Study and Improvement of Robustness of Overlay Networks

Hema Swetha Koppula, Kumar Puspesh and Niloy Ganguly

Department of Computer Science & Engineering
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal - 721 302, India

***Abstract*—The heterogeneity present in the real-world networks like peer-to-peer networks make them particularly vulnerable to attacks as large-scale cascade may be triggered by disabling a set of key nodes. In addition to this vulnerability towards dynamic events, real world networks react quite strongly towards certain types of attacks which may adversely affect their static properties. This brings an obvious concern for the security and robustness of these systems. In this paper, we present empirical results that show how robustness of overlay networks, measured in terms of different parameters like size of largest connected component, number of components and diameter, can be improved by applying various edge modification schemes. We also consider the dynamic effect of node removal along with its static impact on the network.**

I. Introduction

The study of attacks on complex networks is important in order to identify the vulnerabilities of real-world networks, which can be used either for protection (e.g., of infrastructures) or for destruction (e.g., in the control of epidemic diseases). Additionally, it can provide guidance in designing more robust artificial networks (e.g., communication networks). An important property of networked systems is their robustness against various types of failures and attacks on network nodes. Although several design methods have been proposed for creating a network that has optimal robustness according to a given measure, in most real world situations we are often faced with an existing network that cannot be substantially modified or redesigned. Moreover, real world networks are result of many different processes, that may not take the robustness into account. For example we can consider the peer-to-peer networks, which are largely decentralized and highly dynamic systems. One cannot have explicit control over their structure to ensure properties like robustness under various types of disrupting events such as a random failure or an intended attack. The robustness of such networks can be improved by a small degree of modification [1].

The modification could be in the form of either edge addition or edge rewiring. The network can be modified at two different stages to increase the robustness. One is a preventive stage in which the network is made more robust so that it does not breakdown under attack or failure. The second stage is after a disrupting event, by applying some repair strategies to restore the original properties of the network. For applying any kind of edge modification to a network to improve its robustness, it is important to understand how the existing topologies deal with failures and attacks. In this paper, we study the effect of random failure and targeted attack on network nodes in a particular peer-to-peer overlay network, a crawl of Gnutella superpeer network. We study both static and dynamic effects of the node removal and see if by suitably modifying the network we can improve it robustness against failures and attacks without appreciably degrading its performance.

The remainder of this paper is organized as follows. Section 2 provides background and related work on various studies on robustness of complex graphs. Section 3 describes our edge modification schemes and the metrics used to measure robustness, and Section 4 describes the simulation methodology. Section 5 discusses implications of our study and we conclude in Section 6.

II. Background and Related Work

Many authors have studied the effect of failures and attacks on various complex networks. Scale-free networks are known to be sensitive to targeted attacks, which are biased towards higher degree, in comparison with random attacks [2]. This is due to the heterogeneity present in the scale free networks. In these networks, degree distribution i.e. probability of a node having degree k, decreases with power of k [3]. Therefore randomly chosen node is likely to have a low degree, so its removal has little effect on the network. Removal of a high degree node can have a significant effect since such a node may hold a large part of the network together by connecting many other nodes. For Erdos-Renyi random graphs [6], there is not much difference between random failures and targeted attacks due to the homogeneous nature of these networks. In these graphs every pair of nodes is connected with a fixed probability p, independently of every other pair. They have a binomial degree distribution, $P_b(k)$, which approaches a Poisson distribution as the number of nodes becomes large. Hence, there is very less chance of encountering a hub. Therefore, targeted attacks have less effect on these graphs. It is found that these networks are more vulnerable to *random failures* than to *intended attacks*, compared to scale-free networks.

A convenient way to address the robustness of a complex network is to examine how the diameter, size of the largest connected component and number of connected components, which measure the efficiency of communication (or information flow) within the network, are changing under random or intentional attacks. But these measures address only the static properties of the networks. Cascading failures have been reported for numerous networks, which refer to subsequent failure of other parts of the network induced by the failure of or attacks on few key nodes. Researchers have investigated mechanisms leading to cascades of overload failures in complex networks by constructing models incorporating the flow of physical quantities in the network [4]. An important question for many real-world situations is how attacks affect the functioning of a network when the flow of information or other physical quantity in the network are taken into consideration. In particular, the removal of nodes changes the balance of flows and it may trigger a cascading failure, as the one that happened on August 10, 1996 in the western U.S. power grid. Authors have shown that for networks where network flow can redistribute among the nodes, intentional attacks on highly loaded nodes can trigger a large-scale cascade of overload failures [7].

III. Modification Schemes and Metrics

We discuss here the various schemes which are used to increase robustness of networks. In addition to that we discuss some simple measures which can quantify the robustness of any network.

A. Edge Modification Schemes

Various edge modification schemes have been proposed in the literature which aim at improving the robustness of these complex networks [1]. These can be broadly categorized into - Edge Addition schemes and Edge Rewiring schemes. Edge addition schemes result in increased number of edges or connectivity in the network whereas rewiring schemes change the properties of the network while keeping the number of edges constant. In this paper, we have considered the following schemes (Note that 'Random' as used here means randomly chosen with uniform probability and duplication of edges between any two already connected nodes is not allowed)-

1) Random Edge Addition: - An edge is added between any two randomly chosen nodes.

2) Preferential Edge Addition: - An edge is added between two unconnected nodes having the lowest degrees in the network.

3) Random Edge Rewiring: - A random edge is removed and then a random edge is added between two random nodes.

4) Random Neighbor Rewiring: - A node is chosen at random and an edge to a random neighbor is disconnected from that node. The loose end of this edge is connected to a random node.

The *Random neighbor rewiring* is a new edge modification scheme that we have introduced. It is a variation from the previously stated Random neighbor rewiring schemes [1]. If we choose a random neighbor of a randomly chosen node, the probability of the neighbor node having degree k is proportional to kp_k, where p_k is the probability that the randomly chosen node has degree k. Therefore the random neighbors of randomly chosen nodes have higher degree, given that the assortativity is low. In such cases, where assortativity is low, the *Random neighbor rewiring* scheme disconnects the edge connected to a high degree neighbor and reconnects it to a random node, which would be a lower degree node given the power law nature of the scale-free graphs. This tends to bring in a degree of homogeneity into the graph structure, the extent of which depends on the amount of rewiring.

These edge modification schemes can be mapped to different network management processes that take place in unstructured peer-to-peer overlay networks. For example, the superpeers connect to new superpeers which come into the network and disconnect old superpeers with time, in order to exchange network information, as well as to handle the network churn. This process is equivalent to random rewiring if no preference is used in choosing new neighbors. Therefore, studying the effect of these modification schemes on the robustness of the overlay network can help in designing robust network management protocols.

B. Metrics to calculate Robustness

We measure the robustness of the networks on the basis of following parameters

1) *Diameter of the graph*
2) *Size of the largest connected component (LCC)*
3) *Number of components*
4) *Node Failure*

The first three parameters are static measures of robustness of the network, i.e. they do not capture the effect of cascading of the network flow upon a failure or an attack. These three metrics were chosen as they are simple and also capture the essential requirements for a robust network without flow considerations. The *diameter* is a measure of the maximum time for information propagation in the network, whereas the *LCC* and *number of components* measure the availability of the network. The last metric *Node Failure* measures the dynamics of node removals. It shows how many nodes go down due to the overload of flow in the network caused by the previously removed nodes. It is a measure of the cascading effect created due to removing any set of nodes from the network and the breakdown in the information flow caused by it. We show that networks where load can be redistributed among the remaining nodes, targeted attacks on key nodes can lead to breakdown of the whole network.

The various edge modification schemes are studied under the light of how they affect these metrics which are computed as a function of percentage modification for a given percentage of removed nodes. These metrics give us insight into making

the network more robust against attack on nodes by taking proper preventive measures.

IV. Simulation Methodology

Our simulation was mainly concentrated around the preventive measures we introduced in the first section of the paper. We simulated various edge modification schemes on the network graph and then studied the effect of attacks and failures on the resultant graphs. The network graph, modification and attack analysis models are described here.

A. Network Graph

We simulated attack and edge modification schemes and studied their effects upon the peer-to-peer overlay networks. The simulations were performed on the overlay network of size 5000 nodes, obtained by crawling Gnutella. The original network contained more than a million nodes but we selected a connected subset of the original graph for simulation purpose, since the computation of certain metrics is very costly. This subgraph has a heterogenous degree distribution but does not follow power law. Its an hybrid between ER and Power Law graphs. Even though real world networks follow power law and are scale free in nature when the graph is considered as a whole, subgraphs of these networks might not posses these characteristics fully. But they surely have a certain degree of heterogeneity as they are random subgraphs of huge heterogeneous graphs. Since one of the motivations behind the study of the various edge modification schemes is to help in designing robust network management protocols, and since these protocols are most effective when based on local knowledge, it justifies to study the robustness and the effect of the edge modification schemes on random subgraphs of the full network.

B. Edge Modification Model

The edge modification schemes used are *random edge addition*, *preferential edge addition*, *random edge rewiring* and *random neighbor rewiring* as explained in the previous section. First two modification schemes add edges between two nodes which didn't have any edge between them in the original graph. The last two modification schemes try to rewire the edges i.e; number of edges in the network essentially remains the same. Edge modification is applied on the original graph at various percentages (5, 10, 15, 20, 30, 50, 70 %) for each of the four schemes mentioned above.

C. Attack Model

Two types of node removal are studied, *Random Failure* and *Preferential Attack*. In random failure a set of random nodes are removed from the network. In case of preferential attack, a set of nodes with high degree are removed from the network. On each of the original as well as the modified graphs, three levels of failure and attack (5, 10, 15 %) are simulated and the values for the above mentioned metrics were observed. Therefore, the effect of the edge modification is studied by seeing how the measured parameters of the network change with the amount of modification for various levels of failure and attack.

D. Cascaded Failure Model

For studying the cascaded effect of failures, we assume that the number of messages being transmitted through a node is proportional to the betweenness of that node in the network. Also, initially the network is in a stationary state where the load at each node is less than the capacity of that node. Therefore we assign capacities to each node on the basis of its initial betweenness centrality in the network, $(1+\alpha)L$, where L is the initial load (initial betweenness centrality) at each node and α is a small positive fraction. For our simulations we used the value $\alpha = 0.3$. The load at each node at any time step is computed as a function of total number of shortest paths passing through that node. We have used a modification of dijkstra algorithm for computing betweenness centrality of each node [8]. Then a small percentage of nodes is removed using either the Random Failure model or Targeted Attack model. After attack step, loads of the removed nodes are redistributed in the network which changes the betweenness centralities of the remaining nodes. Then each node is checked to see if the load i.e; the betweenness centrality of that node, has exceeded its capacity or not. If yes, the node is treated as failed and removed from the network. This way the cascading of node failures was simulated for a fixed number of time steps or until the network had become stable again.

V. Results

Our results show that both the addition schemes perform better than the rewiring schemes as far as the first three metrics are concerned. Addition of new edges increases redundancy in the paths between any two nodes, and hence increases the *size of largest connected component*, while decreasing the *diameter* and the *number of components*. But edge addition is *costly* as it would lead to extra bandwidth usage in the overlay network. We show some of the results here.

TABLE I
Results of Edge Addition schemes on Gnutella Network (with 5000 nodes)

Random Edge Addition	**0%**	**10%**	**30%**	**50%**
Random failure 5%				
Diameter	12	11	10	9
LCC	4387	4411	4454	4476
# Components	106	81	41	20
Preferential attack 5%				
Diameter	26	23	18	15
LCC	2526	3217	3928	4250
# Components	1528	1007	484	212

Preferential Edge Addition	**0%**	**10%**	**30%**	**50%**
Random failure 5%				
Diameter	12	10	10	9
LCC	4387	4410	4455	4477
# Components	106	82	41	19
Preferential attack 5%				
Diameter	26	23	17	15
LCC	2526	3238	4009	4290
# Components	1528	1009	414	172

TABLE II
RESULTS OF EDGE REWIRING SCHEMES ON GNUTELLA NETWORK (WITH 5000 NODES)

Random Edge Rewiring	0%	10%	30%	50%
Random failure 5%				
Diameter	12	13	15	15
LCC	4387	4391	4384	4369
# Components	106	97	105	118
Preferential attack 5%				
Diameter	26	26	23	21
LCC	2526	3097	3677	3936
# Components	1528	1075	634	437

Random Neighbor Rewiring	0%	10%	30%	50%
Random failure 5%				
Diameter	12	12	12	13
LCC	4387	4338	4264	4275
# Components	106	154	221	210
Preferential attack 5%				
Diameter	26	25	27	21
LCC	2526	2954	3442	3693
# Components	1528	1186	826	615

Table 1 and 2 show some of the simulation results for the various schemes. It can be observed that the number of components increase drastically in case of targeted attack as compared to random failure. As we increase the percentage of rewiring, number of components decrease indicating increased connectivity in the network. Similarly, size of largest connected component (LCC) also grows with the percentage of edges rewired. It can be seen from the results that *Random*

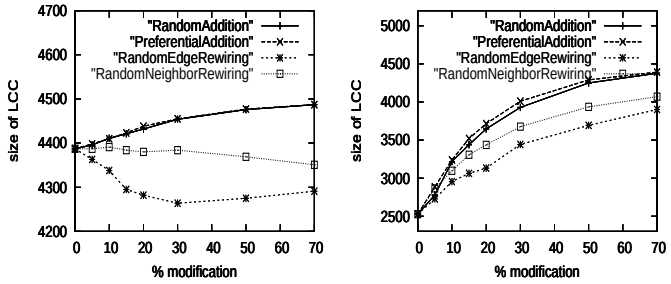

Fig. 1. Size of LCC vs % Edge modification - Preferential Attack 5%(Right) and Random Failure 5% (Left)

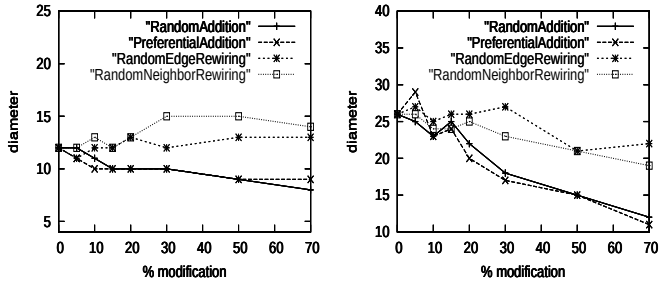

Fig. 2. Diameter vs % Edge modification - Preferential Attack 5%(Right) and Random Failure 5% (Left)

Neighbor Rewiring outperforms other schemes in the static analysis of the network, considering the cost of modifications. This can be explained by the assortativity of the network, having an initial value of -0.19, which means that there is low correlation between the degree of neighboring nodes. Hence, as mentioned before the *Random Neighbor Rewiring* tries to make the network more homogeneous and increases the robustness in terms of availability of the network.

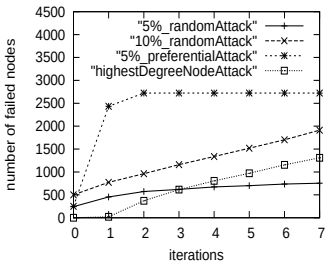

Fig. 3. Total number of nodes failed for random and preferential attacks

Figure 3 shows the failure rate of nodes for random and targeted attacks when cascading is considered. Preferential attack on 5% nodes causes more than half of the nodes to fail in the network (in only two iterations), as expected. It can be seen that the removal of highest degree node is more devastating for the network than attacking 5% nodes of the network randomly.

TABLE III
CASCADING EFFECTS ON REMOVAL OF HIGHEST DEGREE NODE GNUTELLA NETWORK (WITH 5000 NODES)

	0%	10%	20%	50%
Random Edge Addition	1304	904	623	338
Preferential Edge Addition	1304	930	708	295
Random Edge Rewiring	1304	1066	1171	860
Random Neighbor Rewiring	1304	1015	1880	1955

Table 3 shows the cascading effects on removal of the highest degree node from the network. As we had stated earlier, removal of a highly connected node from the network adversely affects the information flow capability of the network. This fact can be easily seen here as removal of the highest degree node from the network causes 1304 nodes to fail in 8 iterations of cascaded analysis. Table 3 also highlights the performance of various edge modification schemes. Clearly, *edge addition schemes* perform better than the *edge rewiring schemes* as they increase the connectivity between nodes. They create more shortest paths between nodes not passing through the highest degree node. Therefore the amount of load to be redistributed after the removal is less, and hence causes less nodes to fail due to the redistribution. The *edge rewiring schemes* do not perform well, as they do not contribute much in shifting the betweenness of the highest degree node to other nodes in the network.

We also evaluated different edge modification strategies when a small fraction of the network nodes are removed. We show the simulation results obtained for 5% random and preferential attacks. Table 4 shows the results for edge addition schemes and we find that when a larger number of nodes in the network

TABLE IV
CASCADED FAILURES WITH EDGE ADDITION SCHEMES ON GNUTELLA NETWORK (WITH 5000 NODES)

Random Edge Addition	Originally	After Cascade, %addition			
Random failure 5%		**0%**	**10%**	**20%**	**50%**
Number of failed nodes	250	755	1080	1232	502
Number of components	106	137	78	118	54
Preferential Attack 5%					
Number of failed nodes	250	2723	3541	4280	4701
Number of components	2526	2373	2135	1854	1194

Preferential Edge Addition	Originally	After Cascade, %addition			
Random failure 5%		**0%**	**10%**	**30%**	**50%**
Number of failed nodes	250	755	1306	434	464
Number of components	106	137	76	70	47
Preferential Attack 5%					
Number of failed nodes	250	2723	3661	4260	4597
Number of components	2526	2373	2116	1794	1095

are *randomly* removed, *preferential addition* is more efficient. Random addition loses out to preferential addition scheme as the randomly chosen nodes which gain edges and contribute in new shortest paths are most likely removed in random failure. In case of *preferential attacks* both the schemes fail to make any improvement in the network.

Rewiring schemes as shown in table 5, also do not perform well in case of *preferential attack* as compared to *random failure*. But it can be seen that at lower modification percentages the *rewiring schemes* are better than *addition schemes*. A high

TABLE V
CASCADED FAILURES WITH EDGE REWIRING SCHEMES ON GNUTELLA NETWORK (WITH 5000 NODES)

Random Neighbor Rewiring	Originally	After Cascade, %rewiring			
Random failure 5%		**0%**	**10%**	**30%**	**50%**
Number of failed nodes	250	755	607	864	663
Number of components	106	137	141	156	168
Preferential Attack 5%					
Number of failed nodes	250	2723	3335	3814	3953
Number of components	2526	2373	2144	1752	1446

Random Edge Rewiring	Originally	After Cascade, %rewiring			
Random failure 5%		**0%**	**10%**	**20%**	**50%**
Number of failed nodes	250	755	805	480	501
Number of components	106	137	117	70	73
Preferential Attack 5%					
Number of failed nodes	250	2723	3215	3702	3968
Number of components	2526	2373	2151	1963	1507

percentage of addition is required to gain more advantage than the rewiring schemes. This observation is particularly important because in case of removing a set of nodes and not just the highest degree node, rewiring is more beneficial and also not as costly as addition. At high modification percentages, edge addition schemes outperform both the rewiring schemes which is expected, but high percentage of addition would also be extremely costly.

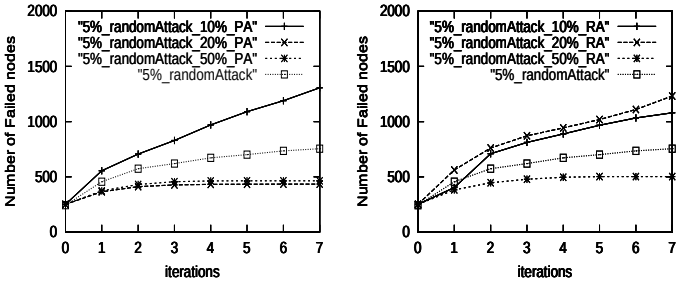

Fig. 4. Total number of nodes failed for 5% random attack with 10% Preferential edge addition(left) and 10% Random Edge addition(right)

VI. FUTURE WORK AND CONCLUSION

In peer-to-peer networks, it is very important to know how to tackle random failures and targeted attacks in an efficient way as they are very common. We have shown that with small modifications we can improve robustness of these networks. We have dealt with the 'preventive' methodology in this paper i.e., trying to modify the network to make it robust against attacks and failures. In our simulation for static analysis, we have noticed that addition schemes perform better than the rewiring schemes as expected, but they are expensive. Considering the cost incurred while rewiring or adding the edges, we see that the *Random neighbor rewiring* performs better than the others as it tries to equalize the degree among all the nodes, making the network more robust against targeted attacks. The cascading effects in the peer-to-peer networks are demonstrated in this paper by taking a simple data flow model. We have also performed the dynamic analysis for the various modification schemes which has given us more insight into the usefulness of the *rewiring schemes* over *addition schemes* when a small fraction of network nodes are removed.

Further theoretical analysis of attack and edge modification model can be done along with the study of changes in the degree distribution due to these schemes. The knowledge of how the various modification schemes affect the robustness of the network can be used to design better distributed network management protocols.

REFERENCES

[1] A. Beygelzimer, G. Grinstein, R. Linsker and I. Rish, *Improving Network Robustness by Edge Modification*, Physica A, Volume 357, Issue 3-4,p.593-612.

[2] P. Crucittia, V. Latorab, M. Marchioric and A. Rapisardab, *Error and Attack Tolerance of Complex Networks*, Nature. 2000 Jul 27, 406(6794):378-82.

[3] R. Albert and A. Barabasi, *Statistical Mechanics of Complex Networks* Reviews of Modern Physics 74, 47 (2002)

[4] Ying-Cheng Lai, A. E. Motter and T. Nishikawa, *Attacks and Cascades in Complex Networks* Lecture Notes in Physics, 2004, Springer.

[5] Jian-jun Wu, Zi-you Gao and Hui-jun Sun, *Cascade and breakdwon in scale-free Networks with community structures* Physical Review E, 2006, APS.

[6] P. Erdos, A. Renyi, *On the Evolution of Random Graphs* Publ. Math. Inst. Hangar Acad. Sci., 5, 1960, 17-61.

[7] A. E. Motter and Ting-cheng Lai, *Cascade-based attacks on complex networks* Physical Review E, 2002, APS.

[8] U. Brandes, *A Faster algorithm for Betweenness Centrality* Journal of Mathematical Sociology, 2001.