Vulnerability Assessment of Complex Networks based on Optimal Flow Measurements under Intentional Node and Edge Attacks

Igor Mishkovski[1] , Risto Kojchev[2], Dimitar Trajanov[2], Ljupco Kocarev[2, 3],

[1] Politecnico di Torino, Turin, Italy
igor.mishkovski@polito.it
[2] Faculty of electrical engineering and information technologies, Skopje, Macedonia
rkojcev@gmail.com, mite@feit.ukim.edu.mk, lkocarev@feit.ukim.edu.mk
[3] Macedonian Academy of Sciences and Arts, Skopje, Macedonia
lkocarev@manu.edu.mk

Abstract. In this paper we assess the vulnerability of different synthetic complex networks by measuring the traffic performance in presence of intentional nodes and edge attacks. We choose which nodes or edges would be attacked by using several centrality measures, such as: degree, eigenvector and betweenness centrality. In order to obtain some information about the vulnerability of the four different complex networks (random, small world, scale-free and random geometric) we analyze the throughput of these networks when the nodes or the edges are attacked using some of the above mentioned strategies. When attack happens, the bandwidth is reallocated among the flows, which affects the traffic utility. One of the obtained results shows that the scale-free network gives the best flow performance and then comes random networks, small world, and the poorest performance is given by the random geometric networks. This changes dramatically after removing some of the nodes (or edges), giving the biggest performance drop to random and scale-free networks and smallest to random geometric and small world networks.

Keywords: Vulnerability**,** NUM, complex networks, attack strategies, measurements, bandwidth allocation.

1 Introduction

In today's everyday life we are surrounded with complex systems. These complex systems can be represented as networks with a certain number of nodes joined together by edges. Commonly cited examples include social networks, technological networks, information networks, biological networks, communication networks, neural networks, ecological networks and other either naturally occurring or man-

made occurring networks. The topology of these complex networks is one aspect that might help understand in details the surrounding complex systems and its exploration started with the graph theory introduced by Erdős and Rényi [1]. Erdős and Rényi introduced random models in order to model the real complex systems and to capture some of the main characteristics of the real complex systems. However, these models could not give a clear picture of the topology of complex systems and there was an increasing need of new more realistic models. Watts and Strogatz found out that many real world networks exhibit what is called the small world property, i.e. most vertices can be reached from the others through a small number of edges, like in social networks. After the introduction of the Watts and Strogatz's model, Barabási and Albert showed that the structure and the dynamics of the network are strongly affected by nodes with a great number of connections [2]. It was found that many real complex networks have a power-law distribution of a node's degree and by that they are in fact scale-free networks. Additionally, many of the systems are strongly clustered with a big number of short paths between the nodes, i.e. they obey the small world property. Another contribution that helped understand the underlying topology of some real complex system, such as ad hoc networks, is made by Penrose introducing the random geometric graphs and their properties [3].

The above mentioned models helped in understanding the dynamic processes that might occur in the network. Epidemic spreading [4,5], nodes' protection so that the network can resist certain attacks or failures [6], gossip [7] or the process or spreading influence in the network [8], synchronization among nodes [9], cascading failures [10] are some examples of dynamic behaviors of complex networks.

Recently, the primary interest in complex networks is the flow properties of the transport entities. In the complex systems there are many types of flows, such as: traffic flows, information flows, energy flows, chemical flows, idea flows, etc. In particular, the most interesting aspect is how the networks structure affects the flow properties, like traffic congestion [11]. In addition to this, many researchers have studied how attacks or failures of nodes affect the traffic performance in the network [12]. This is a present problem in the real-world networks like the power grids, the Internet, telephone networks and transportation networks. In [13] authors study the robustness to random and intentional node attacks. In this study when a node is attacked, the flows which go through the node have to reconfigure their paths which may affect the loads on the other nodes and may start a sequence of overload failures. Their results show that scale-free networks are highly robust to random node failures but fragile to intentional node attacks, while the random graphs are robust under both node attacks. In their results, the flow rates are assumed to be fixed even after the reconfiguration of flow paths. In [14] authors study the effect of random and intentional attacks on the traffic performance in the Internet. They define some indicators to measure the traffic performance and show how they are affected. In [15] authors analyze the total throughput of ad hoc networks with different network interaction models at communication level, such as: random, small world, scale-free, geographic, full mesh and star models. Their results show that the full-mesh network has highest throughput, while scale-free and star networks show lowest throughput.

In this paper we are assessing the vulnerability of complex networks based on optimal flow measurements under intentional node and edge attacks. We are using four models of complex networks as underlying networks: random, small world,

scale-free and geometric model. On these models we calculate the optimal bandwidth allocation solution for a given flow scenario. Then we are measuring the vulnerability of the network by using different strategies for node and edge removal and calculating the reduction of the total flow under the given scenario, network model and attack strategy.

Therefore, the main goal of this work is to measure and analyze the vulnerability of different complex networks under different node and edge strategies by measuring the total flow in the network.

The rest of the paper is organized as follows. In Section 2 we present the network utility maximization problem with its constraints and utility function. Afterwards, in Section 3 we give the description of the various strategies for intentional node and edge attacks. Simulation results and analysis are given in Section 5 and Section 6 concludes this paper.

2 Network Utility Maximization Problem- NUM

Consider a network with m edges, labeled 1, . . . ,m, and n flows, labeled 1, . . . , n. Each flow has an associated nonnegative flow rate f_j; each edge or link has an associated positive capacity c_i. Each flow passes over a fixed set of links (its route); the total traffic t_i on link i is the sum of the flow rates over all flows that pass through link i. The flow routes are described by a routing matrix $R \in R^{m x n}$, defined as:

$$R_{ij} = \begin{cases} 1 & \text{flow } j \text{ passes through link } i \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Thus, the vector of link traffic, $t \in R^m$, is given by $t = Rf$. The link capacity constraint can be expressed as $Rf \leq c$.

The aim of transmitting a flow of packets from their source to the destination is to get some benefit from the information transmission. Thus, it is natural to set a utility function U_j for flow j, and assume that U_i is related to its rate f_j. In this work as a utility function we use a function which provides proportional fairness among the end users:

$$U(f_j) = \log f_j \quad (2)$$

This function is strictly concave, because the second derivative is negative. From the concavity of the utility function it follows that the optimal rates $\{\hat{f}_j\}$ satisfy the following condition:

$$\sum_j \frac{f_j - \hat{f}_j}{\hat{f}_j} \leq 0, \quad (3)$$

This means that if rate of one transmitter rises, the rate of another transmitter will drop, and the drop will be proportionally larger than the rise. This property is known as the law of diminishing returns.

In order to maximize the utility we have to solve the following convex problem:

$$\begin{aligned} &\text{maximize } \sum_{j=1}^{n} \log f_j \\ &\text{subject to } Rf \le c, \end{aligned} \tag{4}$$

with variable f, and the implicit constraint $f \ge 0$.

Some comments about the NUM problem are given in the text below.

An unfair resource allocation is also possible, in which the goal is to maximize the overall throughput without any consideration about the fairness among the end users. If this is the case, then the unfair utility function would be:

$$U(f_j) = f_j \tag{5}$$

Additionally some reformulations and relaxations can be used by which the NUM problem can be decomposed both horizontally and vertically, and can be solved in distributed manner as in [16] and [17]. These decompositions are not needed for our analysis, because we are interested in overall network performance, so we solve the problem in a centralized manner.

In order to represent the performance of the complex network we use the maximum end-to-end throughput (*MT*) as performance indicator. *MT* is the total amount of bits received by all nodes per second and is measured in Mega bits per second (Mbps):

$$MT = \sum_{j \in n} f_j \tag{6}$$

3 Attack Strategies

In order to assess the vulnerability of the network we are considering two kind of intentional attacks: node and edge attack. In the network of computers attacks on nodes can be interpreted as breakdowns of servers by malicious hackers, while the attacks of edges may correspond to the cutting off the communication links. Additionally, the attacker can choose different strategies for node or edge removal, which are based on various centrality measures. These centrality measures can be based on the initial information about the network or on the information obtained by recalculation, when some of the nodes or edges are removed. We call the first ones *initial* and second ones *recalculated*. In the part below we will describe the different centrality measures that we are using for node or edge removal.

3.1 Degree Centrality – DEG

This measure is based on the idea that more important nodes (edges) are more active, that is, they have more neighbors in the graph [18,19]. It may be used for finding the core nodes (or edges) of a certain community. In order to use this measure for edge attack we are defining the edge degree k_e from the local information of the node degrees [14]:

$$k_e \equiv k_v k_w \tag{7}$$

where the edge e connects two nodes v and w with node degrees k_v and k_w, respectively.

3.2 Betweenness Centrality – BTWN

This is a measure of the importance of a node in a network, and is calculated as the fraction of shortest paths between node pairs that pass through the node. Betweenness is, in some sense, a measure of the influence a node has over the flow of information through the network. Let G be a graph given with set of nodes V and set of edges E. Let s and t are be nodes of the graph. σ_{st} is the number of paths that pass from s to t. Let $\sigma_{st}(v)$ be the number of shortest paths that pass through the node v. The central betweenness of node v is:

$$C(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{8}$$

Just like node betweenness denotes the importance of the nodes, the edge betweenness, in the similar way assigns values to links according to their importance. It is calculated as a number of shortest paths that pass through the edge. Let $\delta_{st}(e)$ be the number of shortest paths from s to t that pass through the edge e and δ_{st} be the total number of paths from s to t. The edge betweenness of edge e is:

$$C(e) = \sum_{s \neq t \in V} \frac{\delta_{st}(e)}{\delta_{st}} \tag{9}$$

3.3 Eigenvector Centrality (Pagerank) - PR

With this measure we can find out the importance of nodes according to the adjacent matrix of a connected graph [20,21]. It assigns relative scores to all nodes in the network based on the principle that connections to high-scored nodes contribute more to the score of a node than connections to low-scored nodes. In order to use this centrality measure for finding the importance of edges we first transform the node adjacency matrix into edge adjacency matrix and then we use the pagerank algorithm. The transformation is done in a way that we say that two links are neighboring if they are connected to the same node.

5 Simulation and Results

For our simulations we are using the above mentioned network models, where each network generator generates 5 samples of the 4 network models. Each sample has 100 nodes and average node degree around 6. The number of flows in each scenario is 1000 and each O-D (Origin – Destination) pair is generated randomly. The flow rate f_j

is also generated randomly and it is between 0 and 1. The capacity c_i of the all links is equal to 1.

In order to solve our network utility maximization problem defined with (4) we are using CVX [25]. CVX is a modeling system for disciplined convex programming (DCP). DCP is a methodology for constructing convex optimization problems and is meant to support the formulation and construction of optimization problems that the user intends from the outset to be convex. DCP imposes a set of conventions or rules. Problems which follow the rules can be rapidly and automatically verified as convex and converted to solvable form. Some problems can be reformulated to be made convex and then solved by appropriate methods for convex problems.

The simulation starts with calculating the maximum end-to-end throughput (6) for the given network. Afterwards, we attack a certain node (or edge) by removing it from the network, using one of the mentioned strategies in Section 3. The flows which originate or end at this node are randomly transferred to a different node, while the flows which go through the node reconfigure their routes to find new shortest paths. The removal of the nodes (or links) changes the entries in the adjacency matrix. Using the new routing information we compute a new optimal bandwidth allocation using (4). In these simulation scenarios we use static routing, that means that we do not take into consideration the load balancing. In addition, the simulation for a given network stops when the network falls apart into two or more islands.

In the next part we will show and analyze some of the interesting results we have obtained in our simulations.

In Fig. 1 we show the flow for the ER when nodes are attacked with the suggested recalculated strategies. One can see that the flow is decreasing in the same fashion for all the three strategies. The only difference between the strategies is that the pagerank disconnects the network when the smallest number of nodes is removed. By removing 8% of the nodes the flow decreased by 43%.

Fig. 2 shows the same analysis only now strategies based on initial information are used. We can see quite interesting phenomenon, i.e. by removing the third most important node, the flow increases instead of decreasing. For the explanation of this phenomenon refer to [14]. Additionally, with these attacks based on initial information when removing 8% of the nodes the flow decreased by around 37%.

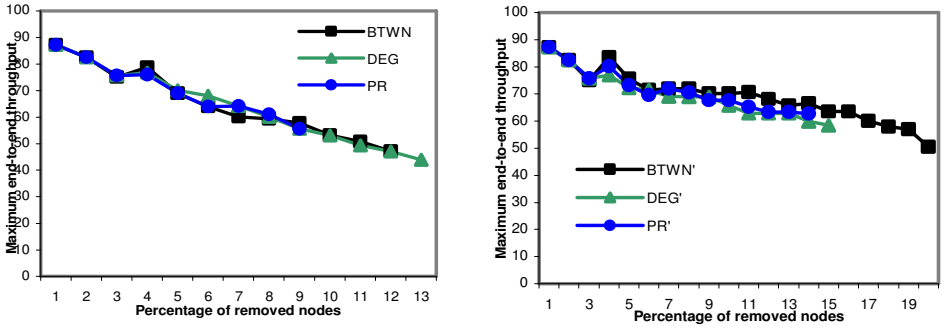
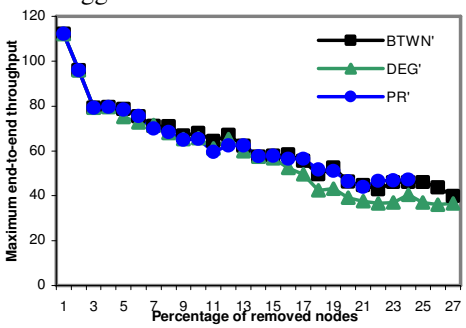


Fig. 1. Maximum end-to-end throughput for the random networks (ER) when attacking nodes using recalculated strategies, such as: betweenness centrality (BTWN), degree centrality (DEG) and pagerank (PR)

Fig. 2. Maximum end-to-end throughput for the random (ER) networks when attacking nodes using strategies based on initial information, such as: betweenness centrality (BTWN), degree centrality (DEG) and pagerank (PR)

For the scale-free networks the recalculated strategies for attack gave the same performance and they disconnect the network only when 5% of the most important nodes were removed. For the attacks based on initial information, from comparing fig. 3 with figures 1, 2, 4, and 5, one can see that the decreasing slop of the flow curve is much bigger than for the rest of the networks.



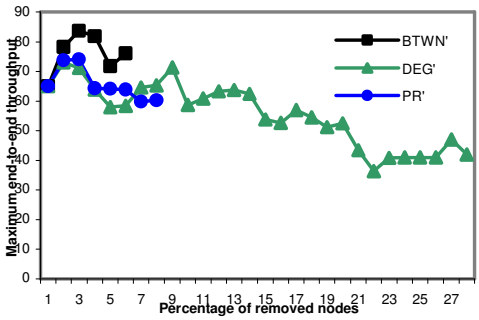


Fig. 3. Maximum end-to-end throughput for the scale-free (SF) networks when attacking nodes using recalculated strategies, such as: betweenness centrality (BTWN), degree centrality (DEG) and pagerank (PR).

Fig. 4. Maximum end-to-end throughput for the random geometric (GR) networks when attacking nodes using strategies based on initial information, such as: betweenness centrality (BTWN), degree centrality (DEG) and pagerank (PR)

Fig. 4 shows how the strategies based on initial information affect the flow in GR networks. It is noticeable that BTWN disconnects the network when the smallest number of nodes is removed. After which came pagerank and the degree strategy needs around 27% of the nodes in order to disconnect the network. In addition, the slope of the flow curve is the smallest, which means that this kind of attacks does not reduce the flow too much, like in the other networks. From Fig. 4 we can notice the same phenomenon, mentioned before, i.e. by removing certain nodes the maximum-end-to-end throughput increases instead of decreasing.

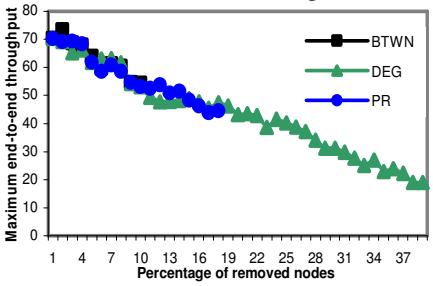


Fig. 5. Maximum end-to-end throughput for the small world (SW) networks when attacking nodes using recalculated strategies, such as: betweenness centrality (BTWN), degree centrality (DEG) and pagerank (PR).

For the small world network when using recalculated strategies by removing 10% of the nodes, the flow decreased only 27%, while when using strategies based on initial information it only decreased for 12% (Fig. 5). This means that this type of network is resistant to intentional node attacks when it comes to measuring the flow in the network. The three types of attack influenced the flow in the same manner. The only difference is that network was disconnected firstly with BTWN, then PR and lastly with DEG.

In Fig. 6 we show the total flow of the four types of network when using PR, based on initial information, as strategy for intentional node attack. The total flow in the networks before removing any node depends on the type of the network. For instance, the highest flow has the SF model, then the ER, SW and the last is the GR model. These results are equal with the results obtained when using game theory and the Method of Successive Averages as a technique for calculating the network vulnerability [26].

The total flow changes dramatically when we start to attack nodes based on the PR technique based on initial information. The highest performance drop has the SF and ER networks. The problem with the GR networks is that they can be easily broken into several disconnected regions (by removing about 1% of the total number of nodes).

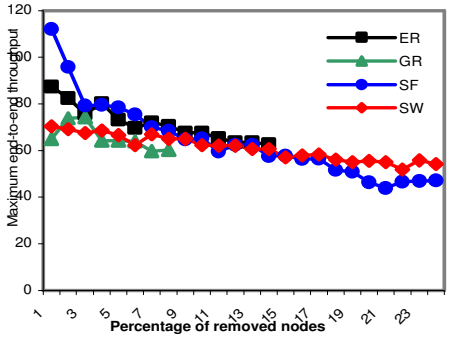


Fig. 6. Maximum end-to-end throughput for all synthetic complex networks when attacking nodes using pagerank based on initial information (PR)

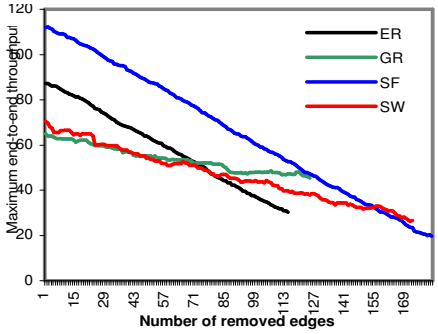


Fig. 7. Maximum end-to-end throughput for all synthetic complex networks when attacking edges using recalculated pagerank (PR)

We encountered almost the same results when instead of nodes we were attacking edges using the recalculated pagerank algorithm (see Fig. 7). It is noticeable that SF and ER networks at the beginning show the best performance, but after removing some of the edges (around 12%) the SW and GR networks outperform the ER network. Then when we continue to remove more edges (around 21%) the GR network performance is close to that of the SF network. At the end when we removed around 27% of the edges the SW outperforms the rest of the networks, when we measure the maximum end-to-end-throughput. When we removed 15% of the edges,

the highest drop in the flow performance showed the ER network (around 43%), then SF (around 42%), then SW (around 36%) ant the lowest drop GR (around 26%). In order to disconnect the network, by attacking the edges with the PR strategy, the most robust to attacks was the SF (around 30% of the edges were needed to disconnect the network), SW (around 28%), ER (around 22%) and GR (around 18%).

6 Conclusion

This brief has studied the attack (node and edge) vulnerability of the different models for complex networks when the maximum end-to-end throughput of the network was taken into consideration. All of the models for complex networks show a considerable decline in performance when they encounter an intentional node or edge attack. One of the obtained results show that the scale-free networks have the highest maximum end-to-end throughput, but when removing nodes or edges the throughput decreases dramatically. The sharp decrease was also the case in the random networks. Additionally, it was shown that among the suggested recalculated and strategies based on initial information there is no big difference when we measure the throughput, whereas they differ in the percentage of nodes (or edges) needed to be removed in order to disconnect the network. The throughput is decreased more when instead of strategies based on initial information we are using recalculated strategies.

As a future work instead of static routing we want to use dynamic routing with load balancing, which takes into account the current flow in the edges, and by that we want to obtain more realistic results. Another improvement would be, instead of removing nodes (or edges), to use certain nodes to generate jam traffic in the network in order to reduce the maximum end-to-end throughput in the network, which presents a more realistic scenario than to remove some important node (or edge) in the network, which can be highly secured and protected.

References

1. P. Erdős, A. Rényi: On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17–61
2. A.-L. Barabási, R. Albert: Emergence of scaling in random networks, Science 286, Oct 1999, pp. 509–512
3. Mathew Penrose: Random Geometric Graphs, Oxford University Press, New York, 2004
4. M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256
5. Z. Dezso, A.-L. Barabási, Halting viruses in scale-free networks, Phys. Rev. E 65 (2002) 055103
6. V. Latora and M.Marchiori: How the science of complex networks can help developing strategies against terrorism, Chaos, Solitons and Fractals 20 (2004), 69–75
7. Nekovee, M. et al.: Theory of rumour spreading in complex social networks. Physica A, 374, pp. 457-470 (2007)

8. D. Kempe, J. M. Kleinberg, and _E. Tardos: Maximizing the spread of influence through a social network. ACM SIGKDD international conference on Knowledge discovery and data mining, 2003
9. P. Checco, M. Biey, G. Vattay, and L. Kocarev: Complex network topologies and synchronization, in Proc. ISCAS '06,Kos, Greece, May 2006, pp. 2641–2644
10. P. Crucitti, V. Latora, and M. Marchiori: Model for cascading failures in complex networks," Phys. Rev. E, vol. 69, p. 045104(R), 2004
11. R. Guimera, A. Diaz-Guilera, F. Vega-Radondo, A. Cabrales, A. Arenas: Optimal network topologies for local search with congestion, Phys. Rev. Lett. 89 (2002) 248701–248704
12. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang: Complex networks: Structure and dynamics, Phys. Rep., vol. 424, pp. 175–308, 2006
13. A. E. Motter and Y.-C. Lai: Cascade-based attacks on complex networks, Phys. Rev. E, vol. 66, p. 065102(R), 2002
14. P. Holme, B.J. Kim, C.N. Yoon and S.K. Han: Attack vulnerability of complex networks, Phys. Rev. E, 65(2002), 056109
15. M. Mirchev, S. Filiposka, N. Trajkovski, D. Trajanov: Network utility maximization in ad hoc networks with different communication patterns, ETAI 2009, Ohrid, Macedonia (2009)
16. F.P. Kelly, A.K. Maulloo, and D.K.H. Tan: Rate control in communication networks: shadow prices, proportional fairness and stability, J. Optical Research Society, Vol. 49, Mar 1998, pp. 237–252
17. S. Kunniyur and R. Srikant: End-to-end congestion control schemes: Utility functions, random losses and ECN marks, IEEE/ACM Transactions on networking, Vol. 11(5), Oct 2003, pp. 689-702
18. L. Freeman: Centrality in social networks: Conceptual clarification, Social Networks, vol. 1, no. 3, pp. 215–239, 1979
19. J. Nieminen: On the centrality in a graph, Scandinavian Journal of Psychology, vol. 15, no. 1, pp. 332–336, 1974
20. P. Bonacich: Factoring and weighting approaches to status scores and clique identification, Journal of Mathematical Sociology, vol. 2, no. 1, pp. 113–120, 1972
21. P. Larry, B. Sergey, R. Motwani et al.: The PageRank citation ranking: Bringing order to the web, Online: http://citeseer. nj. nec.com/page98pagerank. html [04.06. 2003], 1998
22. Michael Karonski and Adrzej Rucinski: The Origins of the Theory of Random Graphs, The Mathematics of Paul Erdos, Berlin, Springer, 1997
23. L. A. N. Amaral, A. Scala, M. Barthe lemy, H. E. Stanley: Classes of small-world networks, PNAS 97:11149-11152
24. A. L. Barabasi: Linked, Penguin Group, London, May, 2003
25. CVX: Matlab Software for Disciplined Convex Programming. Available: http://www.standofd.edu/~boyd/cvx
26. Mishkovski Igor, Sonja Filiposka, Sasho Gramatikov, Dimitar Trajanov and Ljupco Kocarev: Game Theoretic Approach for Discovering Vulnerable Links in Complex Networks, International Joint Conferences on Computer, Information, and System Sciences, and Engineering, University of Bridgeport, USA 5-13 Dec. 2008