An Introduction to Topology-Based Network Protection

Hamid Zargariasl

Mediterraneà University of Reggio Calabria, Italy
hamid.zargariasl@unirc.it

***Abstract*— One of the problems that designers of protecting systems face is the protecting policy against cyber-attacks or viruses in the big and unbounded networks. Installing high performance protecting systems for all nodes of big networks is not feasible and at least is not economically reasonable. In this research, we offer a new method for distributing the protecting devices in some selected nodes in the network according to a limited budget. A network is supposed that all of its nodes are subjected to a cyber-attack and all of them are down except to the protected nodes. According to our suggested algorithm, after attack, the survived network will be connected and links all important nodes to each other. The use of network centrality definitions for selecting the initial nodes is the key point if this study. A real network of 240 moving objects is studied to evaluate the validity of algorithm. The result shows that the imaginary survived network is totally connected and includes all the important nodes of the network.**

Keywords*— *Information Security; Network Security; Social Network Analysis; Network Survival

I. INTRODUCTION

The obvious importance of the information is a powerful reason for researchers to try for addressing its security problems. There are two main groups of studies, one the information security itself and the other is security of network that possesses the information.

This approach covers all network types either bounded or unbounded network, but it is more useful for unbounded networks that comprise a big variety of networks.

In this study, a new method is suggested for reduction the vulnerability of network by using the social network analysis. This study finds a group of nodes for putting the anti attack mechanisms on them so that the whole network will not be no longer more vulnerable. This idea can be considered as an optimization method for protection the network by a limited budget.

We also categorize our study as an attempt for analysis and simulation of survivable networks as mentioned in the survey researches such as [1].

II. REVIEW OF SOCIAL NEWORK ANALYSIS

Social Network Analysis (SNA) was initially dealing with the networks that can be created by associating humans to vertices of graphs and any kind of their relations to edges. Notwithstanding, it is totally changed in recent years and is used in any of the human sciences in which entities can be modeled by nodes and links of a graph.

Network analysis in 1950s was primarily defined [2] and consequently was grown and spread in different fields. It is an interdisciplinary science that has crawled to a big varieties of scientific works including pedagogy [3,4,5,6,7,8,9,10], psychology [11], criminology [12], management [13], medical sciences [14], and so on.

Nodes in these so-called social graphs are referred as actors or vertices and their relation are called links, ties or edges. Nodes may be used as symbol of humans or other entities (like organizations). The networks in which nodes represents only one entity are called 1-mode , while if the nodes can be divided into two distinguished groups with different nature then the networks are called 2-mode networks [15]. For example, if we suppose the connections between a group of students, this is a 1-mode network and if we suppose a network that is showing the belonging of any student to a faculty, this can be modeled as a 2-mode network because of the existence of two groups of nodes (students and faculties). All 2-mode networks can be transformed into 1-mode networks for easing analysis.

Centrality is a frequently used term in social network analysis able to describe the importance of nodes in the networks and is divided into different categories [16].

Network models

Due to the interdisciplinary nature of social network analysis, it is easy to find a lot of studies that hire its rules to solve problems of other scientific fields.

Social network analysis is also closely related to graph theory; graphs that show the social relations among any kind of entity are referred to as social graphs. The basic network models can be classified into three groups:

1-Random graph: It is a network in which all nodes have about the same number of links and distribution of degree is almost flat [17].

2-Small word: It is the same random graph but, the average distances are smaller [18].

3- Scale free: It refers to a network whose degree distribution is in the form of power low where a few nodes have big number of links and the majority of nodes possess a few links [19]. There are a lot of such networks like Internet, World Wide Web, telecommunication routes, and biological network. Network of human social relationship is also a scale free network.

During this study, the social graph of a group of students derived from their cell phone calls were studied and results

showed that the trend of degree distribution - as expected - is power low [20].

There are a lot of tools for network analysis in social network science and they are divided to two groups. The first group is used for analyzing and calculating the parameter values (such as, UCINET [21], SNAP [22], SIENA[23]) and the second is exploited for visualization of the networks (such as, NETDRAW[24], PAJEK [25], SONIA [26]) and to show network changes in form of animation.

III. Centrality Concepts in Social Networks

Centralities are measures to find important and more effective actors (nodes) in the social networks. There are different centralities in social network analysis and any of them have their own definition. In the first view, it seems that counting the connected edges to any node can show its importance in the network. But, other definitions explain that this idea is not comprehensive and there are more important nodes with less ties.

There are different definitions about centrality in the networks and here, the most important centralities in the social network analysis are briefly explained:

Degree centrality of nodes shows the number of connected links to any node. In di-graphs (directed graphs) it is divided to Out Degree and In Degree that show the number of links toward considered node (In Degree) or number of originated ties from node (Out Degree). The values may be shown in normalized form as well.

Closeness centrality [27] indicates how a node is close to other nodes of the network. Higher values shows that the considered node can access a lot of nodes in a few steps. The k-steps closeness centrality indicates how many steps must be traversed to other actors in k steps. For example, the “2 Step Reach Centrality” demonstrates the ability of any node to reach to other nodes in just two steps.

Betweenness centrality [28] implies the position of node to be in the routes of connection of a big number of nodes. A node has higher betweenness centrality if is located in a position in the graph such that a lot of node pairs need to pass that point for communication. Such nodes can also be informed about context and volume of communication between a big fraction of node pairs.

Eigenvector centrality [29] is a form of degree centrality where the connected nodes to its neighbors are considered. The higher values show that, despite having less links directly connected to the node, its neighbors have more links and the considered node has ability to connect to a lot of nodes in a few steps.

Bonacich Power centrality shows Bonacich's power based centrality of nodes in the network.[30].

Average Reciprocal Distance of nodes (ARD) centrality is also used for calculating reaching distances regarding to the both-way routes. It is defined for di-graphs where reaching route in opposite directions are not similar [31].

IV. Methodology

We supposes that if we have a limited budget for protecting a network of objects, which of them must be equiped to the protecting utilities (such as firewall)?

The idea emerges from the surviving a network after any serious attack that blocks all unprotected nodes. The survive nodes may have not connection to other nodes and impropoer selectng the nodes may cause the survived network to be split to different separated networks. The desired group of protected nodes must satisfy two goals: 1) Saves the most important nodes and 2) renders a connected survived network.

Our solution is selecting nodes based on their centrallity in the network and for this objective, we consider a network of 240 objects that only 20 nodes (8.3%) are proteceted.

We suppose that the protecting hardware or software can totally save the functionality of protected nodes and their connecting links in any kind of cyber attacks.

We aim to show that if the selecting of initial nodes (seed) for installing the protecting utilities are based on network centrality, the survived network will be a connected network comprising all important nodes. The degree centrality is considered as an example for our approach.

This approach must be done by adding important nodes selected by experts (no network roles) and must be added to the network of central nodes. In lack of latter nodes, we envision that important nodes are also central in the network. It is imagined that connecting important (technical) nodes and network central nodes can prepare a good initial selecting seed for better surviving the network in any attack.

We summarize our proposed algorithm as:

1) Finding network’s central nodes by network analysis.
2) Indicating the more important nodes based on technical aspects.
3) Evaluating the existence of link between two nodes mentioned in steps 1 and 2.
4) Selecting final nodes by considering a compromising between two groups.

In practice, we examine the idea only for the central nodes and it is supposed that nodes that are selected by experts are within the central nodes.

The Data Set Generation

The completed operations for creating the network of objects’ co-presence can be summarized as follow:

1) Achieving an Internet survey for finding the probability of any object in any place type
2) Using the data of human mobility both real and also simulated one
3) Finding co-presence of human nodes including places type, duration, frequency

4) Substituting object or group of objects with human nodes according to the probability of presence of objects in any place.

For collecting the probabilities of presence of different objects in different places, a survey on the Web conducted by about 450 volunteers (mainly students and staff of the three Universities of the authors) was carried out. The result of this survey indicated the number and type of electronic gadget carried out in any place. In turn, the attending probability of any object in any place has been easily estimated.

Concerning mobility, after having analyzed a large number of data sets, the one collected by the University of Milan [105], hereafter called Milan data set, seemed to be the most appropriate for our purposes, because it offers the opportunity to recover some information about the kind of places where people meet.

Milan data set was directly downloaded from CRAWDAD website. In it, human mobility real data was collected by using sensors, although this refers to two types of places, 44 nodes, and 19 days only. The fixed nodes are considered as places, which were in number of five and indicated by five different IDs. We received from the authors of the research relevant to the Milan data set a description of their sampling nodes and further information such as which were the fixed and moving nodes and which was the type of place where the fixed nodes were installed. The collected data were very informative for our purpose and we could finally find relatively place-aware data, though, it was not a comprehensive data set. Indeed, only two place types are considered. Classrooms and corridors were supposed as working places (W) and cafeteria as entertainment place (S). In this case, there were five fixed sensors: three in cafeteria (nodes 45, 46, 47) and two in saloons (nodes 27, 43). We considered the five fixed nodes as places and looked for co-present nodes, i.e., if two (or more) nodes are seen by a fixed sensor, then we infer that these nodes are co-present; then, the overlapped times of being seen by fixed sensor are assumed to be the co-presence time duration of two nodes. Obviously, the type of place - as indicated previously - is recorded in any co-presence event.

The probabilities collected from the Internet survey are used to associate the right object/objects to any human in a given place. In so doing, we used the following procedure. After having found the co-presence of human nodes (persons) in any place, we exploited the probabilities from the survey to discover objects carried by each of them.

In doing so, if one human owns O_1 objects and the other owns O_2 objects, then a maximum of $O_1 \times O_2$ potential inter-object links can be established. The subset of links actually established among objects is chosen according to the probability that a given object is actually brought along with the user. More specifically, one can consider P_i being the probability that the i-th object of a group of O_l possible objects of the l^{th} person is carried to a specific place.

By using the values of P_i, human nodes may be replaced by objects in any kind of place where they meet, and, consequently, two sets of objects belonging to two connected humans will be connected. This operation can be seen as mapping human graph $Gh(V_h,E_h)$ onto the graph of objects $Go(V_o,E_o)$. As connections between two fixed objects belonging to two persons is not acceptable (people do not bring with them their fixed objects, such as for example TV), we take this feature into due account in our study.

The result of mentioned operation is a graph of 240 objects that any link implies to at least one co-presence of objects (Figure 1).

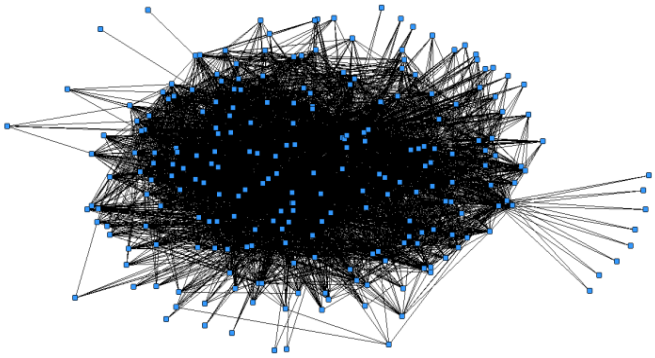

Figure 1 – Network of objects' co- presence

We select 20 nodes which have the higher values in Degree centrality (calculated by UCINET) and delete all other nodes to simulate a real attack to the network. It is imagined that these central nodes are protected by powerful protecting utilities, then their functionality and connections to other nodes can survive after attack.

Figure 2 shows the rest network after deleting all nodes of network.

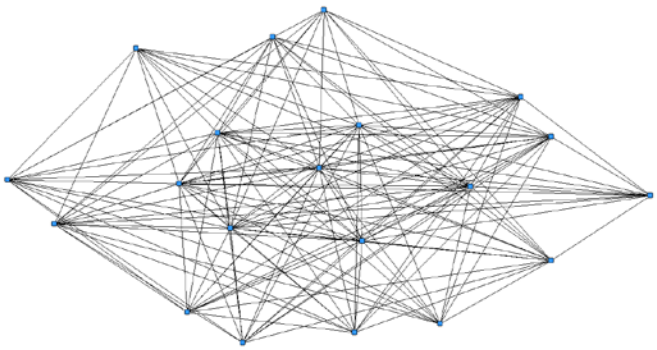

Figure 2. The survived network

It is easily observable that the survived network is connected and can connect all the nodes to each other. It means that installing protecting utilities in central nodes can survive network's connectivity after elimination of all other nodes.

V. Discussion and final remarks

We suggested an algorithm for protecting a big network with immunizing a small group of nodes. This minimization is imposed by limited budget for protecting the network against cyber attacks. We showed that this algorithm causes the survived network after attack to be a connected network (including all important nodes).

We believe that this procedure can be applied in any kind of networks for protecting against attacks. We studied its performance only in one kind of network centrality, but it can be considered in other network centralities for finding the optimized selections.

This study also showed that the protection of central nodes guarantee the connectivity of survived network after any attack.

There may be some combined methods for optimization the selection method that other studies can address them.

It must be added that this algorithm can be used in both bounded and unbounded network and in all real networks.

References

[1] SHEN ChangXiang, ZHANG HuangGuo, FENG DengGuo, CAO ZhenFu, HUANG JiWu, Survey of information security,Sci China Ser F-Inf Sci, June 2007,| vol. 50, no. 3, 273-298.

[2] Bavelas A., Communication patterns in problem-solving groups. Department of Economic and Social Science, Massachusetts Institute of Technology, 1950. grace.evergreen.edu/~arunc/texts/cybernetics

[3] Calvó-Armengol A., Patacchini E., Zenou Y., Peer effects and social networks in education, Review of Economic Studies, v.76, no.10, pages 1239-1267, 2009.

[4] Yang H.-L., Tang J.-H., Effects of social network on student's performance: a web-based forum study in Taiwan, Journal of Asynchronous Learning Networks, v.7, no.3, pages 93-107, 2003.

[5] Zhang Y., Rajabzadeh I., and Lauterbach R., Student network centrality and academic performance: evidence from united nations university, UNU-MERIT Working Paper Series, no.034, 2009. http://www.merit.unu.edu/publications/wppdf/2009/wp2009-034.pdf

[6] Zargariasl H., et al, Analysis the Mobile Social Network of a Class in University, Australian Journal of Basic and Applied Sciences, ISSN 1991-8178, 5(12): 2689-2694, 2011.

[7] Zargariasl H., et al, An Experimental Study on Mobile Social Network Centrality and Educational Performance, 5rd IEEE International Conference on Application of Information and Communication Technology (AICT2011), Baku, Azerbaijan, October, 2011.

[8] Zargariasl H., et al, A study for prediction of the future of undergraduate students by their current social centralities, 8thConference on Applications of Social Network Analysis (ASNA 2011), Swiss, Zurich, September, 2011.

[9] Zargariasl H., et al, A Longitudinal Study on Degree Centrality Changes in a Group of Students, The 4rd IEEE International Conference on Application of Information and Communication Technology (AICT2010), Tashkent, Uzbekistan, 2010.

[10] Zargariasl H., et al, An Experimental Evaluation of Academic Performance of Students and Their Centralities in Mobile Social Network, Journal of Information Technology Institute - Azerbaijan National Academy of Science, Baku, Azerbaijan, 2012.

[11] Zargariasl H., et al, Finding Relationship between Emotional Intelligence and Social Centralities in Undergraduate, 8th Conference on Applications of Social Network Analysis (ASNA 2011), Swiss, Zurich, September, 2011.

[12] Hady W. Lauw Ee-Peng Lim Teck-Tim Tan Hwee-Hwa Pang., Mining Social Network from Spatio-Temporal Events, Computational & Mathematical Organization Theory, 11(2), July, page.97-118, 2005.

[13] Glenn R. Carroll, Albert C. Teo., On the social network of managers, The Academy of Management Journal, Vol. 39, No. 2, 421-440, 1996.

[14] Christakis, N. A. & Fowler, J. H., The spread of obesity in a large social network over 32 years, N. Engl. J. Med. 357, 370–379, 2007.

[15] Stephen P. Borgatti, Social Network Analysis, Two-Mode Concepts in, Encyclopedia of Complexity and Systems Science: 8279-8291, 2009.

[16] Freeman, L.C., Centrality in Networks: I. Conceptual Clarification, Social Networks, 1:215-239, 1979.

[17] P. Erdos, A Rényi, On the evolution of random graphs, the Mathematical Institute of the Hungarian Academy of Sciences 5, 17-61, 1960.

[18] DJ. Watts, SH. Strogatz, Collective dynamics of 'small-world' networks, Nature, 393-440-442, 1998.

[19] AL Barabási, R Albert , Emergence of scaling in random networks, Science, 1999.

[20] Zargariasl H., Scale Free Nature of Mobile Social Networks: An Experimental Study, 10th Conference on Applications of Social Network Analysis (ASNA 2013), Switzerland, Zurich, August, 2013.

[21] Borgatti, S.P., Everett, M.G. and Freeman, L.C., Ucinet for Windows: Software for Social Network Analysis, Harvard, MA: Analytic Technologies, 2002.

[22] J. Leskovec and R. Sosi, SNAP: A general purpose network analysis and graph mining library in C++, http://snap.stanford.edu/snap, 2014.

[23] Ruth M. Ripley, Tom A.B. Snijders, Zsófia Boda, Andras Vörös, and Paulina Preciado, Manual for SIENA version 4.0. Oxford: University of Oxford, Department of Statistics, Nuffield College, 2014. http://www.stats.ox.ac.uk/~snijders/siena/.

[24] Borgatti, S.P., NetDraw: Graph Visualization Software, Harvard: Analytic Technologies, 2002.

[25] Nooy, Wouter d., A. Mrvar and Vladimir Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge University Press, 2005.

[26] Bender-deMoll S, McFarland DA., SoNIA: Social Network Image Animator, 2003. http://sonia.Stanford,edu/

[27] M. Beauchamp, An improved index of centrality, Behavioral Science, 10:161-163, 1965.

[28] L. Freeman, A set of measures of centrality based on betweenness, Sociometry, 40(1):35-41, 1977.

[29] P. Bonacich, Factoring and weighting approaches to status scores and clique identification, Journal of Mathematical Sociology, 2(1):113{120, 1972.

[30] Bonacich, P., Power and Centrality: A Family of Measures Power and Centrality: A Family of Measures, The American Journal of Sociology 92, 5: 1170-1182, 1987.

[31] http://www.analytictech.com/ucinet/help/hs4214.htm