Study and Improvement of Robustness of Overlay Networks

Synopsis submitted in partial fulfillment
for the award of the degree of

Master of Technology
In
Computer Science and Engineering

By

Hema Swetha Koppula

Under the guidance of

Prof. Niloy Ganguly

" YOGA KARMASU
KAUSALAM "

Department of Computer Science & Engineering,
Indian Institute of Technology
Kharagpur

1. 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. 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 super-peer 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 propagation of the node failure in the network depends both on the network structure as well as the routing strategy followed to route messages in the network. Different routing strategies choose different intermediate nodes to pass messages between the same end nodes. This leads to congestion at different nodes and hence causes their failure. Therefore, to understand the effect of the routing strategies, we simulated different routing strategy models on the network and measured the cascading effect.

The remainder of this report is organized as follows. Section 2 describes our edge modification schemes and the metrics used to measure robustness. Section 3 describes the simulation methodology, and Section 4 discusses implications of this study. Section 5 describes the routing schemes used and their effect on the previous results and we conclude in Section 6.

2. Modification Schemes and Metrics

The various schemes which are used to increase robustness of networks are discussed here. In addition to that some simple measures which can quantify the robustness of any network are also discussed.

2.1. 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 report, the following schemes are considered (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 $\mathbf{kp_k}$, where $\mathbf{p_k}$ 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 that 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.

2.2. Metrics to calculate Robustness

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

1. Diameter of the graph; measures the maximum time for information propagation in the network
2. Size of the largest connected component (LCC); measures the availability of the network
3. Number of components; measures the availability of the network
4. Percolation Point; measure of the stability of the network
5. Node Failure; measures the dynamics of node removals i.e., cascading effect

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. While considering the dynamic effect of node removal, the percolation point specifies for how long the network contains a giant component and the number of nodes failed tells us the how many nodes suffer due to an initial node removal. 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.

3. Simulation Methodology

The simulations are mainly concentrated around the preventive measures introduced in the first section of the report. 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.

3.1. Network Graph

Attack and edge modification schemes were simulated and their effects upon the peer-to-peer overlay networks are studied. 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 heterogeneous 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 studying the robustness and the effect of the edge modification schemes on random subgraphs of the full network.

3.2. 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.

3.3. 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.

3.4. Cascaded Failure Model

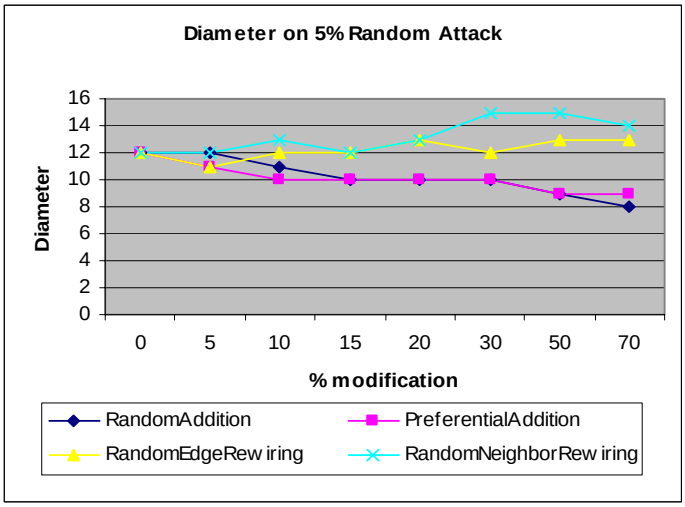
For studying the cascaded effect of failures, we assume that the number of messages being transmitted through a node is proportional to the betweenness of that node in the network. Also, initially the network is in a stationary state where the load at each node is less than the capacity of that node. Therefore we assign capacities to each node on the basis of its initial betweenness centrality in the network, $\mathbf{(1 + \alpha)L}$, where L is the initial load (initial betweenness centrality) at each node and $\boldsymbol{\alpha}$ is a small positive fraction. For our simulations we used the value $\boldsymbol{\alpha = 0.3}$. The load at

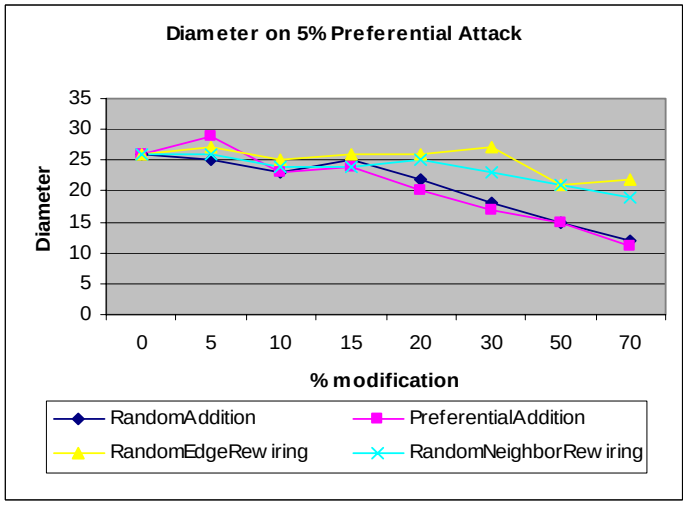
each node at any time step is computed as a function of total number of shortest paths passing through that node. We have used a modification of dijkstra algorithm for computing betweenness centrality of each node[8]. Then a small percentage of nodes is removed using either the Random Failure model or 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.

4. 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.

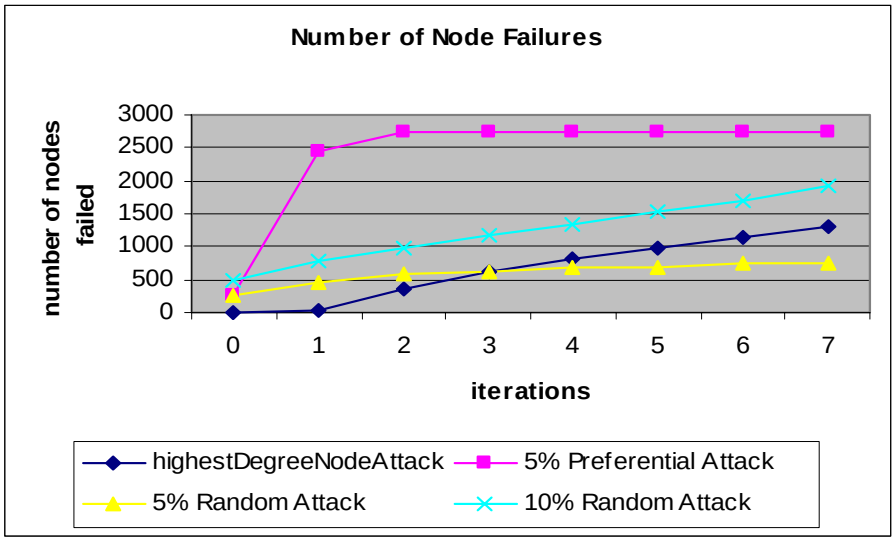
It was observed that the number of components increase drastically in case of targeted attack as compared to random failure. As we increase the percentage of rewiring, number of components decrease indicating increased connectivity in the network. Similarly, size of largest connected component (LCC) also grows with the percentage of edges rewired. Also the percolation point is reached slower when Addition Schemes are used and also gets slower with in increase in the % modification. It was also observed that Random Neighbor Rewiring Performs as well as the Addition Schemes.





It can be seen from the results that *Random 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.



The above graph 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.

The analysis of the performance of various edge modification schemes for the removal of the highest degree node shows that *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 lesser 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. The results for edge addition schemes show that when a larger number of nodes in the network 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 also do not perform well in case of *preferential attack* as compared to *random failure*. But it has been observed that at lower modification percentages the *rewiring schemes* are better than *addition schemes*. A high 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 than 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.

5. Effect of Routing Strategies

5.1. Introduction

The propagation of the node failure in the network depends both on the network structure as well as the routing strategy followed to route messages in the network. Different routing strategies choose different intermediate nodes to pass messages between the same end nodes. This leads to congestion at different nodes and hence causes their failure. We tried to study the cascading effect when random routes were used to communicate. It is import to study the random paths in the network because on absence of global information in the network, the node has to route packets based wholly on local knowledge. The basic routing strategy is to send the message to a random neighbor when no information is available. This is also called a random walk in the network. We also show the simulation results when partial global data is available. That is we follow the shortest path to route when we know it and we use a random path when the shortest path information is not available. We see how the network is affected as a function of deviation from following shortest paths.

5.2. Routing Model

We considered a model in which, at a time instant, some random source nodes try to communicate with random destination nodes by sending packets. This time instant is assigned for the transmission of these packets from source to destination. Each node in the network has a capacity. Two case were analyzed, one with all nodes having constant capacity and the second where the capacity of the node is proportional to its degree. A node is considered congested if the number of packets routed through this node at a time instant exceeds its capacity. Congested nodes are considered failed, since these nodes can't be used for routing immediately. As in the earlier analysis, we don't consider the recovery of nodes from the congested state. Therefore they are removed from the network. We continue to do the above for about 100 iterations and observe how the network is affected.

We first tried to identify the number of pairs of nodes allowed to communicate at a time instant by simulating the model with different values for this number and then identifying a value which stabilized the total node removals. The value obtained was 100, therefore 100 random pairs are allowed to communicate before the congested nodes are checked for and removed from the network. The routing scheme we considered for routing packets in the network is "random walk". Two variations of random walks were considered. We compared the results to those when shortest paths are used to route the packets. Also we tried to see how deviation from the shortest path routes to random routes affects the network, by considering deviation percentages of 20, 40, 60 & 80.

5.3. Random Walks

The basic definition of a random walk on a graph is as follows: given a graph and a starting point, we select a neighbor of it at random, and move to this neighbor; then we select a neighbor of this point at random, and move to it etc. The (random) sequence of points selected this way is a random walk on the graph. It has been shown that the more links a node has to other nodes in the network, the more often it will be visited by a random walker[9]. We try to prove this by calculating the betweenness centrality of the nodes based on the random walks[10]. Since the number of times a node is chosen is proportional to the degree, this random betweenness measure of a node should also be proportional to the degree.

We considered two variations of random walks. One is the random walk as defined above but with the restriction that a node visited once cannot be visited again when going from the source node to the destination. The other is to choose a random path from the source to destination from all the paths between two nodes. Here also we consider paths which contain a node only once. This is done by constructing a random spanning tree from the source node and then finding the path to the destination node in the spanning tree. This is also equivalent to random walk but it is not clear directly if the number of times a node is selected depends on its degree. We try to see if the two variations of the random walk are similar to each other in terms of there relation to the degree. Also we compare the betweenness of nodes when selected using these variations obtained from the simulations to the theoretical random walk where there is no condition on the number of times a node can appear in the path.

5.4. Cascading effect

We have simulated the routing model on the gnutella graph used for the previous analysis. We have used both the random walking variations and also the shortest path routes to see how many nodes are getting removed, and how the network is effected by these removals, ie., the size of the LCC and the number of components after every iteration. We also simulated the effect of the amount of global information present by applying the various percentages of deviation from the number of shortest paths used for routing. It was observed that the number of nodes removed while following random paths is higher than when routing using all shortest paths. The degree of nodes removed while following random paths was seen to be higher than when using shortest paths. Due to this the network disintegrated into small pieces much earlier when random paths are used. This shows that the high degree nodes (hub-like nodes) are selected in random walks more often and hence their removal causes more network disruption, as we saw in the earlier analysis of node removals.

We also tried to formulate the dependence of the degree of the node removed when random walks are chosen. The probability of a node to get congested and eventually removed from the network is shown to be proportional to the square of its degree. Therefore higher degree nodes tend to fail much more than the lower degree nodes. This explains why using random paths leads to high cascading effect compared to the shortest paths.

6. 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 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 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. The knowledge of how the various modification schemes affect the robustness of the network can be used to design better distributed network management protocols.

The effect of routing on the dynamics of the network has been studied. This gives us an insight into how different routing strategies can lead to congestion at different nodes. This knowledge is useful in selecting a suitable routing scheme, given a network topology, which leads to efficient network communication. Therefore, in cases where we have no control on the topology of the network or changing the topology is very costly, we can implement a routing strategy with increases the throughput of the network.

7. References

[1] A.Beygelzimer, G.Grinstein, R.Linsker and I.Rish - *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 Breakdown 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.

[9] J. D. Noh, H. Rieger - *Random Walks on Complex Networks*, Physics Review Letter 92, 118701.

[10] M.E.J Newman – *A measure of betweenness centrality based on random walk*, Social Networks, 2005- Elsevier.