Analyzing the Vulnerability of Superpeer Networks Against Attack

B. Mitra
Dept. of CSE
IIT Kharagpur, India
bivasm@cse.iitkgp.ernet.in

F. Peruani
ZIH, Technical University of Dresden, Germany
peruani@mpipks-dresden.mpg.de

S. Ghose, N. Ganguly
Dept. of CSE
IIT Kharagpur, India
{sujoy, niloy}@cse.iitkgp.ernet.in

ABSTRACT

In this paper, we develop an analytical framework to measure the vulnerability of superpeer networks against attack. Two different kinds of attacks namely deterministic and degree dependent attack have been introduced here. We formally model the superpeer networks with the help of bimodal structure and different attacks with the help of graph dynamics. Our analysis shows that fraction of superpeers and their connectivity have profound impact upon the stability of the network. The results obtained from the theoretical analysis are validated through simulation. The agreement between the simulation results and theoretical predictions is almost perfect.

Categories and Subject Descriptors: C.2.0 [General]: Security and protection

General Terms: Measurement, Security.

Keywords: Superpeer networks, attacks, complex network theory.

1. INTRODUCTION

The growing popularity of peer-to-peer networks makes them a very likely candidate for being a substrate for future internet scale information systems. In peer to peer networks, a huge number of peers are connected among themselves by some logical links forming an overlay above the physical network. Currently superpeer topologies have emerged as the most influencing topology among various overlay networks [12, 14]. Most of the commercial systems like KaZaA [1] have also adopted superpeers in their design. In this system, superpeer nodes with high bandwidth connect to each other forming the upper level in the network hierarchy. A large number of peers are connected with superpeers to get service from them.

Understanding the effect of attacks upon the large scale superpeer networks is becoming a major challenge for the p2p network community. The most prominent attack that affects the stability of the network is Denial Of Service (DoS) attack [11]. In the p2p networks, DoS drown important peers in fastidious computation so that they fail to provide any service requested by other peers. DoS attacks become far more effective when the attack is launched in distributed fashion, the feature more popularly termed as distributed denial-of-service (DDoS) attack. The perpetrator in DDoS remotely controls personal computers and directs attacks on important peers in the network through them. Worm propagation, file poisoning, sybil attack, eclipse attack are some of the important attacks that also affect the stability of the p2p networks [13].

A survey in the literature reveals that most of the commercial superpeer networks can be represented as large scale complex graphs. Attacks upon networks can also be modeled as different kinds of dynamics that take place in these complex graphs. Some analysis of graph dynamics have been done mainly by physicists. The effect of random failures and intentional attacks in various kinds of graphs are discussed by Cohen in [3, 4]. In [9], Newman *et al.* introduced the concept of generating function formalism. Using it, Callaway [2] found the exact analytic solutions for percolation[1] on random graphs with arbitrary degree distribution.

In this paper, we utilize many of the aforesaid results of percolation theory to develop a *generalized analytical framework* to measure the stability of superpeer networks against various kinds of attacks. *The attack is modeled in terms of removal of important nodes from the network.* We characterize the importance of a node mainly by its connectivity and bandwidth. Two different attack models are proposed. In the naive model, nodes are removed in the sequence of their degrees while in the second (sophisticated) model, every node has a probability of removal with probability tilted towards high degree nodes. The disintegration of *giant component* [8] helps us to measure the stability of the attacked network. We also perform simulation to validate the theoretical results.

The rest of the paper is organized as follows. In section 2, we develop an analytical framework to measure the stability of superpeer networks. Section 3 defines and models various environmental parameters like superpeer topology and attacks. Here we also explain the simulation environment generated to mimic large superpeer networks and specify

CCS'07, October 29–November 2, 2007, Alexandria, Virginia, USA.

[1]Percolation indicates the existence of a critical probability p_c such that below p_c the network is composed of isolated clusters but above p_c, a giant cluster spans the entire network (i.e. the network is almost fully connected).

how to measure the stability of the network. In section 4 we utilize the developed formalism to assess the stability of superpeer networks against attacks and validate the theoretical results with the help of simulations. Finally section 5 concludes the paper.

2. DEVELOPING ANALYTICAL FRAMEWORK USING GENERATING FUNCTION FORMALISM

In this section, we derive an analytical framework for measuring the stability of overlay structures undergoing any kind of disturbances in the network. With the help of this framework, we find the critical condition for break down of the connectivity of the network. We assume that we have an infinite system, and so before any failure or attack the biggest cluster size in the system is infinite. Theoretically the question that we want to answer is how severe should be the failure or attack to make the biggest cluster size in the system finite.

We start out by giving some definitions. Let p_k be the probability of finding a randomly chosen vertex with degree k. Let q_k be the probability that a node of degree k survives the failure or attack. Correspondingly $f_k = 1 - q_k$ is the probability that a node of degree k is removed. In our framework, p_k models the ensemble of overlay structures and f_k models the disruptive events that take place in the network. We are going to establish the relationship between stability and p_k and q_k i.e. $(1 - f_k)$ using the generating function formalism.

Generating function has been widely used to model various stochastic processes [9, 2]. A brief introduction of generating function follows.

Generating function: A generating function $G(x)$ is formally a power series of x which encodes some probability distribution. Let us assume that $G(x)$ generates the degree distribution of the network given by p_k, then the generating function takes the form

$$G(x) = \sum_{k=0}^{\infty} p_k x^k \tag{1}$$

The connection between the generating function and the probability distribution it generates is given by

$$p_k = \lim_{x \longrightarrow 0} \frac{1}{k!} \frac{d^k G(x)}{dx^k} \tag{2}$$

Another important property of generating functions is that the average of the index of the probability, i.e., for $G(x)$ the average degree z of a vertex, can be expressed simply by

$$z = \langle k \rangle = \sum_{k=0}^{\infty} k p_k = G'(1) \tag{3}$$

Using this formalism we can formulate the generating function $H_0(x)$ which generates the distribution of the component sizes to which a randomly selected *node* belongs to. Subsequently the average size of the components can be calculated from $H_0'(1)$. When this average component size becomes infinity, it indicates the emergence of giant component and hence we can derive the critical condition for the stability of the giant component. However to formulate $H_0(x)$, we have to use a set of generating functions that are specified below.

Some useful generating functions:

- $H_1(x)$ generates the distribution of the component sizes that are reached by choosing a *random edge* and following it to one of its ends.
- $F_1(x)$ generates the probability distribution of the outgoing edges of the *first neighbor* of a randomly chosen *node* after the process of removal of some portion of nodes is completed.
- $F_0(x)$ is the generating function associated with the probability of a *node* having degree k to be present in the network after the disruptive event.

Derivation of $F_0(x)$

The generating function $F_0(x)$ specifies the probability of finding a node of degree k to be present in the network after the failure or attack. Since $p_k q_k$ is the probability of finding a node of degree k to be present after the disruptive event, applying the definition of generating function (Eq. 1), we find that $F_0(x)$ takes the form

$$F_0(x) = \sum_{k=0}^{\infty} p_k q_k x^k \tag{4}$$

Derivation of $F_1(x)$

To reach the first neighbor of a randomly chosen node, we have to pick up one of its outgoing links randomly and follow it until we reach the other end. Hence the probability distribution generated by $F_1(x)$ is same as the probability distribution of the outgoing edges of a node reached by following a random edge. Therefore we derive the generating function $F_1(x)$, with the help of another generating function $A(x)$ which is based upon the probability of finding a randomly chosen edge connected to a node of degree k.

Derivation of $A(x)$:

If we think of an edge connecting two nodes i and j as actually two edges; one going from i to j, and another from j to i, then total number of such edges in the system becomes $\sum_{k=0}^{\infty} k n_k$, where n_k is the number of nodes with degree k in the system, which can be expressed as $n_k = N p_k$ (N being the total number of nodes in the system). The expected number of edges connected to nodes of degree k which remained present after the node removal event is $k n_k q_k$. So, the probability of finding a randomly chosen edge connected to a node of degree k becomes

$$\begin{aligned} p_{on}(k) &= \frac{k n_k q_k}{\sum_{k=0}^{\infty} k n_k} \\ &= \frac{k p_k q_k}{\sum_{k=0}^{\infty} k p_k} = \frac{k p_k q_k}{z} \end{aligned} \tag{5}$$

In consequence the generating function associated to the probability $p_{on}(k)$ is

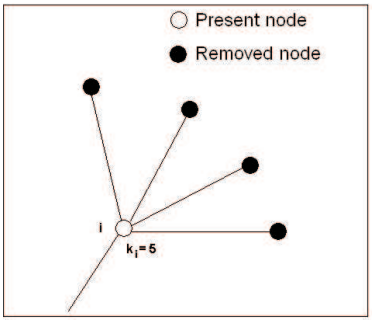
$$A(x) = \sum_{k=0}^{\infty} p_{on}(k) x^k = \sum_{k=0}^{\infty} \frac{k p_k q_k}{z} x^k$$

Since $\sum_{k=0}^{\infty} k p_k q_k x^k$ can be expressed as $x F_0'(x)$ therefore with the help of Eq. (3)

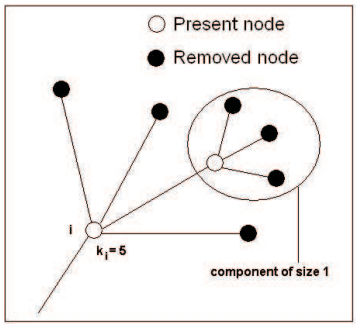
$$A(x) = x F_0'(x) / G'(1) \tag{6}$$

Derivation of $F_1(x)$:

The generating function $F_1(x)$ is based upon the probability distribution signifying the outgoing degree of a node reached



(a) Calculation of s_1



(b) Calculation of s_2

Figure 1: Schematic diagram explains the calculation of s_1 and s_2. White node indicates the node reached by following a random edge and black nodes indicate the removed nodes.

following a random edge. We know that a node having degree k arrived following a random edge has only $k-1$ outgoing links that leaves from that node. Hence probability of finding an existing node (that survives after the disruptive event) of $k-1$ outgoing edges reached following a random edge is $p_{on}(k) = \frac{kp_kq_k}{z}$ as defined in Eq. (5). Therefore probability distribution of the outgoing edges of the first neighbor of a randomly chosen node can be generated by

$$F_1(x) = \sum_{k=1}^{\infty} p_{on} x^{k-1} = \sum_{k=1}^{\infty} \frac{kp_kq_k}{z} x^{k-1} = F_0'(x)/z \qquad (7)$$

Derivation of $H_1(x)$

The function $H_1(x)$ generates the distribution of cluster sizes reached by following an edge chosen uniformly at random. Without loss of generality, we assume that following an edge, we can reach either a non existent node (node removed during deletion) or an existent node. The probability of following the randomly chosen edge and finding an existing/present node of degree zero is zero, the probability of finding an existing node of degree one is p_1q_1/z, the probability of finding an existing node of degree two is $2p_2q_2/z$, and so on. So the probability of finding a node following a random edge is $\sum_{k=0}^{\infty} kp_kq_k/z = F_1(1)$. In consequence, the probability of finding an edge that leads to a node which has been removed is $1-F_1(1)$. Clearly this is also the probability of following a randomly chosen edge that leads to a zero size component. Therefore if s_0 is the coefficient that accompanies x^0 in $H_1(x)$ then $s_0 = 1-F_1(1)$.

To find the full expression of $H_1(x)$ we have still to look for the probabilities that accompany non-zero size components. We find those probabilities next with the help of induction method.

Calculation of s_1, s_2 etc: We calculate the probability s_1 of finding by following a randomly chosen edge a component of size 1. This is nothing else than the sum of the probabilities of following an edge and finding a node of degree k which has its other $k-1$ edges connected to zero size components (all the nodes in these components are removed) (Fig. 1(a)). This can be expressed as:

$$\begin{aligned} s_1 &= \sum_{k=1}^{\infty} \frac{kp_kq_k}{z}(1-F_1(1))^{k-1} \\ &= F_1(H_1(0)) = \lim_{x\longrightarrow 0} \frac{1}{1!}\frac{d(s_0+xF_1(H_1(x)))}{dx} \end{aligned}$$

where $p_{on}(k) = kp_kq_k/z$ and $(1-F_1(1))^{k-1}$ is the probability of taking randomly $k-1$ edges and finding that all of them are attached to zero size components.

Knowing this we can easily calculate s_2, the probability of finding a component of size 2 by following a randomly chosen edge. s_2 is the sum of the probability of following a randomly chosen edge that leads to a node of degree k which is connected to $k-2$ zero size components, and has also an edge that leads to a component of size 1 (Fig. 1(b)). This can be expressed as

$$\begin{aligned} s_2 &= \sum_{k=2}^{\infty} \frac{(k-1)kp_kq_k}{z}(1-F_1(1))^{k-2}s_1 \\ &= F_1'(H_1(0))H_1'(0) = \lim_{x\longrightarrow 0} \frac{1}{2!}\frac{d^2(s_0+xF_1(H_1(x)))}{dx^2} \end{aligned}$$

where $(1-F_1(1))^{k-2}s_1$ is the probability of taking randomly $k-1$ edges and find that $k-2$ edges are attached to zero size components, and one to a size 1 component. The term $k-1$ in s_2 indicates that there are $k-1$ possible configurations for these edges.

Similarly we can find the probability of finding a component of size 3 by following a randomly chosen edge

$$s_3 = \lim_{x\longrightarrow 0} \frac{1}{3!}\frac{d^3(s_0+xF_1(H_1(x)))}{dx^3}$$

and so on. This suggests a self-consistence equation for $H_1(x)$ that generates the distribution of component sizes of nodes that are reached by randomly chosen edge after the disruptive event

$$\begin{aligned} H_1(x) &= s_0 + xF_1(H_1(x)) \\ &= 1 - F_1(1) + xF_1(H_1(x)) \qquad (8) \end{aligned}$$

It can be easily verified that Eq. (8) leads to the correct expressions of s_0, s_1,..., s_n by applying Eq. (2).

Derivation of $H_0(x)$
Along similar lines we can obtain the generating function $H_0(x)$ of the distribution of the component size to which a randomly chosen node belongs to. The probability that a randomly chosen node belongs to a component of size zero after the disruptive event is $1-F_0(1)$. Similarly the probability of a randomly chosen node to belong to some nonzero size component depends on the size of the components where all its first neighbors belong to. Hence the expression for $H_0(x)$ takes the form:

$$H_0(x) = (1 - F_0(1)) + xF_0(H_1(x)) \tag{9}$$

Finally from Eq. (9) and recalling the definition of average given by Eq. (3), we can obtain the average size of the components:

$$H_0'(1) = \langle s \rangle = F_0(1) + \frac{F_0'(1)F_1(1)}{1 - F_1'(1)} \tag{10}$$

As mentioned above, we are interested in knowing the threshold at which the average cluster size becomes infinite. Clearly Eq. (10) diverges when $1 - F_1'(1) = 0 \Rightarrow F_1'(1) = 1$, and this critical condition sets the threshold between finite and infinite cluster sizes[2]. Finally replacing $F_1'(1)$ by its definition (Eq. (7)), we obtain a critical condition for giant component formation

$$\sum_{k=0}^{\infty} kp_k(kq_k - q_k - 1) = 0 \tag{11}$$

The significance of the Eq. (11) lies in the fact that it states the critical condition for the stability of giant component with respect to any type of graphs (characterized by p_k) undergoing any type of failure or attack (characterized by q_k). Formulating this general formula is one of the primary contributions of the paper. Using this formalism, we investigate the stability situation of various superpeer networks.

3. ENVIRONMENT DEFINITION

In this section, we formally model the superpeer networks and different kinds of attacks. The stability of the network will be analytically derived based on the defined model. In theory, the stability is presented in terms of percolation threshold. To reproduce that effect in simulation, the stability metric is defined in detail.

3.1 Modeling superpeer networks

The different types of superpeer networks can be modeled using the uniform framework of probability distribution p_k, where p_k is the probability that a randomly chosen node has degree k. In this paper, we consider a simple model of superpeer networks - strict bimodal structure. We believe strict bimodal structure is simple enough to understand; at the same time it captures the essence of commercial superpeer networks. In strict bimodal structure, superpeer networks can be modeled by bimodal degree distribution where a large fraction (r) of peer nodes with small degree k_l are connected with superpeers and few superpeer nodes ($1 - r$) with high degree k_m are connected to each other. Therefore only two separate degrees are allowed in this kind of network. Formally

$$p_k > 0 \quad if \quad k = k_l, k_m; \qquad p_k = 0 \quad otherwise$$

k_l & k_m are degrees of peers and superpeers respectively. Therefore $p_{k_l} = r$ and $p_{k_m} = (1 - r)$.

3.2 Different kinds of attack models

The attack is modeled in terms of removal of nodes from the network. As defined in the previous section, let q_k be the probability that a vertex of degree k be present in the network after the removal of a fraction of nodes. In our framework q_k is used to specify two kinds of attacks which we consider in this paper namely deterministic attack and degree dependent attack. In the deterministic attack, superpeers are targeted before attacking any peer. The peers are attacked only after removal of all the superpeers from the network. In the degree dependent attack, both peers and superpeers are attacked simultaneously, but the probability of superpeers being attacked is much more than that of the peers. Each of the attack is quantitatively defined next.

- In the deterministic attack, the nodes having high degrees are progressively removed. Formally

  $q_k = 0$ when $k > k_{max}$
  $0 \leq q_k < 1$ when $k = k_{max}$
  $q_k = 1$ when $k < k_{max}$.

  This removes all the nodes from the network with degree greater than k_{max} and a fraction of nodes having degree k_{max}.

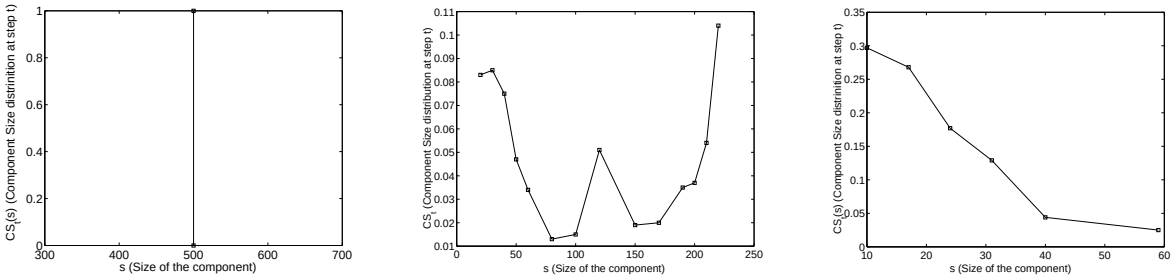
- In the degree dependent attack, the nodes having higher degrees are more likely to be removed. Therefore probability of removal of a node having degree k is proportional to k^γ where $\gamma \geq 0$ is a real number. With proper normalization $f_k = \frac{k^\gamma}{C}$ where C is a normalizing constant. Hence the fraction of nodes having degree k which survives after this kind of attack is $q_k = (1 - \frac{k^\gamma}{C})$.

3.3 Stability metric

The stability of superpeer networks are primarily measured in terms of a certain fraction of nodes (f_c) called percolation threshold [8], removal of which disintegrates the network into large number of small, disconnected components. Below that threshold, there exists a connected component which spans the entire network. This connected component is also termed as the giant component. The value of the percolation threshold f_c theoretically signifies the stability of the network, higher values indicate greater stability against attack.

We take cue from condensation theory used by physicists to develop the metric to measure the percolation threshold experimentally [5, 10]. During the experiment, we remove a fraction of nodes f_t from the network in step t and check whether we reach the percolation point. If not then in the next step $t + 1$ we remove $f_{t+1} = f_t + \epsilon$ fraction of nodes

[2]We present an intuitive explanation for this critical condition of giant component disruption. $F_1'(1)$ represents the average outgoing links of the first neighbor of a randomly chosen node. After the node removal process, if this average number of outgoing links is more than one, then the network should percolate, i.e. it is possible to find an infinite cluster of connected nodes. But if it is less than one, then it is very likely that by following a random edge, we land in a node that has no outgoing link and thus no chance of reaching another existing node.



(a) Initial component size distribution (only single giant component of size 500).

(b) Intermediate component size distribution

(c) Component size distribution at percolation point

Figure 2: The above plots represent the change in the component size distribution during percolation process and indicates the percolation point.

from the network and check again. This process is continued until we reach the percolation point. After each step, we find out the status of the network in terms of the number and size of the components formed. We collect the statistics of s and n_s where s denotes size of the components and n_s, number of components of size s and define the normalized component size distribution $CS_t(s) = sn_s / \sum_s sn_s$ at step t. We compute $CS_t(s)$ for all the steps starting from $t = 1$ and observe the behavior of $CS_t(s)$ after each step (Fig. 2). Initially the $CS_t(s)$ shows unimodal character confirming a single connected component (Fig. 2(a)) or bimodal character (Fig. 2(b)) confirming a large component along with a set of small components. As the fraction of nodes removed from the network increases gradually, the network disintegrates into several components. This leads to the change in the behavior of $CS_t(s)$ whereby at a particular step t_n, $CS_{t_n}(s)$ becomes monotonically decreasing function indicating t_n as the percolation point (Fig. 2(c)). Therefore t_n is considered as the time step where percolation occurs and the total fraction of nodes removed at that step f_{t_n} specifies the percolation threshold.

3.4 Network Generation

In our simulation, the superpeer network is represented by a simple undirected graph. In order to generate the topology, every node is assigned a degree according to the bimodal degree distribution. Thereafter the edges are generated using the "matching method" [7]. Some of the edges are then rewired using "switching method" to generate sufficient randomness in the graph [6]. In our experiment, we simulate the superpeer networks by generating graphs with 5000 nodes.

4. STABILITY OF SUPERPEER NETWORKS AGAINST ATTACK

In this section, we formally analyze the effect of attacks on the superpeer networks. The theoretically derived results are verified with the help of simulation. The two kinds of attacks, namely deterministic attack and degree dependent attack are discussed separately.

4.1 Stability analysis against deterministic attack

Two cases may arise in the deterministic attack

Case 1 The removal of a fraction of superpeers is sufficient to disintegrate the network.

Case 2 The removal of all the superpeers is not sufficient to disintegrate the network. Therefore we need to remove some of the peer nodes along with the superpeers.

We analyze these two cases separately with the help of our analytical framework. From Eq. (11) the critical condition for the stability of the superpeer networks can be rewritten as

$$\sum_{k=k_l,k_m} k(k-1)p_k q_k = \langle k \rangle$$

The equation can be further expanded as below to differentiate between peers and superpeers

$$\sum_{k=k_l} k(k-1)p_k q_k + \sum_{k=k_m} k(k-1)p_k q_k = \langle k \rangle \qquad (12)$$

Case 1: In this case, removal of a fraction of superpeers is sufficient to disintegrate the network. If f_{sp} is the critical fraction of superpeer nodes, removal of which disintegrates the giant component then $q_k = 1$ for $k = k_l$ and $q_k = 1 - f_{sp}$ for $k = k_m$. Hence according to Eq. (12),

$$\sum_{k=k_l} k(k-1)p_k + \sum_{k=k_m} k(k-1)p_k(1-f_{sp}) = \langle k \rangle$$

$$\Rightarrow f_{sp} = 1 - \frac{\langle k \rangle - k_l(k_l-1)p_{k_l}}{k_m(k_m-1)p_{k_m}}$$

As the fraction of superpeer nodes in the network is $(1-r)$, then percolation threshold for case 1 becomes $f_{tar} = (1-r) \times f_{sp}$

$$\Rightarrow f_{tar} = (1-r)\left(1 - \frac{\langle k \rangle - k_l(k_l-1)r}{k_m(k_m-1)(1-r)}\right) \qquad (13)$$

Case 2: Here we have to remove f_p fraction of peer nodes along with all the superpeers to breakdown the network.

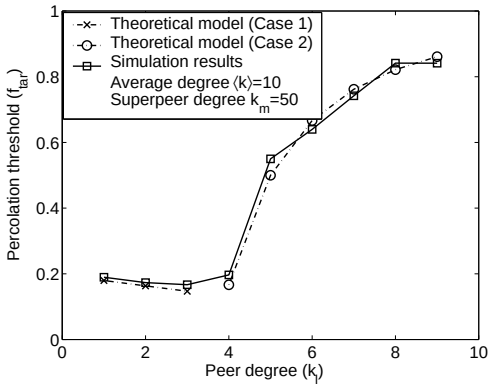


Figure 3: Stability of the superpeer networks in face of deterministic attack (Comparative study between theoretical and simulation results).

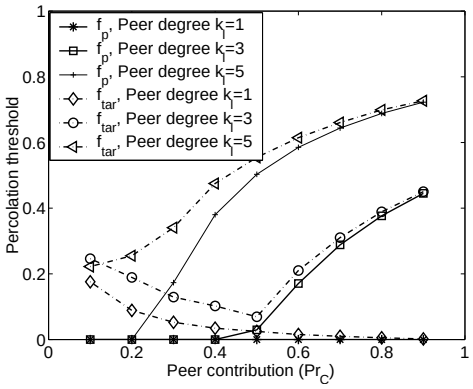


Figure 4: The plot represents the impact of peer contribution Pr_C upon the stability of the network against deterministic attack.

Therefore $q_k = 1 - f_p$ for $k = k_l$ and $q_k = 0$ for $k = k_m$. Hence according to Eq. (12),

$$\sum_{k=k_l} k(k-1)p_k(1-f_p) = \langle k \rangle$$

$$\Rightarrow f_p = 1 - \frac{\langle k \rangle}{k_l(k_l-1)p_{k_l}}$$

Therefore the total fraction of nodes required to be removed to disintegrate the network for case 2 becomes $f_{tar} = rf_p + (1-r)$.

$$\Rightarrow f_{tar} = r\left(1 - \frac{\langle k \rangle}{k_l(k_l-1)r}\right) + (1-r) \qquad (14)$$

Transition point: The transition from case 1 to case 2 can be easily marked by observing the value of the percolation threshold f_{tar}. While calculating using Eq. (13) (case 1), if the value of f_{tar} exceeds the fraction of superpeers in the network $(1-r)$, it indicates that removal of all the superpeers is not sufficient to disrupt the network. Hence subsequently we enter into case 2 and start using Eq. (14) to find the percolation threshold.

During our simulation, initially only high degree superpeer nodes in the network are removed gradually until the percolation point is reached. If the percolation point is not reached even after removing of all the superpeers, we remove a fraction of peers along with the superpeers to breakdown the network. We perform each experiment for 500 times and take the average of the percolation threshold obtained in each of them. Superpeer networks with average degree $\langle k \rangle = 10$ and superpeer degree $k_m = 50$ are considered for case study. We increase the peer degree k_l gradually (the peer fraction changes accordingly) and observe the change in the percolation threshold f_{tar} (Fig. 3).

Observations:

a. In networks with peer degree $k_l = 1, 2$ and 3, the removal of only a fraction of superpeers causes breakdown of the network. Moreover, the increase of peer degree from 1 to 2 and 3 further reduces the fraction of superpeers in the network which results networks with $k_l = 2, 3$ more vulnerable. Normal wisdom would expect the attack vulnerability of the network to decrease with the decrease of fraction of superpeers. But the opposite happens here. The reason is in this zone (at $k_l = 2, 3$), although peers have a larger share in the network, yet it is not large enough to form effective connections within themselves. Therefore the stability of the network is still entirely dependent on the high degree superpeers, hence now attacking even a smaller fraction breaks down the network.

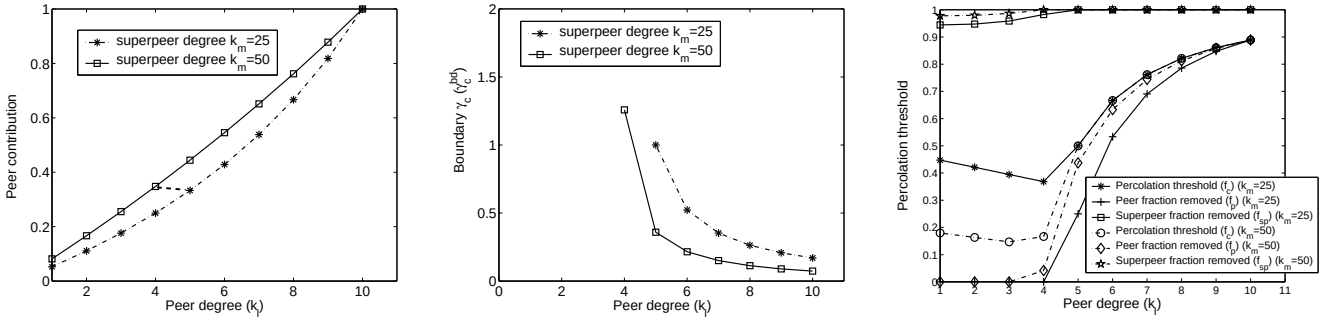
b. However as peer degree increases beyond 4, a fraction of peers is required to be removed even after removal of all the superpeers to dissolve the network. This is because the high degree peers connect among themselves and they are not entirely dependent on superpeers for connectivity. This results in the steep increase of stability of the network with peer degree $k_l \geq 5$.

Peer contribution: In addition to peer degree, we study the stability of the network with respect to a new metric namely 'peer contribution'. The peer contribution controls the total bandwidth contributed by the peers which determines the amount of influence superpeer nodes exerts on the network. The peer contribution Pr_C is defined by two parameters - peer degree and fraction of peers in the network. Hence $Pr_C = \frac{rk_l}{\langle k \rangle}$ where $\langle k \rangle = rk_l + (1-r)k_m$. We generate three set of networks having peer degree $k_l = 1, 3$ and 5 respectively for individual peer contribution Pr_C $(0.1 \leq Pr_C \leq 0.9)$. In order to do that, we choose fraction of peers r uniformly at random and adjust superpeer degree k_m accordingly to keep the peer contribution Pr_C and peer degree k_l constant. This procedure is followed to generate one hundred networks for each set. We restrict superpeer degree $k_m \geq 20$ in order to generate realistic superpeer networks. We theoretically compute the fraction of peers required to be removed f_p and percolation threshold f_{tar} for individual network and calculate their average for individual k_l. This expected fraction of peers required to be removed f_p and percolation threshold f_{tar} is plotted with respect to the peer contribution Pr_C (Fig. 4).

Observations:

a. It can be observed from Fig. 4 that superpeer networks having peer degree $k_l = 1$ can be disintegrated without attacking peers at all for any peer contribution Pr_C. This kind of attack belongs to case 1 of the deterministic attack.

b. The peers of the superpeer networks having peer con-



(a) Behavior of peer contribution Pr_C due to the change in peer degree k_l.

(b) Behavior of boundary critical exponent due to the change in peer degree.

(c) Fraction of peers and superpeers required to be removed to breakdown the network and its impact upon percolation threshold f_c.

Figure 5: The above plots illustrate the impact of peer degree upon the stability of superpeer networks in face of degree dependent attack. The fraction of peers r is adjusted with peer degree k_l to keep the average degree $\langle k \rangle$ fixed.

tribution $Pr_C \leq 0.2$ does not have any impact upon the stability of the network. This is true for low as well as high degree peers.

c. The influence of high degree peers increases with the increase of peer contribution. At $Pr_C = 0.3$, a fraction of peers is required to be removed to disintegrate the networks having peer degree $k_l = 5$. The impact of high degree peers upon the stability of the network becomes more eminent as peer contribution $Pr_C \geq 0.5$. In this region, a significant fraction of peers is required to be removed for all the networks having peer degree $k_l = 3, 5$. This kind of attack belongs to case 2 of the deterministic attack.

d. Increase in peer contribution $Pr_C \geq 0.4$ brings the percolation threshold f_{tar} and fraction of peers needed to be attacked f_p close to each other which implies that stability of these networks is primarily dependent upon the stability of the peers.

e. It is interesting to observe that peer contribution Pr_C has two opposite effects upon stability of the networks depending on the peer degree k_l. The percolation threshold f_{tar} increases with peer contribution Pr_C for $k_l = 3, 5$, but gradually reduces for $k_l = 1$. The reason behind this is, stability of the networks with peer degree $k_l = 1$ is entirely dependent upon superpeers. Since increase in peer contribution decreases superpeer contribution, it decreases stability of these networks also. On the other hand, peers having degree $k_l \geq 3$ are strongly connected among themselves, hence stability of these networks is more dependent upon peer contribution. Hence percolation threshold f_{tar} increases with peer contribution Pr_C.

4.2 Stability analysis against degree dependent attack

In this kind of attack, probability of removal of a node of degree k is directly proportional to k^γ where $\gamma \geq 0$ is a real number. Hence with proper normalization, $f_k = \frac{k^\gamma}{C}$ where C is the normalizing constant. Therefore probability of survival of a node having degree k after a degree dependent attack is

$$q_k = 1 - \frac{k^\gamma}{C}$$

As mentioned in bimodal degree distribution, let r be the fraction of peers with degree k_l and rest be superpeers of degree k_m. If $\langle k \rangle$ is the average degree of the network then

$$p_{k_l} = r = \frac{k_m - \langle k \rangle}{k_m - k_l} \qquad p_{k_m} = (1 - r) = \frac{\langle k \rangle - k_l}{k_m - k_l}$$

From Eq. (11) the critical condition for the stability of the giant component can be rewritten as

$$\sum_{k=k_l,k_m} k(k-1)p_k q_k = \langle k \rangle$$

$$\begin{aligned} &\Rightarrow \quad \langle k^{\gamma+2} \rangle - \langle k^{\gamma+1} \rangle = C(\langle k^2 \rangle - 2\langle k \rangle) \\ &\Rightarrow \quad r k_l^{\gamma+1}(k_l - 1) + (1-r)k_m^{\gamma+1}(k_m - 1) = \\ &\qquad C(\langle k \rangle (k_m + k_l) - k_m - 2\langle k \rangle) \end{aligned} \tag{15}$$

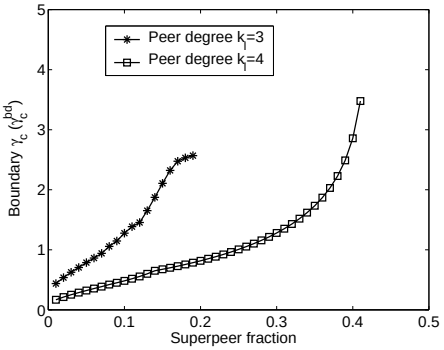
where θ^{th} moment of the bimodal degree distribution can be written as $\langle k^\theta \rangle = k_m^\theta p_{k_m} + k_l^\theta p_{k_l}$. The solution of the Eq. (15) yields a particular value of γ, say γ_c (termed as critical exponent) and the percolation threshold becomes

$$f_c = r\frac{k_l^{\gamma_c}}{C} + (1-r)\frac{k_m^{\gamma_c}}{C} \tag{16}$$

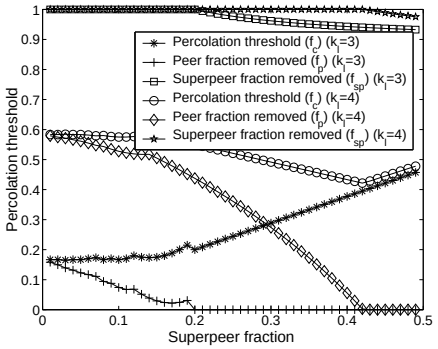
In order to evaluate the disintegration point, proper assignment of the value of normalizing constant C is necessary. Since f_k should be ≤ 1 $\forall k$, hence the minimum value of $C = k_m^\gamma$. Assuming this condition, the Eq. (15) becomes

$$\begin{aligned} r k_l^{\gamma+1}(k_l - 1) + (1-r)k_m^{\gamma+1}(k_m - 1) \quad &\geq \\ k_m^\gamma(\langle k \rangle (k_m + k_l) - k_m - 2\langle k \rangle) \end{aligned} \tag{17}$$

The solution set of the above inequality (say S_{γ_c}) can be bounded (where $0 \leq \gamma_c \leq \gamma_c^{bd}$) or unbounded (where $0 \leq \gamma_c \leq +\infty$). Each critical exponent $\gamma_c \in S_{\gamma_c}$ specifies the fraction of peers and superpeers required to be removed to



(a) Behavior of γ_c^{bd} with respect to the change in superpeer fraction.



(b) Fraction of peers and superpeers required to be removed to breakdown the network and its impact upon percolation threshold f_c.

Figure 6: The above plots mainly illustrate case 1 of degree dependent attack. The superpeer degree k_m is adjusted with the change of superpeer fraction to keep the average degree fixed.

breakdown the network. Assuming equality of Eq. (17) and hence obtaining minimum value of C, each γ_c results in the corresponding normalizing constant

$$C_{\gamma_c} = \frac{rk_l^{\gamma_c+1}(k_l-1) + (1-r)k_m^{\gamma_c+1}(k_m-1)}{\langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle} \qquad (18)$$

Hence the fraction of peers and superpeers needed to be attacked

$$f_p^{\gamma_c} = \frac{k_l^{\gamma_c}}{C_{\gamma_c}} \qquad f_{sp}^{\gamma_c} = \frac{k_m^{\gamma_c}}{C_{\gamma_c}} \qquad (19)$$

respectively and the total fraction of nodes removed $f_c^{\gamma_c}$ is obtained from Eq. (16). The $f_c^{\gamma_c}$ depends upon the critical exponent $\gamma_c \in S_{\gamma_c}$ and normalizing constant C_{γ_c}. The nature of the solution set S_{γ_c} has profound impact upon the behavior of $f_p^{\gamma_c}$, $f_{sp}^{\gamma_c}$ and as well as $f_c^{\gamma_c}$. The breakdown of the network can be due to one of the three situations noted below

1. The removal of all the superpeers along with a fraction of peers.
2. The removal of only a fraction of superpeers.
3. The removal of some fraction of both superpeers and peers.

Each situation arises due to the following reasons
1: Networks having a bounded solution set S_{γ_c} where $0 \leq \gamma_c \leq \gamma_c^{bd}$ exhibit this kind of behavior at the maximum value of the solution $\gamma_c = \gamma_c^{bd}$. Here the fraction of superpeers removed becomes $f_{sp}^{\gamma_c^{bd}} = 1$ and fraction of peers removed $f_p^{\gamma_c^{bd}} = \frac{k_l^{\gamma_c^{bd}}}{C_{\gamma_c^{bd}}}$.
2: Some networks have an open solution set S_{γ_c} where $0 \leq \gamma_c \leq +\infty$. At $\gamma_c \to \infty$, $f_p^{\gamma_c}$ converges to 0 and $f_{sp}^{\gamma_c}$ converges to some x where $0 < x < 1$.
3: Intermediate critical exponents $\gamma_c \in S_{\gamma_c}$ signifies the fractional removal of both peers and superpeers.

One of the major contributions of this paper is to provide an uniform attack framework, which is able to successfully capture all the important features of the deterministic attack[3] as well as provide flexibility to attack the nodes in a more suitable way. Next we perform a theoretical case study that justifies our claim. We show the impact of peer degrees upon the stability of the networks against the degree dependent attack (Fig. 5). We consider superpeer networks with two superpeer degrees $k_m = 25, 50$ and fixed average degree $\langle k\rangle = 10$. The change in peer contribution Pr_C due to the change of peer degree k_l is shown in Fig. 5(a). In Fig 5(b), we show the behavior of boundary critical exponent γ_c^{bd} with the change of peer degree. The nonexistence of γ_c^{bd} implies the unbounded solution set for that particular network. Fig. 5(c) describes the fraction of peers and superpeers required to be removed (f_p and f_{sp} respectively) as well as percolation threshold (f_c) for various peer degrees.
Observations:
a. The peer contribution Pr_C increases with superpeer degree k_m for a particular peer degree k_l (Fig. 5(a)). In order to keep the average degree and peer degree constant, the network with higher superpeer degree results higher fraction of peers which increases the peer contribution.
b. Fig. 5(b) shows that the solution set of inequality (17) remains unbounded for the networks having peer degree $k_l \leq 4$ with superpeer degree $k_m = 25$ and peer degree $k_l \leq 3$ with superpeer degree $k_m = 50$. This implies that removal of only a fraction of superpeers disintegrate these networks (Fig. 5(c)). This kind of attack resembles case 1 of the deterministic attack. Apart from the mathematical conclusion, we explain the phenomena from the point of relative contribution of peers and superpeers in the network. The low peer degree results in low peer contribution (Fig. 5(a)) and high superpeer contribution. Hence removal of only a fraction of superpeers is sufficient to breakdown these networks.

[3]Case 1 and case 2 of the degree dependent attack resembles qualitatively as well as quantitatively with the case 2 and case 1 of the deterministic attack respectively.

c. The gradual increase in peer degree increases the peer contribution and at $k_l = 5$ ($k_m = 25$), the high peer contribution ensures the necessity to remove a fraction of them to breakdown the network (Fig. 5(c)). This kind of attack resembles case 2 of the deterministic attack. The inequality (17) gets a bounded solution set for the network (Fig. 5(b)). Same thing happens for network with superpeer degree $k_m = 50$ at peer degree $k_l = 4$. Note that the peer contributions get almost same values for these two networks (Fig. 5(a)). This ensures that peer contribution has profound impact upon the stability of the network specially with the networks having high peer degree k_l.

Next we illustrate the different categories of degree dependent attack with the help of individual case study.

Case study 1: First we consider superpeer networks with peer degrees $k_l = 3, 4$ average degree $\langle k \rangle = 5$ and theoretically study the stability of the networks due to the change in the fraction of superpeers. The results of the case study are noted in Fig. 6. It can be observed that the solution set of these networks upto a threshold superpeer fraction sp^{th}, ($sp^{th} = 0.19$ and 0.41 for $k_l = 3$ and $k_l = 4$ respectively) remains bounded and the behavior of the boundary critical exponent γ_c^{bd} due to the change of fraction of superpeers is shown in Fig. 6(a). The fraction of superpeers and peers needed to be attacked for these networks is presented in Fig. 6(b). These networks exhibit the properties of case 1 of degree dependent attack hence the removal of all the superpeers is necessary to disintegrate the network along with a fraction of peers. Fig. 6(b) also represents some instances of case 2 where only some fraction of superpeers are needed to be removed.

Observations:

a. Impact upon the fraction of peers removed

The increase of superpeer fraction slowly increases γ_c^{bd} (Fig. 6(a)) which in turn gradually decreases the fraction of peers removed $f_p^{\gamma_c^{bd}}$ (Fig. 6(b)). The amount of removal of peers also depends upon the peer degree k_l. The high degree peers strongly connect among themselves to enhance their influence upon the stability, hence $f_p^{\gamma_c^{bd}}$ gets higher values for these networks.

b. Impact upon percolation threshold

Let the percolation threshold for the networks having superpeer fraction $sp_1(= 1 - r_1)$ and $sp_2(= 1 - r_2)$ (where $sp_1 < sp_2$) be $f_{c1}^{\gamma_c^{bd}}$ and $f_{c2}^{\gamma_c^{bd}}$ respectively. Hence the percolation threshold for these two networks are

$$f_{c1_{(2)}}^{\gamma_c^{bd}} = r_{1_{(2)}} f_{p1_{(2)}}^{\gamma_c^{bd}} + (1 - r_{1_{(2)}})$$

Therefore the change in the percolation threshold when the superpeer fraction changes from sp_1 to sp_2 is

$$\begin{aligned} \triangle f_c^{\gamma_c^{bd}} &= r_1 f_{p1}^{\gamma_c^{bd}} - r_2 f_{p2}^{\gamma_c^{bd}} + ((1 - r_1) - (1 - r_2)) \\ &= \triangle \left(r f_p^{\gamma_c^{bd}} \right) + \triangle(1 - r) \end{aligned} \quad (20)$$

The Eq. (20) shows that the change of percolation threshold $f_c^{\gamma_c^{bd}}$ becomes influenced by two opposite forces; in one hand the increase of superpeer fraction in the network makes $\triangle sp = \triangle(1 - r) < 0$. On the other hand, the fraction of peers in the network as well as fraction of them required to be removed decreases (Fig. 6(b)) which makes $\triangle \left(r f_p^{\gamma_c^{bd}} \right) > 0$. Depending upon the weightage of influence, $f_c^{\gamma_c^{bd}}$ either decreases ($k_l = 3$) or increases ($k_l = 4$) slowly when the fraction of superpeers is less than sp^{th}.

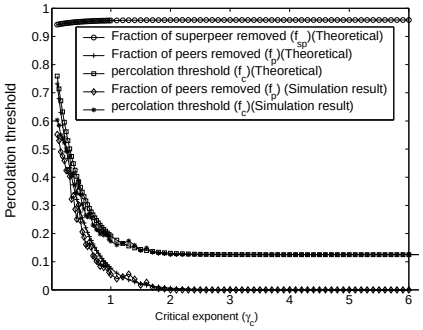


Figure 7: The above plot illustrates the case 2 of degree dependent attack.

Case study 2: We illustrate the degree dependent attack where removal of only a fraction of superpeers is sufficient to disintegrate the network. The case study is performed with a network having superpeer degree $k_m = 25$, average degree $\langle k \rangle = 5$ and peer degree $k_l = 2$ and results are validated with the help of simulation. The unbounded solution set S_{γ_c} of the network signifies that it belongs to case 2 of the degree dependent attack. We plot the theoretically calculated (Eq. (18), (19)) fraction of peers and superpeers required to be removed to breakdown the network for each critical exponent γ_c (Fig. 7). In simulation, we initially remove a fraction of superpeers $f_{sp}^{\gamma_c}$ (calculated theoretically) and then start removing peers gradually to breakdown the network. The minimum peer fraction, removal of which causes the breakdown of the network produces simulated $f_p^{\gamma_c}$. We perform the simulation on graphs of 5000 nodes and repeat each experiment for 500 times and take the average of the removed peer fraction. We compare simulated result with theoretically calculated $f_p^{\gamma_c}$ (Fig. 7).

Observations:

a. The fraction of peers removed gradually decreases with the increase of critical exponent γ_c which in turn decreases the value of $f_c^{\gamma_c}$. As $\gamma_c \to \infty$, the $f_p^{\gamma_c} \to 0$ with $f_{sp}^{\gamma_c} \to x$ (where $0 < x < 1$) and eventually $f_{sp}^{\gamma_c}$, $f_c^{\gamma_c}$ both reach some steady value. It signifies that the removal of only a fraction of superpeers is sufficient to breakdown the network (Fig. 7).

b. The nonexistence of the boundary critical exponent γ_c^{bd} for the networks having more than the threshold superpeer fraction sp^{th}, signifies that the solution set of these networks is unbounded and that the percolation process belongs to case 2 (Fig 6(a)). It can be observed that fraction of peers required to be removed for these networks becomes zero (Fig. 6(b)) and removal of only a fraction of superpeers disintegrates the network.

c. It is important to note that removal of only a fraction of superpeers is sufficient to disintegrate any network with peer degree $k_l = 1, 2$, irrespective of superpeer degree and its fraction. Mathematically it can be explained as follows

1. The inequality (17) becomes independent of γ as $k_l = 1$. Hence at $k_l = 1$, the solution set S_{γ_c} becomes unbounded which implies that removal of only a fraction of superpeers causes breakdown of the network.

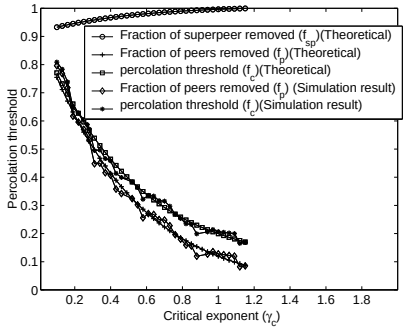


Figure 8: The above plot illustrates the case 3 of the degree dependent attack.

2. For $k_l = 2$, $2k_l \geq k_l^2$

$$\Rightarrow \quad 2rk_l \geq rk_l^2$$
$$\Rightarrow \quad (1-r)k_m + 2rk_l - rk_l^2 \geq 0$$
$$\Rightarrow \quad (1-r)k_m(k_m-1) \geq \langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle$$
$$\Rightarrow \quad rk_l^{\gamma+1}(k_l-1) + (1-r)k_m^{\gamma+1}(k_m-1) \geq k_m^{\gamma}(\langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle) \qquad (21)$$

Since the above inequality holds for any values of γ, it indicates that any network with $k_l = 2$ has unbounded solution set.

Case study 3: Degree dependent attack also allows to disintegrate the network by removing a fraction of both peers and superpeers (designated as case 3 of degree dependent attack). We investigate the amount of peers and superpeers needed to be removed to dissolve the network due to the change in γ_c. We deduce the results for a network having superpeer degree $k_m = 25$, average degree $\langle k\rangle = 5$ and peer degree $k_l = 3$ and results are validated with the help of simulation (Fig. 8). The simulation set up is same as described for case 2 of the degree dependent attack.

Observations:

a. Our analytical results show that this network has bounded solution set S_{γ_c} of the inequality (17) and all the critical exponents γ_c less than the boundary critical exponent $\gamma_c^{bd} = 1.171$ results in this kind of breakdown. It is evident from both theoretical and simulation results that the removal of any combination of $(f_p^{\gamma_c}, f_{sp}^{\gamma_c})$ (obtained from the curves in Fig. 8) where $0 \leq \gamma_c < \gamma_c^{bd}$, results in the breakdown of the network.

b. Networks with unbounded solution set (Fig 7) has finite values of γ_c (here $\gamma_c < 1.5$ (approx)) where the removal of both fraction of peers and superpeers are necessary to disintegrate the network.

5. CONCLUSION

In this paper, we have developed an analytical framework to measure the stability of the superpeer networks against attack. The wide range of attacks have been modeled in two different ways, deterministic as well as in a more general degree dependent manner. We have applied those models in our analytical framework to measure the stability of superpeers networks. In the deterministic attack, networks having peer degree $k_l \leq 3$ are very much vulnerable and removal of only a fraction of superpeers causes the breakdown the network. But as the peer degree increases, the stability of the network increases as well. In the degree dependent attack we have formulated a critical condition whose solution set provides the critical exponent γ_c. The peers and superpeers required to be removed is dependent upon this critical exponent γ_c and normalizing constant C_{γ_c}. The degree dependent attack model provides us with a more general scenario where various situations can be obtained only by changing the parameter γ. We believe that this can be further extended to understand the effect of churn in the network as well as combined function of both churn and attack.

6. REFERENCES

[1] Kazaa website. http://www.kazaa.com.

[2] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review E*, 85(21), 2000.

[3] R. Cohen, K. Erez, D. Avraham, and S. Havlin. Resilience of the internet to random breakdown. *Physical Review Letters*, 85(21), 2000.

[4] R. Cohen, K. Erez, D. Avraham, and S. Havlin. Resilience of the internet under intentional attack. *Physical Review Letters*, 86(16), 2001.

[5] S. N. Majumdar, M. R. Evans, and R. K. P. Zia. Nature of the condensate in mass transport models. *Physical Review Letters*, 94(180601), 2005.

[6] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *eprint arXiv:cond-mat/0312028*, 2003.

[7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs : Simple building blocks of complex networks. *Science*, 298, 2002.

[8] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7, 1998.

[9] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their application. *Physical Review E*, 2001.

[10] F. Peruani, A. Deutsch, and M. Baer. Nonequilibrium clustering of self-propelled rods. *Physical Review E*, 74(030904(R), 2006.

[11] B. Pretre. Attacks on peer-to-peer networks. In *Ph.D thesis*. Swiss Federal Institute of Technology (ETH), Zurich, 2005.

[12] Y. J. Pyun and D. S. Reeves. Constructing a balanced, log(n)-diameter super-peer topology. In *Proceedings of the* 4^{th} *International Conference on Peer-to-Peer Computing*. Zurich, Switzerland, August 2004.

[13] J. Saia. Attack-resistant peer-to-peer networks. In *NIPS 2003 Workshop on Robust Communication Dynamics in Complex Networks*. Whistler, Canada, December 12-13 2003.

[14] B. Yang and H. Garcua-Molina. Designing a super-peer network. In *Proceedings of the International Conference on Data Engineering (ICDE)*. Los Alamitos, CA, March 2003.