir çox verilənlər toplusu əlyetərdir: DARPA, KDD’99, Internet Trafik Arxivi, LBNL, CAIDA, DEFCON, PREDICT, ISCX 2012 və s. DARPA verilənlər topluları şəbəkə təhlükəsizliyi məqsədləri üçün orta ölçülü ABŞ aviabazasında müşühidə edilən şəbəkə trafiki imitasiya edilməklə yaradılmışdır. KDD təlim verilənləri toplusunda imitasiya edilmiş hücumlar dörd kateqoriyaya bölünmüşdür:

- *Xidmətdən imtina hücumu (**Denial of Service, DoS**)* – hücum edən müəyyən servisləri həddindən artıq yükləyir ki, qanuni istifadəçilərə xidmətdən imtina edilsin.

- *User to Root Attack (**U2R**):* Hücum edən müəyyən eksploytlardan istifadə edərək sistemdə normal istifadəçi hesabındanadministrator hesabına yüksəlməyə cəhd edir.

- Remote to Local Attack (**R2L**): hücum edən kompüterdə müəyyənboşluqlardan istifadə edərək həmin kompüterdə lokal istifadəçi hesabına giriş əldə etməyə cəhd edir.

- ***probing attack:*** Probe hücum informasiya təhlükəsizliyini pozmaq məqsədilə kompüter şəbəkəsi haqqında informasiya toplamağa cəhd edir.

DARPA və KDD verilənlər topluları köhnəlmiş hesab edilir, bir sıra nöqsanlara malikdirlər, lakin hələ də istifadə edilirlər. Nöqsanları aradan qaldırmaq məqsədilə KDD`99-un təkmilləşdirilmiş versiyası – NSL-KDD verilənlər toplusu işlənmişdir [21]. Təlim toplusunda izafi yazılar,test toplusunda təkrar yazılar yoxdur və s. Bunları nəzərə alaraq məqalədə 125973 yazıdan təşkil olunmuş NSL-KDDTrain dataset-dən və 22544 yazıdan ibarət NSL-KDDTest verilənlərindən istifadə edilmişdir. NSL-KDDTrain dataset-də hər bir yazı nümunəsi normal və ya anomaliya kimi qeyd olunmuş 42 atributdan ibarətdir. [22-24]-də verilənlər, atributlar, onların adları, təsviri və s. haqqında ətraflı məlumat verilmişdir.

C. WEKA (Waikato Environment for Knowledge Analysis)

Verilənlərin intellektual analizi üçün mükəmməl alətlərin olması çox vacibdir. Verilənlərin analizi və maşın təlimi alqoritmlərini yerinə yetirmək üçün yaradılan Weka açıq proqram təminatının ilk versiyası 1993-cü ildə Yeni Zelandiyanın Vaikato universitetində Java proqramlaşdırma dilində yazılmışdır. Bu da onun istənilən kompüter platformada istifadəsinə imkan verir. WEKA tədqiqatçılara ilkin emal alətləri, çoxsaylı klassifikasiya və klasterizasiya, reqresiya metodları təqdim edir və nəticələrin vizuallaşdırılması imkanı verir. Ötən illər ərzində proqram təminatı inkişaf etdirilmiş, tədqiqatçılara ən müasir imkanlar yaradılmışdır. Qeyd etmək lazımdır ki, son 3.8.1 versiyasının Big data-ya tətbiqi reallaşmişdır [25].

IV. EKSPERİMENTLƏR

Bu bölmədə aparılmış eksperimentlər və onun nəticələri ümumiləşdirilmişdir. Eksperimentlər Windows 8.1(64bit) əməliyyat sistemi, İntel(R) Core(TM) i5-2400 prosessora, 4GB operativ yaddaşa malik kömpüterdə aparılmışdır və şəbəkə trafikində anomaliyaları aşkarlamaq üçün NSL-KDD verilənlər dəstinin təlim və test **.arff** fayllarından istifadə olunmuşdur. Bu fayllar “protocol_type”, “service”, “flag”, “src_bytes”, “dst_bytes”, “land”, “wrong_fragment”, “urgent”, “hot”,dst_host_count və s. kimi 42 atributlardan ibarətdir.

Məqalədə anomaliyaların aşkarlanması modeli bütün əlamətlər vektoru əsasında J48, LogitBoost, IBk, AdaBoost, RandomTree klassifikatorları test edilmişdir. Əvvəldə qeyd edildiyi kimi təklif edilən modelin effektivliyinin qiymətləndirilməsi WEKA mühitində NSL-KDD açıq verilənlər bazaları üzərində aparılmışdır. Klassifikatorlar ansamblının yaradılmasında Stacking metodu tətbiq edilmişdir. Klassifikator ansamblında meta klassifikator olaraq SVM Radial Basis Function götürülmüşdür. Klassifikatorların aşkarlama dəqiqliyi dəqiqlik (*precision)*, tamlıq (*recall*), yanlış müsbət hallar (false positive rate-*FPR)*, doğru müsbət hallar(true positive rate - *TP*), *f-ölçü (f-measure)*, doğruluq (*accuracy)* metrikaları əsasında qiymətləndirilmişdir.

Nəticəni aşağıdakı kimi imitasiya etmək olar. Cədvəl 1-dən də göründüyü kimi təklif edilən yanaşmanın bütün metrikalar üzrə aşkarlama dəqiqliyi digər metodlarla müqayisədə yüksəkdir. Təklif edilən modelin anomaliyaları aşkarlama dəqiqliyi 83 faizdən bir qədər artıqdır.

CƏDVƏL 1. KLASSİFİKATORLARIN DƏQİQLİYİNİN MÜQAYİSƏSİ

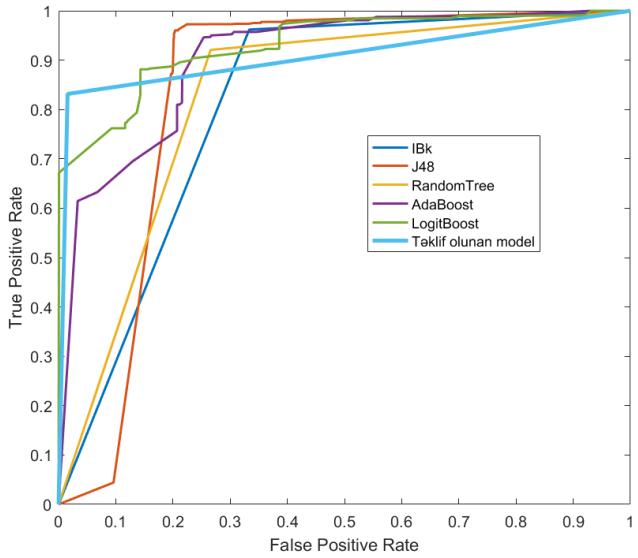
Metodlar	**TP**	**FP**	**Precision**	**Recall**	**F-measure**	**Accuracy**
Ada Boost	78.4%	17.4%	83.4%	78.4%	78.3%	78.44%
Random Tree	81.4%	16.0%	83.7%	81.4%	81.4%	81.36%
IBk	79.4%	16.5%	84.1%	79.4%	79.2%	79.364%
J48	81,5%	14.6%	85.8%	81.5%	81.5%	81.53%
Logit Boost	74.7%	21.0%	79.7%	74.7%	74.5%	74.72%
Təklif olunan model	82.9%	12.0%	87.5%	82.9%	83.2%	**83.09%**

Şəkil 2-də təqdim olunan ROC (Receiver operating characteristic) əyrisi vizuallaşdırma aləti kimi istifadə olunmuşdur. Bunun üçün Matlab proqram təminatında istifadə olunmuşdur. Əyridə *x* oxu yalanları, *y* oxu isə mənfi yalanları əks etdirir.

Araşdırmaların nəticəsi olaraq deyə bilərik ki, anomaliyaların aşkarlanmasında ümumi (universal istifadə olnan) yanaşma yoxdr. Bu sahədə mövcd metodların (*statistik, neyron şəbəkələr, maşın təlimi metodları, optimallaşma metodları və s.*) spektri çox genişdir.

Şəbəkə trafikinin anamaliyaları şəbəkənin təhlükəsizliyi üçün çox ciddi fəsadlar yarada bilər. Şəbəkə trafikini analizinin nəticələri şəbəkə mühəndislərinə və inzibatçılarına daha etibarlı çəbəkə yaratmağa, mümkün yüklənmələrdən qaçmağa və qabaqlayıcı tədbirlərin görülməsinə kömək ola bilər. Bunun

üçün daha effektiv intellektual analiz metodlarının işlənməsinə ehtiyac vardır.



Şəkil 2. ROC əyrisi

NƏTİCƏ

Təqdim olunan multi-klassifikator modelinin eksperimentlərinin nəticəsi modelin təhlükəsizlik obyektlərində tətbiqinə imkan verir. Gələcəkdə bu modelin aşkarlama dəqiqliyinin yüksəldilməsi istiqamətində əlamətlərin kiçildilməsi, optimallaşma və s. kimi yanaşmalar işlənilməlidir.

MİNNƏTDARLIQ

Bu iş Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun maliyyə yardımı ilə yerinə yetirilimişdir – **Qrant № EİF-KETPL-2-2015-1(25)-56/05/1**

ƏDƏBİYYAT

[1] M.Ş. Hacırəhimova, "Big Data texnologiyaları və informasiya təhlükəsizliyi problemləri", İnformasiya texnologiyaları problemləri, 2016, №1, s.49–56.

[2] Global State of Information Security. Survey 2017. https://www.pwc.com/.../global-state-of-information-security

[3] V.Chandola, A. Banerjee, V. Kumar, “Anomaly Detection : A Survey”, ACM Computing Surveys, 2009, pages 1-72.

[4] C.C. Aggarwal and P.S. Yu “Outlier Detection for High Dimensional Data”, Proceedings of the ACM SIGMOD Conference, 2001, pp. 37-46.

[5] R. Zuech et al., “Intrusion detection and Big Heterogeneous Data: a Survey”, Journal of Big Data (2015) 2:3, p. 41. DOI 10.1186/s40537-015-0013-4

[6] M.Bai, X.Wang, J.Xin,G.Wang, “An efficient algorithm for distributed density-based outlier detection on big data”, Neurocomputing, 2016, vol.181, pp. 19–28.

[7] H. Kim et al.,.”Behavior-based anomaly detection on big data”, Australian Information Security Management Conference, 2015, pp.73-80.

[8] O. Tene, J.Polonetsky, “Big Data for all: Privacy and user control in the age of analytics”, Northwestern Journal of Technology and Intellectual Property, 2013, vol.11, no.5, pp.239–273.

[9] Y.N. Imamverdiyev, L.V. Sukhostat, “Network trafic anomalies detection based on informative features”, Radio Electronics, Computer Science, Control, 2017, no.3, pp.113-119.

[10] R.M.Əliquliyev, M.Ş.Hacırəhimova, "Big Data" fenomeni: problemlər və imkanlar, İnformasiya texnologiyaları problemləri, 2014, №2, s. 3-16.

[11] S.Akbar, “Intrusion detection system methodologies based on data analysis”, International Journal of Computer Applications, 2010, vol.5, No.2, pp.10-20. http://citeseerx.ist.psu.edu/viewdoc/download

[12] V. Hodge and J. Austin, “A survey of outlier detection methodologies,” Artificial Intelligence Review, 2004, vol. 22, no. 2, pp. 85–126.

[13] J.Wang, D.Rossell, etal. Network anomaly detection: A survey and comparative analysis of stochastic and deterministic Methods. https://arxiv.org/pdf/1309.4844.pdf

[14] W. Wang et al.“Statistical wavelet-based anomaly detection in big data with compressive sensing”, EURASIP Journal on Wireless Communications and Networking , 2013:269.

[15] H. H. Pajouh, G.H. Dastghaibyfard1, S. Hashemi, “Two-tier network anomaly detection model: a machine learning approach”, Journal of Intelligent Information Systems, February 2017,vol 48, no.1, pp.61–74.

[16] Y.N. Imamverdiyev, L.V. Sukhostat, “Anomaly detection in network traffic using extreme learning machine”, Proceedings of the IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016, pp.419.

[17] M. Thottan, Chuanyi Ji Anomaly detection in IP networks IEEE Transactions on Signal Processing,2003,vol. 51,no. 8,pp.2191-2204

[18] A.L Buczak., E.Guven, “A survey of Data Mining and Machine Learning methods for cyber security intrusion detection” //IEEE Communications Surveys & Tutorials, 2016, vol. 18, no. 2, pp.1153-1176

[19] Р.М. Алыгулиев, Я.Н.Имамвердиев, Ф.Д.Абдуллаева, “Обнаружение аномалий в облачных Big Data данных”, Сборник материалов XIII Международной научно-технической конференции “Оптико-электронные приборы и устройства”, 2017, стр. 35-37.

[20] RM Əliquliyev, R.M. Alıquliyev, F.C. Abdullayeva, “Bulud infrastrukturunun keyfiyyət göstəricilərində anomaliyaların real zamanda aşkarlanması metodu "Proqram mühəndisliyinin aktual elmi-praktiki problemləri" I respublika konfransının materialları, 2017, səh. 30-36.

[21] KDD data set, 1999; http://kdd.ics.uci.edu/databases/kddcup99/

[22] NSL-KDD data set for network-based intrusion detection systems, 2017. Electronic resours - http://nsl.cs.unb.ca/NSL-KDD/

[23] L. Dhanabal, S.P. Shantharajah, A study on NSL-KDD dataset for intrusion detection system based on classification algorithms, International Journal of Advanced Research in Computer and Communication Engineering 4 (2015) 446–452.

[24] S. Revathi, Dr. A. Malathi, “A detailed analysis on NSL-KDD dataset using various machinelearning techniques for intrusion detection”, International Journal of Engineering Research & Technology (IJERT), 2013, vol.2, no.12, pp.1848–1853.

[25] https://www.cs.waikato.ac.nz/~ml/weka/

ANOMALY DETECTION MODEL IN INFORMATION SECURITY OBJECTS BASED ON BIG DATA ANALYTICS

Aliguliyev R.M[1], Hajirahimova M. Sh.[2]
[1,2]Institute of Information Technology of ANAS,
Baku, Azerbaijan
[1]*r.aliguliyev@gmail.com*, [2]*makrufa@science.az*

Abstract – Anomaly detection is one of the main issues in data analysis and used widely for detecting network threats. The article offers a more precise and simple multi-classifier model for anomaly detection in network traffic based on Big Data. Experiments have been performed on the NSL-KDD data set by using the Weka. The offered model has shown decent results in terms of anomaly detection accuracy.

Keywords – anomaly, Big data analytics, information security, ensemble of classifiers, IDS, NSL-KDD, WEKA.