Static and Dynamic Analysis of the Internet's Susceptibility to Faults and Attacks

Seung-Taek Park[1]   Alexy Khrabrov[2]   David M. Pennock[2]   Steve Lawrence[2]

C. Lee Giles[1,2,3]   Lyle H. Ungar[4]

[1]Department of Computer Science and Engineering
[3]School of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802 USA
separk@cse.psu.edu, giles@ist.psu.edu

[2]NEC Research Institute
4 Independence Way
Princeton, NJ 08540 USA
{alexy,dpennock,lawrence}@research.nj.nec.com

[4]Department of Computer and Information Science
University of Pennsylvania, 566 Moore Building, 200 S. 33rd St
Philadelphia, PA 19104
ungar@cis.upenn.edu

Abstract— We analyze the susceptibility of the Internet to random faults, malicious attacks, and mixtures of faults and attacks. We analyze actual Internet data, as well as simulated data created with network models. The network models generalize previous research, and allow generation of graphs ranging from uniform to preferential, and from static to dynamic. We introduce new metrics for analyzing the connectivity and performance of networks which improve upon metrics used in earlier research. Previous research has shown that preferential networks like the Internet are more robust to random failures compared to uniform networks. We find that preferential networks, including the Internet, are more robust only when more than 95% of failures are random faults, and robustness is measured with the average diameter. The advantage of preferential networks disappears with alternative metrics, and when a small fraction of faults are attacks. We also identify dynamic characteristics of the Internet which can be used to create improved network models, allowing more accurate analysis for the future Internet, for example facilitating the design of network protocols with optimal performance in the future, or predicting future attack and fault tolerance. We find that the Internet has been becoming more preferential as it evolves. The average diameter has been stable or even decreasing as the number of nodes has been increasing. The Internet has been becoming more robust to random failures over time, but has also become more vulnerable to attacks.

Method Keywords— Simulation, Network measurement, Graph theory, Statistics

Index Terms— fault, attack, robustness, scale-free network, exponential network, evolving network, Internet, node failure, network topology, network models

I. Introduction

Many biological and social mechanisms—from Internet communications [1] to human sexual contacts [2]—

can be modeled using the mathematics of networks. Depending on the context, policymakers may seek to impair a network (e.g., to control the spread of a computer or bacterial virus) or to protect it (e.g., to minimize the Internet's susceptibility to distributed denial-of-service attacks). Thus a key characteristic to understand in a network is its robustness against failures and intervention. As networks like the Internet grow, random failures and malicious attacks can cause damage on a proportionally larger scale – an attack on the single most connected hub can degrade the performance of the network as a whole, or severe millions of connections. With the ever increasing terrorist threat, attack and fault tolerance becomes an important factor in planning network topologies and strategies for sustainable performance and damage recovery.

A network consists of nodes and links (or edges), which often are damaged and repaired during the lifetime of the network. Damage can be complete or partial, causing nodes and/or links to malfunction, or to be fully destroyed. As a result of damage to components, the network as a whole deteriorates: first, its performance degrades, and then it fails to perform its functions as a whole. Measurements of performance degradation and the threshold of total disintegration depend on the specific role of the network and its components. Using random graph terminology [3], disintegration can be seen as a phase transition from degradation – when degrading performance crosses a threshold beyond which the quality of service becomes unacceptable.

Network models can be divided into two categories according to their generation methods: static and evolving (growing) [4]. In a static network model, the total number of nodes and edges are fixed and known in advance, while in an evolving network model, nodes and links are

added over time. Since many real networks such as the Internet are growing networks, we use two general growing models for comparison – growing exponential (random) networks, which we refer to as the GE model, where all nodes have roughly the same probability to gain new links, and growing preferential (scale-free) networks, which we refer to as the Barabási-Albert (BA) model, where nodes with more links are more likely to receive new links. Note that [5] used two general network models, a static random network and a growing preferential network.

For our study, we extend the modeling space to a continuum of network models with *seniority*, adding another dimension in addition to the *uniform* to *preferential* dimension. We extend the simulated failure space to include mixed sequences of failures, where each failure corresponds to either a fault or an attack. In previous research, failure sequences consisted either solely of faults or attacks; we vary the percentage of attacks in a fault/attack mix via a new parameter β which allows us to simulate more typical scenarios where nature is somewhat malicious, e.g., with $\beta \approx 0.1$ (10% attacks).

We analyze both static and dynamic susceptibility of the Internet to faults and attacks. In static analysis, we first reconfirm previous work of Albert et al. [5]. Based on these results, we address the problems of existing metrics, the average diameter and the S metric, and propose new network connectivity metrics, K and DIK. Second, we put that result to test by diluting the sequence of faults with a few attacks, which quickly strips scale-free networks of any advantage in resilience. Our study shows that scale-free networks including the Internet do not have any advantage at all under a small fraction of attacks ($\beta > 0.05$ (5%)) with all metrics. Moreover, we show that the Internet is much more vulnerable under a small fraction of attacks than the BA model – even 1% of attacks decrease connectivity dramatically. In dynamic analysis, we trace the changes of the Internet's average diameter and its robustness against failures while it grows. Our study demonstrates that the Internet has been becoming more preferential over time and its susceptibility under attacks has been getting worse. Our results imply that if the current trend continues, the threat of attack will become an increasingly serious problem in the future.

Finally, we analyze 25 Internet topologies examined from November, 1997 to November, 1999, and perform a detailed analysis of dynamic characteristics of the Internet. These results provide insight into the evolution of the Internet, may be used to predict how the Internet will evolve in the future, and may be used to create improved network models.

II. PREVIOUS WORK

Network topology ties together many facets of a network's life and performance. It is studied at the overall topology level [6], link architecture [7], [8], and end-to-end path level [9], [10]. Temporal characteristics of a network are inseparable consequences of its connectivity. This linkage is apparent from [11], [12], [13]. Scaling factors, such as power-law relationships and Zipf distributions, arise in all aspects of network topology [6], [14] and web-site hub performance [15].

Topology considerations inevitably arise in clustering clients around demanding services [16], strategically positioning "digital fountains" [17], and mobile positioning [18] *et ad infinitum*. In QoS and *any*cast, topology dictates growing overlay trees, reserved links and nodes, and other sophisticated connectivity infrastructure affecting overall bandwidth through hubs and bottlenecks [19], [20], [21]. Other special connectivity infrastructures include P2P netherworlds [22] and global, synchronizable storage networks with dedicated topology and infrastructure for available, survivable network application platforms such as the Intermemory [23], [24], [25].

An important aspect which shows up more and more is fault control [26]. Several insights have come from physics, with the cornerstone work by Barabási [5], and further detailed network evolution models, including small worlds and Internet breakdown theories [27], [28], [4], [29], [30], [31], [32].

Albert, Jeong, and Barabási [5] examine the dichotomy of *exponential* and *scale-free* networks in terms of their response to errors. They found that while exponential networks function equally well under random faults and targeted attacks, scale-free networks are more robust to faults but susceptible to attacks. Because of their skeletal hub structure, preferential networks can sustain a lot of faults without much degradation in *average distance*, $\overline{d}$, a metric also introduced in [5] to aggregate connectivity of a possibly disconnected graph in a single number. According to the same metric, uniform networks behave in the same way under faults and attacks when a small percentage (about 5%) of the network is destroyed, while preferential networks weather faults much longer but fail pathetically under attacks on their hubs.

III. NETWORK MODEL AND SIMULATION ENVIRONMENT

Network models can be divided into two categories according to their generation methods: static and evolving (growing) [4]. In an evolving model, nodes are added over time – time goes in steps, and at each time step a node and m links are added. The probabilities in such a network are time-dependent (because the total number of nodes/edges changes with each time-step). In a static network model, the total number of nodes and edges are fixed and known in advance. Note that this difference between the models affects the probability of each node to gain new edges –

old nodes have a higher probability than new nodes to gain new edges in an evolving network model. Both classes of models can be placed at the edges of a *seniority* continuum, defined as follows. Seniority is a probability σ that all of the m edges of this iteration will be added immediately, or at the end of time. A seniority value of 1 corresponds to a pure time-step model, and a seniority value of 0 represents a pure static model.

In our simulations, we use a modified version of the model in [33] for comparison with the Internet. The model contains a parameter, α, which quantifies the natural intuition that every vertex has at least some baseline probability of gaining an edge. In [33], both endpoints of edges are chosen according to a mixture of probability α for preferential attachment and $1 - \alpha$ for uniform attachment. Let k_i be the degree of the *i*th node and m denotes the number of edges introduced at each time-step. If m_0 represents the number of initial nodes and t denotes the number of time-steps, the probability that an endpoint of a new edge connects to vertex i is

$$\Pi(k_i) = \alpha \frac{k_i}{2mt} + (1 - \alpha) \frac{1}{m_0 + t}.$$

An α value of 0 corresponds to a fully uniform model, while α values close to 1 represent mostly preferential models.

When an evolving network is generated, we initially introduce a seed network with two nodes and an edge between them ($n_0 = 2$, $e_0 = 1$).[1] Then, at each time-step, after a new node is introduced, new edges can be located with two different edge increment methods: external-edge-increment [5], [1] and internal-edge-increment [33]. In a growing exponential network with the external-edge-increment method, a new node is connected to a randomly chosen existing node. However, with internal-edge-increment, new edges are added between two arbitrary nodes chosen randomly. In our experiment, unlike [33], we apply external-edge-increment instead of internal-edge-increment because preferential networks generated by internal-edge-increment contain too many isolated nodes. Note that when α equals 1, preferential networks in our experiments are the same as the Barabási-Albert (BA) model in [5], [1], which is very similar to the network in [33] with $\alpha = 0.5$.

Failures can be characterized as either *faults* or *attacks* [5]. Faults are random failures, which affect a node independent of its network characteristics, and independent of one another. On the other hand, attacks maliciously target specific nodes, possibly according to their features (e.g., connectivity, articulation points, etc.), and perhaps form-
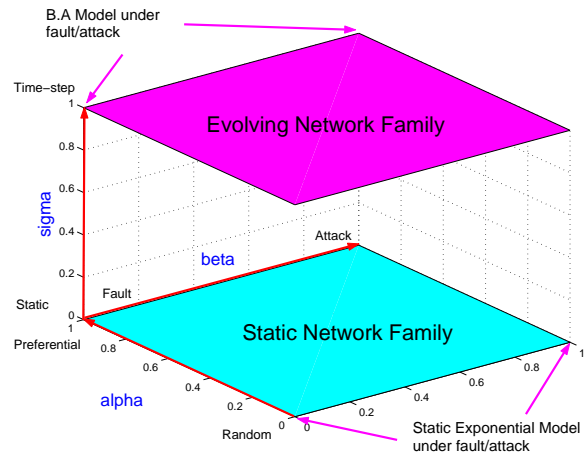

Fig. 1. Phase space of the network models in our study. We conducted experiments with both the evolving network family (pure time-step models) and the static network family. We focus on the evolving network family because most real networks are considered to be evolving networks.

ing a strategic sequence. The topology of the network affects how gracefully its performance degrades, and how late disintegration occurs. To measure robustness of networks against mixed failures, we use β for characterizing failures. With probability $1 - \beta$, a failure is a random fault destroying one node chosen uniformly. Otherwise (probability β), the failure is an attack that targets the single most connected node. When β equals 1, all failures are attacks, and when β equal 0, all failures are faults.

Figure 1 shows the phase space of different network models. We conducted experiments with both the evolving network family (pure time-step models) and the static network family. However, in this paper we mainly compare the robustness of two different types of evolving networks: evolving exponential (uniform) networks and evolving scale-free (preferential) networks, because many real networks, such as the Internet and the World Wide Web, are considered to be evolving networks.

We implemented our simulation environment in C++ with LEDA [34][2]. The networks are derived from LEDA's *graph* type, with additional features and experiments as separate modules. We do not allow duplicate edges and self-loops in our models and we delete all self-loop links from the Internet.

Like [5], the Internet's robustness against failures can be measured from a snapshot of the Internet. We call this kind of analysis *Static Analysis*. However, the Internet is a growing network and its topology changes continuously. Does the growth mechanism of the Internet affect its robustness? How is the Internet's robustness changing while it is growing? Will performance and robustness of the Internet improve in the future? To answer these questions, we analyze historical Internet topologies. We call

---

[1] A seed network is needed to generate a network using the preferential model – the probabilities of new links for all initial nodes at $t = 1$ are zero if there are no initial links.

[2] Library of Efficient Data types and Algorithms (LEDA), available at http://www.algorithmic-solutions.com.

this *Dynamic Analysis*. In this paper, we mainly compare the robustness of the Internet with two different network models, the BA model and a growing exponential network model (GE model).

IV. STATIC ANALYSIS OF THE INTERNET'S SUSCEPTIBILITY TO FAULTS AND ATTACKS

A. Metrics

As noted in [35], finding a good connectivity metric remains an open research question. [5] introduced two important metrics, $\overline{d}$ and S. The average diameter or average shortest path length, $\overline{d}$, is defined as follows: let $d(v, w)$ be the length of the shortest path between nodes v and w; as usual, $d(v, w) = \infty$ if there is no path between v and w. Let Π denote the number of distinct node pairs (v, w) such that $d(v, w) \neq \infty$ where $v \neq w$.

$$\overline{d} = \frac{\sum_{(v,w) \in \Pi} d(v, w)}{|\Pi|}$$

where $v \neq w$. To evaluate the reliability of the $\overline{d}$ metric, we started with measuring the robustness of three different evolving networks under faults or attacks only. Our experiments are somewhat different from [5]. We compared behaviors of the growing scale-free network (the BA model) and the Internet with those of the growing random network (the GE model), while [5] used static exponential networks for comparison.

As we expected, our results are very similar to [5]; A growing exponential network performs worse under faults, but better under attacks. However, as we can see in Figure 2(a), $\overline{d}$ is not always representative of the overall connectivity because it ignores the effect of isolated nodes in the network. Note that $\overline{d}$ is decreasing rapidly after a certain threshold under attacks only, showing that when the graph becomes sparse, $\overline{d}$ is less meaningful. The other metric, S, is defined as the ratio of the number of nodes in the giant connected component divided by the total number of nodes. One might notice the different characteristics of the two metrics. Shorter average diameter means shorter latency. It demonstrates how fast a network can react when an event occurs, providing an indication of the performance of a network. On the other hand, S mainly considers the networks' connectivity, showing how many nodes are connected to the largest cluster.

Since the S metric only considers the relative size of the largest connected component, and does not characterize the entire network, we created a new metric, K, that describes the whole network connectivity. K is defined as follows: let Ψ be the number of distinct node pairs, and Π is defined as above. Then

$$K = \frac{|\Pi|}{|\Psi|}$$

K measures all connected node-pairs in a network. In Figure 2, we can see that the Internet shows the best robustness under faults according to the diameter. However, if we use the K or S metrics, the Internet is most vulnerable even under faults.

One weakness of the K metric is that it does not consider the effect of redundant edges. The K value for a connected graph with n nodes and n-1 edges[3] ($K = 1$, $\overline{d} \geq 1$) is the same as that of a fully connected graph[4] ($K = 1$, $\overline{d} = 1$) even though the diameter and connectivity of each graph is quite different. To solve this problem, we introduce a modified diameter metric, which we call *Diameter-Inverse-K (DIK)*. *DIK* is defined as:

$$DIK = \frac{\overline{d}}{K}$$

The *DIK* metric uses the K metric as a penalty parameter for sparse graphs and measures both the expected distance between two nodes and the probability of a path existing between two arbitrary nodes. Figure 2 demonstrates that $\overline{d}$ significantly decreases when it reaches a certain threshold, while *DIK* continuously increases. Note that the Internet is most vulnerable even under faults if we measure network connectivities with S or K.

B. Robustness against Mixed Failures

In real life, it is somewhat unrealistic to expect that failures are either all faults or all attacks. One may expect that failures are a mixture of attacks and faults, e.g., only a small fraction of failures are attacks while most failures denote faults. In the following experiments, network destruction was performed until 10% of the total number of nodes was destroyed, using different values of β (probability of attack). We performed 10 runs in each case with different seed numbers. The results in Figure 3 are the average of the ten runs. We define the average diameter ratio as $\overline{d_f}/\overline{d_o}$ where $\overline{d_o}$ denotes the average diameter of the initial network, and $\overline{d_f}$ is the average diameter after 10% of the nodes have failed. Similarly, the DIK ratio is defined as DIK_f/DIK_o where DIK_o is the DIK value of the original network, and DIK_f is the DIK value after 10% of the nodes have failed. Figure 3 shows that: (a) Although there seems to be an advantage for scale-free networks under pure faults, their disadvantage under attacks is much larger, and even a small fraction of attacks, $\beta > 0.05$ (5%), in a mix of failures removes any overall advantage of the scale-free networks. (b) The K metric is even more unforgiving to the scale-free networks, showing no advantage under any $\beta \geq 0.01$ (1%). Note that the Internet shows the worst robustness even under faults

---

[3] A graph where all nodes are connected to the giant connected component.

[4] A graph where all nodes are connected to all other nodes.

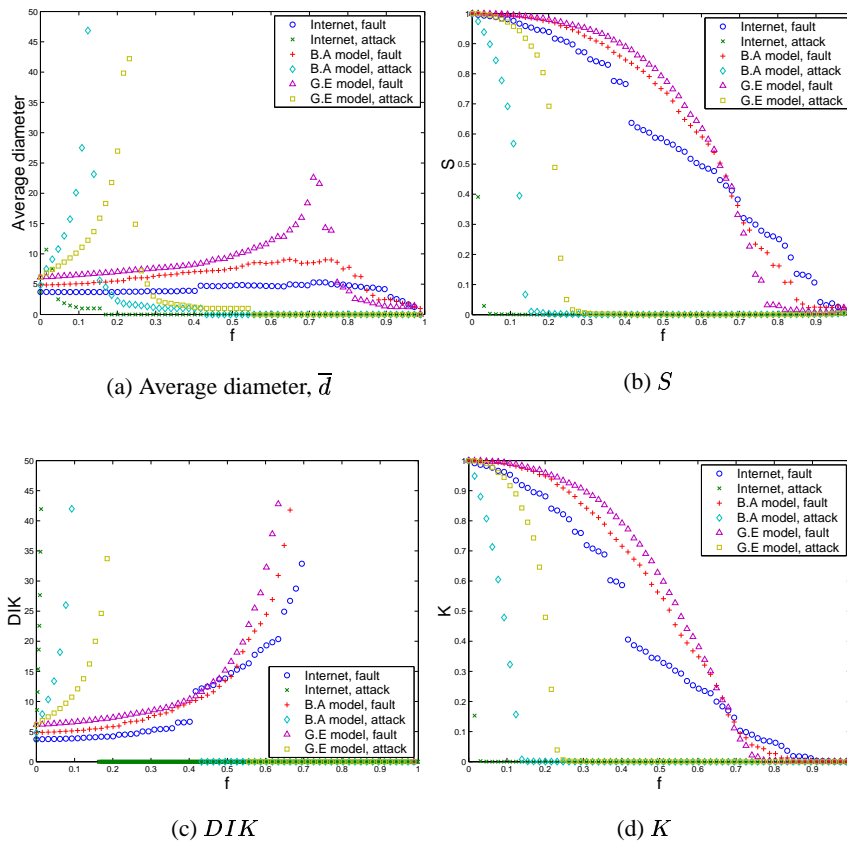(a) Average diameter, $\overline{d}$

(b) S

(c) DIK

(d) K

Fig. 2.   Robustness against faults/attacks; We used the AS (Autonomous System) level topology of the Internet with 6474 nodes and 13895 edges from http://moat.nlanr.net/Routing/rawdata/, which was examined on Jan. 2, 2000. After removing self-loops, the number of edges decreased to 12572. For growing network models, we set m equal to two and generated networks with 6474 nodes. f denotes the number of failure nodes divided by the total number of nodes in the original network. Two nodes and an edge between them are initially introduced when we generate the network ($n_0 = 2$, $e_0 = 1$). (a) and (c): (a) shows $\overline{d}$ for the Internet, and for the BA and GE models. Note that $\overline{d}$ significantly decreases when it reaches a certain threshold, while DIK continuously increases. (b) and (d): The S and K metrics do not agree with the previous observations using $\overline{d}$. The Internet is most vulnerable under both attacks and faults using these metrics. Even though S and K behave very similarly, S only considers the relative size of the giant connected component, while K considers all node pairs which are connected. We set DIK to zero when $\overline{d}$ and K becomes zero. Note that smaller is better for $\overline{d}$ and DIK, but larger is better for S and K.

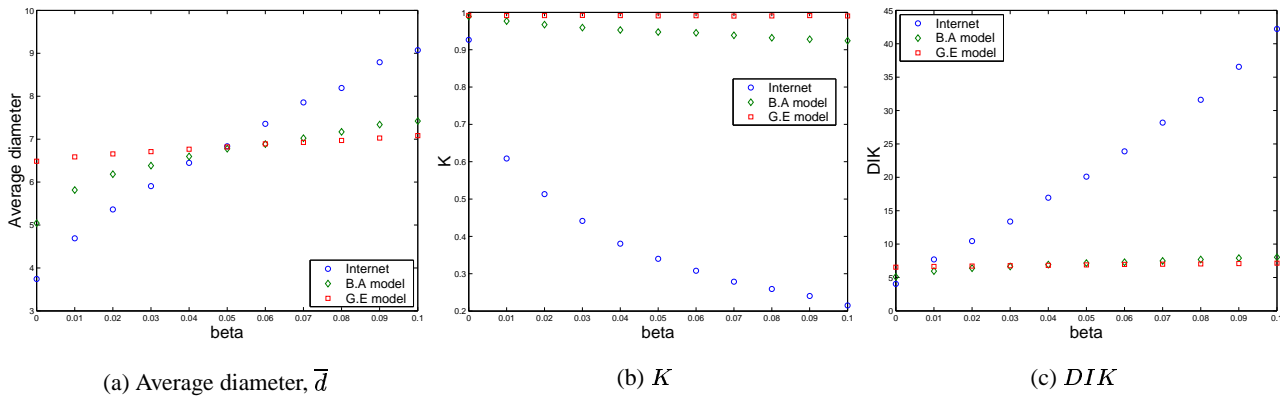(a) Average diameter, $\overline{d}$

(b) K

(c) DIK

Fig. 3.   Robustness of the Internet, and the BA and GE models under mixed failures after 10% of total nodes are destroyed. (a): The average diameter of the Internet and the BA model increases rapidly compared with the GE model as β is increasing. The advantage of smaller $\overline{d}$ disappears when $\beta > 0.05$ (5%). Figure (c) demonstrates this trend more clearly. Note that even 1% of attacks significantly hurts robustness of the Internet. (b): The K metric is even more unforgiving to the scale-free networks, and shows no advantage under any $\beta \geq 0.01$ (1%). The Internet shows the worst robustness even under faults only. The results shown are the average of ten runs. Note that smaller is better for $\overline{d}$ and DIK, but larger is better for K.

only. Figure 3(c) clearly shows the vulnerability of the Internet under a small fraction of attacks. DIK is increasing very rapidly and even 1% of attacks significantly hurts its

robustness.

We also measured the effect of preferential attachment and observed the following trends. First, more preferen-
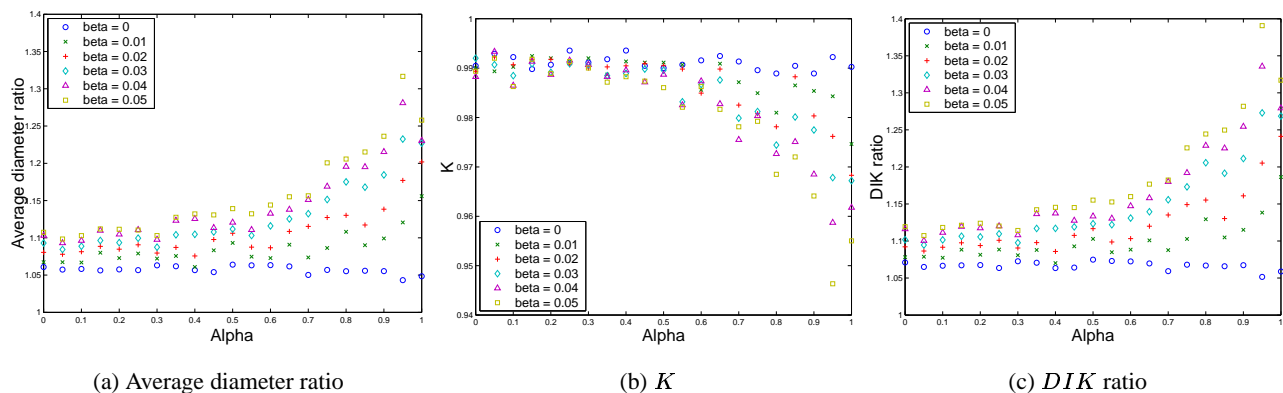
(a) Average diameter ratio	(b) K	(c) DIK ratio

Fig. 4. Robustness of the various network models ($0 \leq \alpha \leq 1$) under mixed failures after 10% of total nodes are destroyed. Note that larger α means more preferential networks and smaller $\overline{d}$ and DIK, but larger K means greater robustness. Each network contains 1000 nodes. (a) and (c): $\overline{d}$ ratio and DIK ratio increments are growing when β is increasing. However, a small fraction of attacks ($\beta \geq 0.01$ (1%)) cancels out this advantage of the scale-free network and damages preferential networks more. (b): With the K metric, preferential networks do not show any noticeable advantage even under attack. The results shown are the average of ten runs.
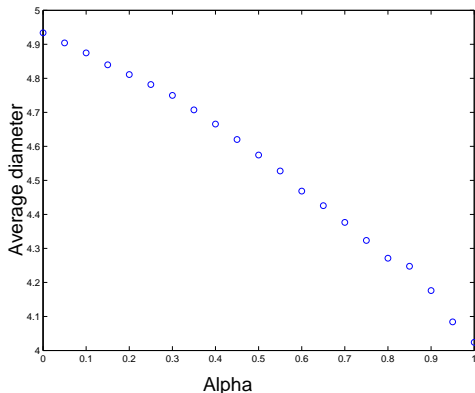

Fig. 5. Relationship between preferentiality and average diameter; While α is increasing, the average diameter of the networks generated is decreasing. Results are the average of 10 different networks with different seed numbers.

tial networks have shorter average diameters. We generated networks with various α and observed this trend, as shown in Figure 5. The most preferential network with n nodes and $n - 1$ edges has all nodes connected to the most popular node. The diameter from the most popular node to others is one and the diameter between any two nodes except the most popular node is two, therefore the average diameter is less than two, and the network has the smallest diameter of all possible networks with n nodes and $n - 1$ edges. Second, more preferential networks are more robust under faults only, but more vulnerable under even a small fraction of attacks if we measure robustness using the average diameter. Figure 4 demonstrates that when α is close to 1, even a small fraction of attacks ($\beta \geq 0.01$ (1%)) cancels out the advantage of the scale-free networks and hurts their topologies more. Note that if the average diameter reaches a certain threshold, it decrease rapidly and becomes meaningless. Third, with the K metric, a preferential network does not show any noticeable advantage even under attack, and an exponential network dominates all kinds of failures.

V. DYNAMIC ANALYSIS OF THE INTERNET'S SUSCEPTIBILITY TO FAULTS AND ATTACKS

In this section, we measure changes in the Internet's robustness against failures over time. We sampled eight Internet topologies from different points in time from http://moat.nlanr.net/Routing/rawdata/. Self-loop links were removed. First, we measured the average diameter. We also generated the BA model and the GE model and measured their average diameters. While the number of nodes in the Internet increased, the average diameter actually decreased, which can not be explained by the BA model. Both the BA and GE models predict an increasing average diameter as the number of nodes increases, as shown in Figure 6.

Next, we trace the robustness of the Internet while it is growing. For each Internet topology, we destroy 10% of the total number of nodes and measure robustness with three different metrics – average diameter, K, and DIK. Figure 7 (a) and (d) show the robustness of the Internet with the average diameter. The average diameter ratio of the Internet is decreasing while the number of nodes is increasing under pure faults. Note that the average diameter ratios of other network models are fluctuating and do not show any clear trend. Figure 7(d) is misleading because the Internet topology becomes too sparse after 10% of the nodes are removed. Note that the average diameter is meaningless when a graph contains many isolated nodes. With the K and DIK metrics, we observe a clear trend: the Internet becomes more robust under faults, but more vulnerable under attacks while it grows. In other words, the Internet has been becoming more preferential over time and the growth mechanism of the Internet focuses on maximizing overall performance (decreasing average diameter) rather than robustness against attacks, and the Internet's susceptibility under attacks will be a more serious problem in the future if this trend continues.
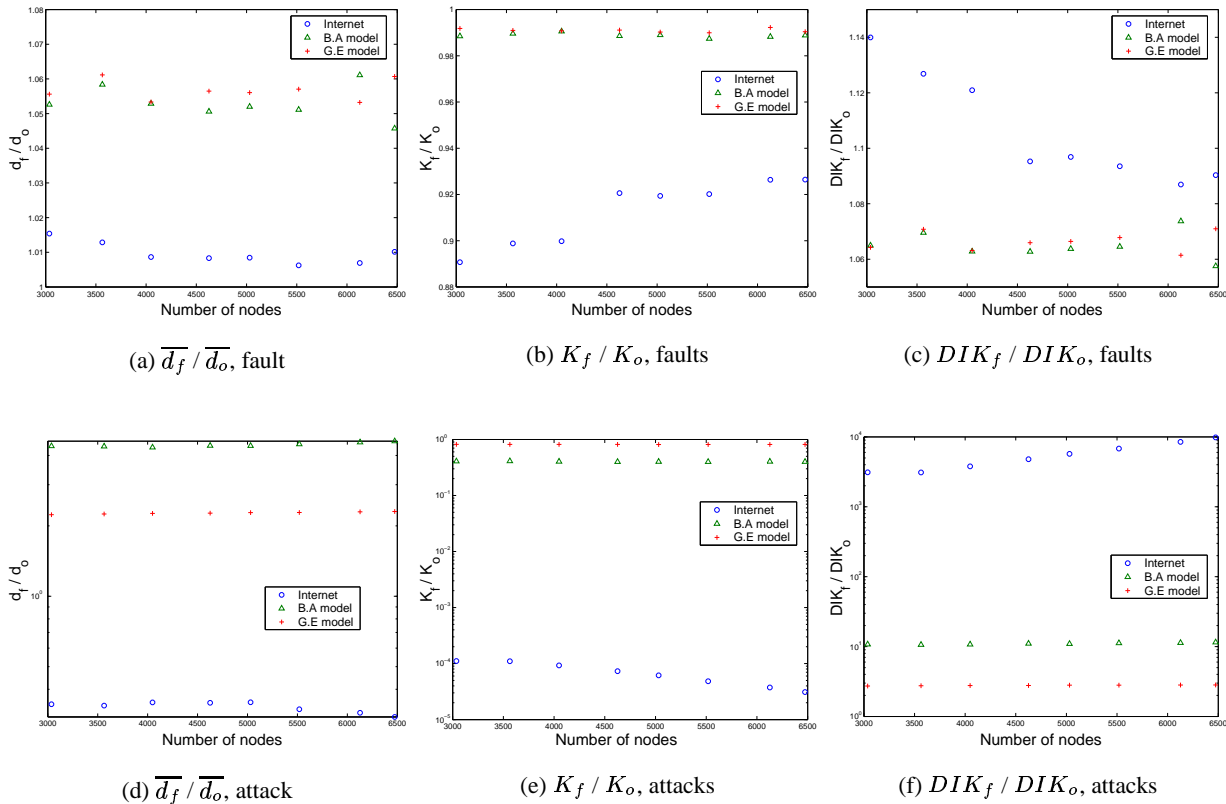
Fig. 7. Dynamic characteristics of the Internet; $\overline{d_o}$, K_o and DIK_o are defined as the average diameter, K, and DIK of the original networks and $\overline{d_f}$, K_f and DIK_f denote the diameter, K, and DIK after 10% of the nodes are removed. Results are the average of ten runs. (a) and (d): (a) shows that the average diameter ratio of the Internet is decreasing while the number of nodes are increasing under pure faults. (d) is misleading because the Internet topology becomes too sparse after 10% of nodes are removed. (b) and (e): While the Internet is growing, the K ratio of the Internet is increasing under faults but decreasing under attacks. (c) and (f): (f) also agrees with previous observations that the Internet becomes more robust under faults but more vulnerable under attacks while it is growing. Note that smaller is better for $\overline{d}$ and DIK, but larger is better for S and K.

VI. DYNAMIC CHARACTERISTICS OF THE INTERNET

Characteristics of the Internet topology and its robustness against failures have been widely studied in [6], [14], [5], [1], focusing on extracting common regularities from several snapshots of the Internet topology. Those characteristics, e.g., power-law of the degree distribution, can be called *Static Characteristics* because of their consistency over time. Several Internet topology generators have been introduced [36], [14], [37], [1], [38], [39] based on the static characteristics. Recent research has argued that the performance of network protocols can be seriously effected by the network topology [40]. Unfortunately, usage of existing Internet topology generators to develop network protocols is significantly limited for the following reasons. The Internet is a growing network and its topology and characteristics can change over time. For example, the clustering coefficient of the Internet has been growing and the average diameter of the Internet has been decreasing during the past few years [39]. We call these *Dynamic Characteristics* of the Internet. Since current Internet topology generators are designed using only the static characteristics of the Internet, they suffer from a

lack of ability to predict future Internet topology. Currently, the best method to simulate network protocols is using the actual Internet topology instead of using Internet topology generators, however this limits our ability to develop, for example, network protocols that best fit future conditions. Most existing Internet topology generators fail to explain some of the dynamic characteristics of the Internet that we have found. For example, we found that the average degree of the Internet is growing but most Internet topology generators do not show this characteristic. We believe that the dynamic characteristics should be considered when an Internet topology generator is designed.

In order to analyze the dynamic characteristics of the Internet topology in detail, we sampled 25 Internet topologies from http://moat.nlanr.net/Routing/rawdata/[5]. We first analyze the number of total nodes, node births, and node deaths in the Internet topologies. Since we cannot guarantee that our data set covers whole Internet topologies,

---

[5]They were examined on the 15th of each month from November, 1997 to November, 1999. Since most Internet topology generators and previous work does not consider self-loop links, we removed all self-links (around 9% of total links).
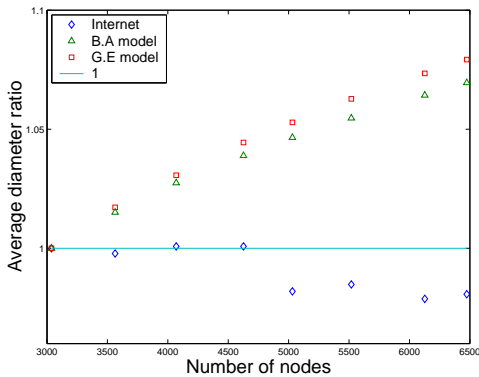
Fig. 6.   Diameter ratio while a network is growing. We sampled eight topologies of the Internet, examined on 11/15/1997 (3037 nodes), 04/08/1998 (3564 nodes), 09/08/1998 (4069 nodes), 02/08/1999 (4626 nodes), 05/08/1999 (5031 nodes), 08/08/1999 (5519 nodes), 11/08/1999 (6127 nodes), and 01/02/2000 (6474 nodes), and measured their diameters. For comparison, we also generated the BA and GE models and measured their average diameters. We generated each network model ten times with different seed numbers and calculated average values. Each d_i is divided by d_o, the diameter of the first network with 3037 nodes. d_o is 3.78 for the Internet, 4.51 for the BA model, and 5.20 for the GE model. Note that as the networks are growing, the diameter of the BA and GE models increases, while the diameter of the Internet decreases, indicating a growth mechanism that maximizes performance (minimizing diameter and latency).

and a node may not be discovered because of a temporary failure, we define that a node is dead only when it does not appear in future Internet topologies. For example, a node in November, 1997 is considered to be deleted only when it never appears from December, 1997 to November, 1999.

Figure 8(a) shows clear regularity in the number of total nodes, added nodes, and deleted nodes over the period of November, 1997 to November, 1999. We also measured the number of total links, added links, and deleted links as shown in Figure 8(b). The total number of nodes and edges increases quadratically and we can predict the number of nodes in the near future with the equations given in (a) and (b).

We also reconfirm the previous result that the Internet has been becoming more preferential. For this purpose, we introduce a new metric *Preferentiality (PF)*. Let n be the number of nodes in the network and r be a rank of a node according to its degree. The highest degree node has rank one and any two nodes cannot have the same rank. k_r and k_{avg} denote the degree of rank r nodes and the average degree. Then *PF* is defined as:

$$PF = \frac{\sum_{r=1}^{n}(r * k_r)}{k_{avg} * \frac{n*(n+1)}{2}}$$

This metric shows how preferential the network is. PF values close to 0 mean that the network is extremely preferential, and PF values close to 1 means the network is extremely random. The PF value of the Internet is $0.40 \sim 0.39$ while that of the BA and GE models are

0.61 and 0.70. Our results show two interesting observations: First, the Internet is a much more preferential network than the BA model and its susceptibility under attacks is much larger than even the general scale-free networks such as the BA model. Second, the Internet has been becoming more preferential as it grows and its vulnerability has been increasing. Figure 8(c) clearly shows this trend.

Average degrees of the Internet topologies are shown in Figure 9(a). In most of the time-step based Internet topology generators including [1], [38], [39], the number of links added at each time-step is fixed. However, the average degree of the Internet is increasing linearly while the number of nodes is increasing, which implies that previous time-step and fixed number of node/edge addition approaches may not generate the proper Internet topologies. Links can be introduced by two main processes. When a new node is introduced, new links are introduced which connect the new node to existing nodes. We previously defined this process as *external edge increment*. Otherwise, links can be added between two existing nodes, defined as *internal edge increment* earlier. In a few cases, we found that a link is introduced between two new nodes and those cases are ignored. In general, 1.3 links per new node are added by external edge increment and 1.9 links per new node are added by internal edge increment. A total of 3.2 links per new node are added over time. Note that internal edge increment affects link increment more than external edge increment. $71 \sim 66\%$ of new nodes are introduced with a single link and $25 \sim 31\%$ of new nodes are added with two links. Only $3 \sim 4\%$ of new nodes are introduced with more than two links. Like link births, a link can be discarded two ways. When a node is dead, links connected to the node are broken. Also, a link can die when any one of the connected nodes decides to be disconnected from the other. We define the former as *external edge death* and the latter as *internal edge death*. Node death is not the main factor in link death – link death frequently happens without node death. Around 79% of dead links are broken due to internal edge death. In general, the average number of dead links due to node death increased slightly over time. This can be explained since the average degree of the Internet is increasing. Around $1.1 \sim 1.5$ links were broken when a node was discarded. The average number of internal edge deaths is more than three times larger than that of external edge deaths in the same time period. $8.7 \sim 5.5$ links per node death are discarded during November, 1997 to November, 1999. Note that the average number of link deaths per node death was increasing; the death rate of links overcomes the death rate of nodes. One of the interesting observations for link and node death is that more than 76% of dead nodes had degree one, but less than 18% of dead nodes had degree two.

(a) Number of nodes (b) Number of links (c) Preferentiality of the Internet
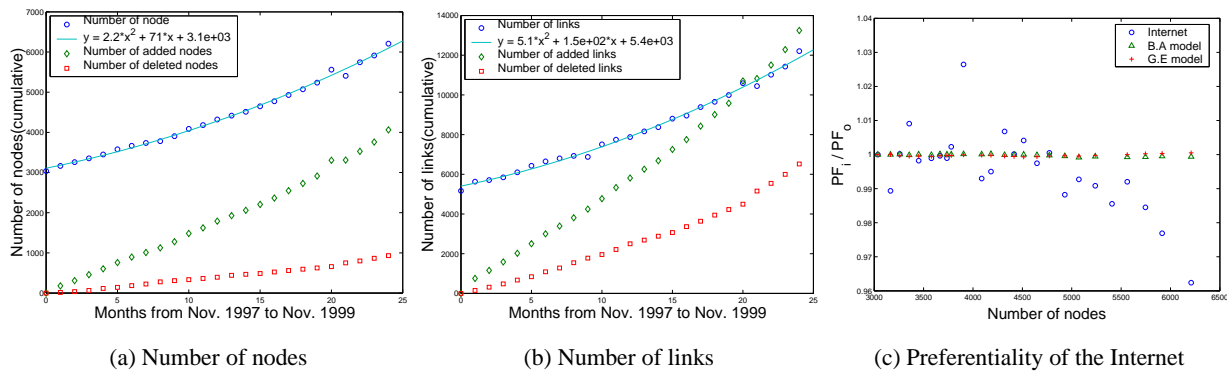
Fig. 8.   Dynamic characteristics of the Internet – number of nodes and links, and preferentiality of the Internet. (a) and (b): The number of nodes/links is increasing quadratically. (c): PF_o is the PF value of a network with 3037 nodes and PF_i represent the PF value of a network with i nodes. PF_o of the Internet, the BA model, and the GE model are 0.404, 0.610, and 0.699. Lower values of PF indicate a more preferential network. Note that, in general, the Internet was becoming more preferential. Results for the network models are averaged over 10 runs with different seed numbers.

Note that there are almost the same number of nodes with degree one and two in the Internet. It clearly shows that less popular nodes are more likely to die.

Figure 10(a) shows that the fraction of nodes with degree one in the Internet is decreasing while that of nodes with degree two is increasing. Note that around 70% of nodes have degree one or two. Figure 10 (b) and (c) clearly show the limitations of the BA model-like topology generators. First, there are no nodes with degree one. The results strongly imply that the proposal of [39] is a better model for the Internet. In this proposal, when a new node is added, a node and a single link are introduced with probability α, but a node and two links are introduced with probability $1 - \alpha$. Second, several complex mechanisms are involved with the growth of the Internet. Our simple growing network model only incorporates external edge increment, missing other important factors in the Internet's growth – internal edge increment, internal edge death, and external edge death.

VII. FUTURE WORK

Our study may be extended in various ways, for example:

- *Internet topology generator*
  Currently, we are designing a new Internet topology generator which fits not only the static characteristics but also the observed dynamic characteristics of the Internet. This generator can be used for simulation to develop network protocols aiming to have optimal performance in the future.
- *Metrics*
  New overall connectivity or QoS metrics can be created, for example one possibility is k-disjoint paths: how many paths are there, on average, between any two nodes, which have at least k different edges? Novel approaches are also desirable, soliciting actual survivability/performance degradation metrics from other network practitioners.

- *Overall performance degradation caused by local network congestion*
  Instead of attacking the most popular nodes, selected edges can be blocked. If user requests in the network increase, the number of requests in the most popular links will increase and may be blocked by network congestion. How will the network as a whole be affected by local network congestion?

VIII. CONCLUSIONS

In our study, we first re-evaluated two basic connectivity metrics, average diameter and *S*. The average diameter may be a good metric for measuring the performance of networks, but is not always representative of the overall network connectivity. The *S* metric only considers the relative size of the largest component and ignores other components. To analyze the Internet's susceptibility to faults and attacks, we introduced two new metrics, *K* and *DIK*. Unlike *S*, *K* measures all connected node-pairs in a network. Also, unlike average diameter, *DIK* is still valuable in sparse graphs, and incorporates both the average expected distance between two nodes, and the probability of a path existing between two arbitrary nodes. We also examined the robustness of the Internet under mixed failures. We found that any advantage of scale-free networks, including the Internet, disappeared when a small fraction of failures are attacks, or when using metrics other than the average diameter. We also conducted dynamic analysis of the Internet's susceptibility to attacks and faults, and discovered two interesting results; First, the Internet is much more preferential than the BA model, and its susceptibility under attacks is much larger than even general scale-free networks such as the BA model. Second, the growth mechanism of the Internet stresses maximizing performance, and the Internet is evolving to an increasingly preferential network. If this trend continues, attacks on a few important nodes will be a more serious threat in

(a) Average degree

(b) Average number of links added per node birth

(c) Percentage of new nodes with degree k

(d) Cause of link death

(e) Average number of links discarded per node death
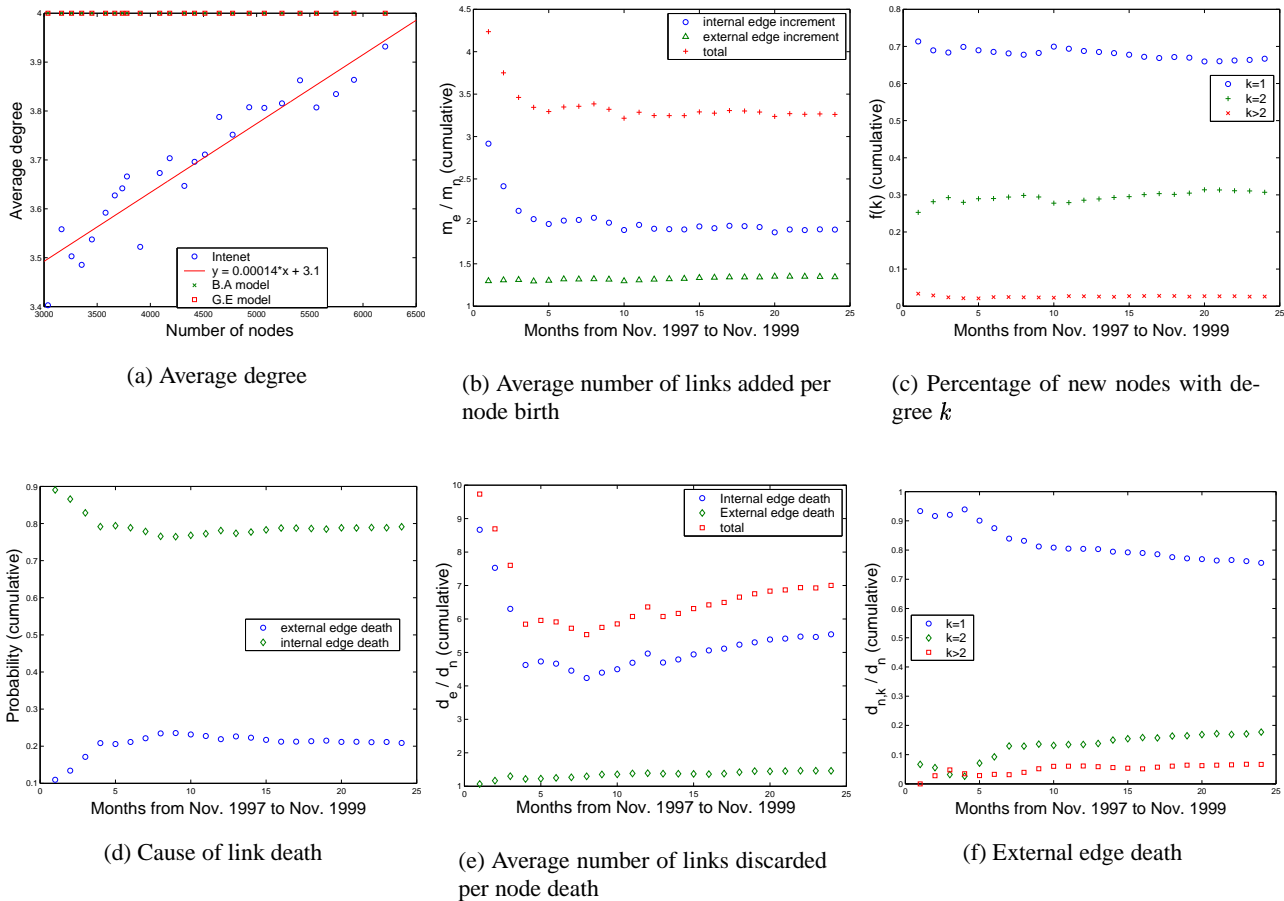
(f) External edge death

Fig. 9. Dynamic characteristics of the Internet – average degree, creation of nodes and links, and death of nodes and links; (a) In most time-step based Internet topology generators including [1], [38], [39] , the number of links added at each time-step is fixed. However, the average degree of the Internet is increasing linearly while the number of nodes is increasing. (b): m_n and m_e denotes the number of nodes and links added since November, 1997. In general, 1.3 links per new node are added by external edge increment, and 1.9 links per new node are added by internal edge increment. A total of 3.2 links per new node are added over time. Note that internal edge increment affects link increment more than external edge increment. (c): For external edge increment, $71 \sim 66\%$ of new nodes are introduced with a single link and $25 \sim 31\%$ of new nodes are added with two links. Only $3 \sim 4\%$ of new nodes are introduced with more than two links. (d) and (e): External edge death is not a main factor in link death. Only about 21% of dead links were due to node deletion and 79% of link death were made without node death. d_n and d_e denote the number of nodes and links deleted since November, 1997. The number of internal edge deaths per node death are more than three times larger than that of external edge death in the same time period. $8.7 \sim 5.5$ links per node death are discarded during November, 1997 to November, 1999. (f): $d_{n,k}$ denotes the number of degree k nodes deleted since November, 1997. More than 76% of destroyed nodes have degree one even though there are almost the same number of nodes with degree one and two in the Internet. This figure clearly shows that less popular nodes are more likely to die.

the future. Finally, we addressed dynamic characteristics of the Internet in detail, finding that:

- The number of nodes and links has been increasing quadratically over time.
- The average degree of the Internet has been increasing linearly.
- $71 \sim 66\%$ of new nodes are introduced with single links and $25 \sim 31\%$ of new nodes are introduced with two links. Only $3 \sim 4\%$ of new nodes are introduced with more than two links.
- Two edge increment mechanisms – external edge increment and internal edge increment – affect link birth. In general, 1.3 links per new node are added by external edge increment, and 1.9 links per new node are added by internal edge increment. A total

of 3.2 links per new node are added over time.
- Node death is not the main factor in link death. Link death frequently happens without node death. Only about 21% of dead links are due to node death, while 79% occur without node death.
- Less popular nodes are more likely to die. More than 76% of dead nodes have degree one, but less than 18% of dead nodes have degree two. Note that there are almost the same number of degree-one nodes and degree-two nodes. Only 6% of dead nodes have degree more than two.

The observed characteristics of the Internet topology shows that existing network generators not only lack the ability to predict future Internet topologies, but also are not able to explain or generate many characteristics.
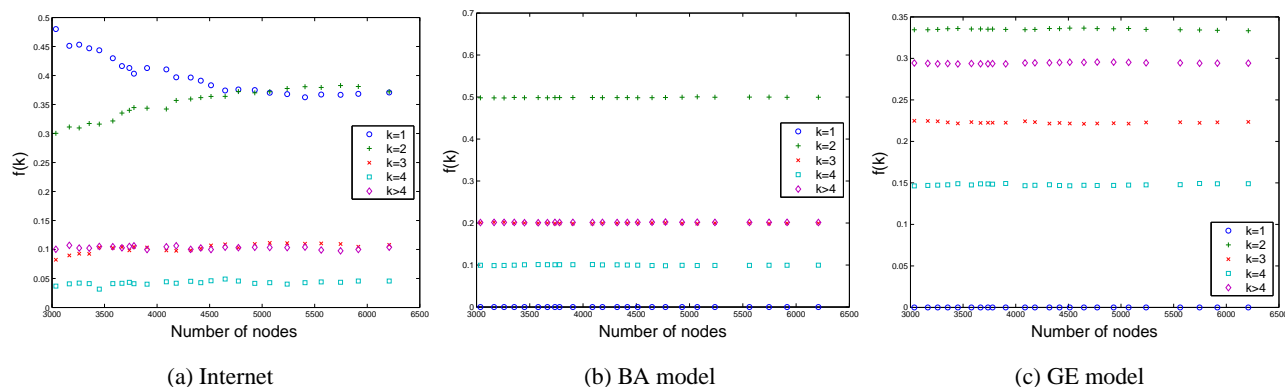
(a) Internet   (b) BA model   (c) GE model

Fig. 10.   Dynamic characteristics of the Internet – relative size of nodes with degree k; (a): $f(k)$, the fraction of nodes with degree k. In the Internet, the fraction of nodes with degree one is decreasing while that of nodes with degree two is increasing. Note that around 70% of nodes have degree one or two. (b) and (c): These plots clearly show limitations of the BA model-like topology generators; First, there are no nodes with degree one. Second, the relative fraction of the same degree nodes does not change in our models – changes in Internet topology over time can not be explained by our network model.

Based on our analysis, we are currently designing new Internet topology generators.

REFERENCES

[1] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[2] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, "The web of human sexual contacts," *Nature*, vol. 411, pp. 907–908, 2001.

[3] B. Bollobás, *Random Graphs*, Cambridge Mathematical Library. Cambridge University Press, 2001.

[4] S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of networks," arXiv:cond-mat/0106144, 2001, submitted to Adv. Phys.

[5] R. Albert, H. Jeong, and A. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology," in *SIGCOMM*, 1999, pp. 251–262.

[7] B. Lowekamp, D. R. O'Hallaron, and Thomas Gross, "Topology discovery for large Ethernet networks," in *SIGCOMM*, 2001.

[8] D. S. Alexander, M. Shaw, S. Nettles, and J. M. Smith, "Active Bridging," in *SIGCOMM*, 1997, pp. 101–111.

[9] M. Allman and V. Paxson, "On Estimating End-to-End Network Path Properties," in *SIGCOMM*, 1999, pp. 263–274.

[10] E. Cohen, B. Krishnamurthy, and J. Rexford, "Improving End-to-End Performance of the Web Using Server Volumes and Proxy Filters," in *SIGCOMM*, 1998, pp. 241–253.

[11] A. Veres, Z. Kenesi, S. Molnár, and G. Vattay, "The Propagation of Long-Range Dependence in the Internet," in *SIGCOMM*, 2000.

[12] K. Lai and M. Baker, "Measuring link bandwidths using a deterministic model of packet delay," in *SIGCOMM*, 2000, pp. 283–294.

[13] A. B. Downey, "Using Pathchar to Estimate Internet Link Characteristics," in *SIGCOMM*, 1999, pp. 222–223.

[14] A. Medina, I. Matta, and J. Byers, "On the Origin of Power Laws in Internet Topologies," *ACM Computer Communication Review*, April, 2000.

[15] V. N. Padmanabhan and L. Qui, "The content and access dynamics of a busy web site: findings and implications," in *SIGCOMM*, 2000, pp. 111–123.

[16] B. Krishnamurthy and J. Wang, "On network-aware clustering of web clients," in *SIGCOMM*, 2000, pp. 97–110.

[17] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital Fountain Approach to Reliable Distribution of Bulk Data," in *SIGCOMM*, 1998, pp. 56–67.

[18] S. Cheshire and M. Baker, "Internet Mobility 4x4," in *SIGCOMM*, 1996.

[19] S. B. Fredj, T. Bonald, A. Proutiere, G. Régnié, and J.W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level," in *SIGCOMM*, 2001.

[20] L. Breslau and S. Shenker, "Best-Effort versus Reservations: A Simple Comparative Analysis," in *SIGCOMM*, 1998, pp. 3–16.

[21] E. Amir, S. McCanne, and R. H. Katz, "An Active Service Framework and Its Application to Real-Time Multimedia Transcoding," in *SIGCOMM*, 1998, pp. 178–189.

[22] A. Oram, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, O'Reilly, 2001.

[23] A. Khrabrov, S. Sobti, and P. N. Yianilos, "Synchronizable Databases for the Web," Tech. Rep., NEC Research Institute, 4 Independence Way, Princeton, NJ, December 2000.

[24] Y. Chen, J. Edler, A. Goldberg, A. Gottlieb, S. Sobti, and P. Yianilos, "A Prototype Implementation of Archival Intermemory," in *Proceedings of the fourth ACM Conference on Digital libraries (DL '99)*, 1999.

[25] A. V. Goldberg and P. N. Yianilos, "Towards an Archival Intermemory," in *Proc. IEEE International Forum on Research and Technology Advances in Digital Libraries (ADL'98)*. April 1998, pp. 147–156, IEEE Computer Society.

[26] A. Reddy, R. Govindan, and D. Estrin, "Fault isolation in multicast trees," in *SIGCOMM*, 2000, pp. 29–40.

[27] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Breakdown of the Internet under intentional attack," *Physical Review Letters*, vol. 86, 2001, arXiv:cond-mat/0010251.

[28] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Physical Review Letters*, vol. 85, 2000, arXiv:cond-mat/0007048.

[29] S. Dorogovtsev, J. Mendes, and A. Samukhin, "Structure of growing networks with preferential linking," *Physical Review Letters*, vol. 85, no. 21, 2000.

[30] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "Giant strongly connected component of directed networks," arXiv:cond-mat/0103629, 2001, Phys. Rev. E 64, 025101 (R) (2001).

[31] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs," *Physical Review Letters*, vol. 85, no. 25, pp. 5468–5471, December 2000.

[32] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-Field So-

lution of the Small-World Network Model," *Physical Review Letters*, vol. 84, no. 14, pp. 3201–3204, April 2000.

[33] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles, "Winners don't take all: Characterizing the competition for links on the web," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 99, no. 8, pp. 5207–5211, 2002.

[34] K. Mehlhorn and S. Näher, *LEDA: A Platform for combinatorial and geometric computing*, Cambridge University Press, 1999.

[35] A. Broder, R. Kumar, Prabhakar Raghavan F. Maghoul, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener, "Graph Structure in the Web," in *Proceedings of WWW9 Conference*, 2000.

[36] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet Topology Generator," 2000.

[37] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *Proceedings of the 32rd Annual ACM Symposium on Theroy of Computing*, 2000, pp. 171–180.

[38] R. Albert and A. Barabási, "Topology of evolving networks: local events and universality," Tech. Rep., LANL ArXiv, 2000.

[39] T. Bu and D. Towsley, "On Distinguishing between Internet Power Law Topology Generators," in *Proceedings of INFOCOM*, 2002.

[40] C. Labovitz, A. Ahuja, R. Wattenhofer, and V. Srinivasan, "The Impact of Internet Policy and Topology on Delayed Routing convergence," in *INFOCOM*, 2001, pp. 537–546.

[41] L. A. Adamic and B. A. Huberman, "The nature of markets in the World Wide Web," *Quarterly Journal of Electronic Commerce*, vol. 1, no. 1, pp. 5–12, 2000.

[42] L. A. Adamic and B. A. Huberman, "Power-law distribution of the World Wide Web," *Science*, vol. 287, pp. 2115a, 2000.

[43] R. Albert, H. Jeong, and A. Barabási, "Diameter of the World Wide Web," *Nature*, vol. 401, pp. 130–131, 1999.

[44] A. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A*, vol. 272, pp. 173–187, 1999.

[45] A. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: The topology of the World Wide Web," Tech. Rep., University of Notre-Dame, 1999.

[46] P. Barford, A. Bestavros, A. Bradley, and Mark Crovella, "Changes in web client access patterns," *World Wide Web, Special Issue on Characterization and Performance Evaluation*, vol. 2, pp. 15–28, 1999.

[47] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," in *Proceedings of the Ninth International World Wide Web Conference*, 2000.

[48] J. L. Casti, "Bell curves and monkey languages," *Complexity*, vol. 1, no. 1, pp. 12–15, 1995.

[49] C. R. Cunha, A. Bestavros, and M. Crovella, "Characteristics of WWW client-based traces," Tech. Rep. TR-95-010, Department of Computer Science, Boston University, 1995.

[50] S. Glassman, "A caching relay for the World Wide Web," *Computer Networks and ISDN Systems*, vol. 27, pp. 165–173, 1994.

[51] G. Hoff and M. Mundhenk, "Finding scientific papers with HPSearch and Mops," Tech. Rep. D-54286, Department of Mathematics/Computer Science, University Trier, Germany, 1999.

[52] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. M. Lukose, "Strong Regularities in World Wide Web Surfing," *Science*, vol. 280, pp. 95–97, 1998.

[53] B. A. Huberman and L. A. Adamic, "Growth dynamics of the World Wide Web," *Nature*, vol. 401, pp. 131, 1999.

[54] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, 2000.

[55] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, "The web as a graph: Measurements, models, and methods," in *Proceedings of the the International Conference on Combinatorics and Computing*, 1999.

[56] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Trawling the web for emerging cyber-communities," in *Proceedings of the Eighth International World Wide Web Conference*, 1999.

[57] S. Lawrence and C. L. Giles, "Accessibility of information on the web," *Nature*, vol. 400, pp. 107–109, 1999.

[58] R. M. May, "How many species are there on Earth?," *Science*, vol. 214, pp. 1441–1449, 1988.

[59] J. E. Pitkow, "Summary of WWW characterizations," *Computer Networks and ISDN Systems*, vol. 30, pp. 551–558, 1998.

[60] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.

[61] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[62] A. Woodruff, P. M. Aoki, E. Brewer, P. Gauthier, and L. A. Rowe, "An investigation of documents from the World Wide Web," *Computer Networks and ISDN Systems*, vol. 28, pp. 963–980, 1996.

[63] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "The Origin of Power Laws in Internet Topologies Revisited," in *Proceedings of INFOCOM*, 2002.

[64] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications," in *SIGCOMM*, 2001.

[65] K. Park and H. Lee, "The effectiveness of probabilistic packet marking for IP traceback under denial of service attack," in *SIGCOMM*, 2000.

[66] W. Ashmawi, R. Guerin, S. Wolf, and M. Pinson, "On the Impact of Policing and Rate Guarantees in Diff-Serv Networks: A Video Streaming Application Perspective," in *SIGCOMM*.

[67] S.N. Dorogovtsev and J.F.F. Mendes, "Comment on "Breakdown of the Internet under Intentional Attack"," *Physical Review Letters*, vol. 87, 2001, arXiv:cond-mat/0109083.

[68] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, "Pseudofractal Scale-free Web," arXiv:cond-mat/0112143, 2001.

[69] S.N. Dorogovtsev, A.N. Samukhin, and J.F.F. Mendes, "Multifractal properties of growing networks," arXiv:cond-mat/0106142, 2001.

[70] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "Anomalous percolating properties of growing networks," arXiv:cond-mat/0106141, 2001, Phys. Rev. E 64, 066110 (2001).

[71] S.N. Dorogovtsev and J.F.F. Mendes, "Language as an Evolving Word Web," arXiv:cond-mat/0105093, 2001, Proc. Royal Soc. London B, 268, 2603 (2001).

[72] S.N. Dorogovtsev and J.F.F. Mendes, "Scaling properties of scale-free evolving networks: Continuous approach," arXiv:cond-mat/0012009, 2000, Phys. Rev. E 63, 056125 (2001).

[73] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "Generic scale of the "scale-free" growing networks," arXiv:cond-mat/0011115, 2000, Phys. Rev. E 63, 062101 (2001).

[74] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "Growing network with heritable connectivity of nodes," arXiv:cond-mat/0011077, 2000.

[75] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "WWW and Internet models from 1955 till our days and the "popularity is attractive" principle," arXiv:cond-mat/0009090, 2000.

[76] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, "Structure of Growing Networks: Exact Solution of the Barabási–Albert's Model," cond-mat/0004434, 2000, Phys. Rev. Lett. 85, 4633 (2000).

[77] S.N. Dorogovtsev and J.F.F. Mendes, "Scaling Behaviour of Developing and Decaying Networks," arXiv:cond-mat/0005050, 2000, EuroPhys. Lett. 52, 33 (2000).

[78] S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of reference networks with aging," arXiv:cond-mat/0001419, 2000, Phys. Rev. E62, 1842 (2000).

[79] R. Govindan and H. Tangmunarunkit, "Heuristics for Internet Map Discovery," in *INFOCOM (3)*, 2000, pp. 1371–1380.