ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computer Networks xxx (2006) xxx–xxx

Computer Networks

www.elsevier.com/locate/comnet

Internet resiliency to attacks and failures under BGP policy routing

Danny Dolev [a], Sugih Jamin [b], Osnat (Ossi) Mokryn [c,*], Yuval Shavitt [c]

[a] School of Engineering and Computer Science, The Hebrew University, Jerusalem 91904, Israel
[b] Department of EECS, University of Michigan, Michigan, United States
[c] School of Electrical Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

Received 4 August 2004; received in revised form 26 June 2005; accepted 7 November 2005

Responsible Editor: J. Hou

Abstract

We investigate the resiliency of the Internet at the Autonomous System (AS) level to failures and attacks, under the real constraint of business agreements between the ASs. The agreements impose policies that govern routing in the AS level, and thus the resulting topology graph is directed, and thus the reachability between Ases is not transitive. We show, using partial views obtained from the Internet, that the Internet's resiliency to a deliberate attack is much smaller than previously found, and its reachability is also somewhat lower under random failures. We use different metrics to measure resiliency, and also investigate the effect of added backup connectivity on the resiliency.

Keywords: Directed graph; AS relationships; Valley free routing

1. Introduction

In recent years there is a growing interest in the resiliency of the Internet, as it represents the network's availability in times of instabilities or under extreme conditions. Research in this field took two distinct paths. One is the stability of routing protocols in case of errors and failures [1,2], and the other, which also draws attention outside the computer networking community, focuses on the resiliency of the Internet to random failures and attacks on strategic locations [3–5]. Such events can happen as a result of a disaster, or a manipulated online attack on key Internet elements.[1] In this research we focus on the latter.

Research in the field was motivated by the finding that the Internet AS topology can be classified as scale free, belonging to a class of networks for which the connectivity resembles a power law

* Corresponding author. Tel.: +972 54 270020.
E-mail address: osnaty@eng.tau.ac.il (Osnat (Ossi) Mokryn).

[1] While the collapse of an entire large ISP seems unlikely, it actually happened a few times in the recent past for the largest AS in the Internet, UUNet. On April 22nd and October 3rd 2002 the UUNet network collapsed due to software problems in its routers, and in January 25th 2003 due to a DoS attack [6].

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2005.11.010

distribution [7–10]. In physics terminology, the susceptibility of the Internet to node deletion is considered in terms of network *phase transition*, representing the transition from a connected phase to a disconnected phase. The research in this field [3–5,11,12] showed that the Internet has a high tolerance to random failures, and does not break until more than 95% of the nodes have failed. On the other hand, it was found that the Internet is highly sensitive to deliberate attacks that target its most connected nodes. Under such an attack, the network transitions to a set of small disconnected components, after the removal of a small fraction of the highly connected nodes. Cohen et al. [4] have shown that the rate of transition under a deliberate attack depends on the minimal connectivity, hence on the average degree. They have also shown that the average path length grows dramatically under such an attack, almost approaching the critical point of transition in which the network disintegrates.

A significant drawback of the works in [3–5,11,12] is that they treat the Internet as an undirected graph. However, routing in the Internet between the ASs is governed by policies that are set locally with the aid of BGP, the inter-network routing protocol, according to business agreements [13]. The implication of policy-based routing is that not every two nodes (ASs) that have a physical path connecting them can indeed exchange information; a valid path that conforms to the policies of the ASs along it must exist. These considerations and agreements create a network far different from the one used in all the above listed works, and calls for revisiting the question on the resiliency of the Internet. In addition, the data used for obtaining the above results was of partial views of the Internet. These partial views, obtained mainly through dumps of BGP announcements, lack in connectivity due to two main reasons. The first is that these views are taken from a few sites in the Internet. While they contain most of the nodes, they lack in connectivity information, since they contain mostly links that are on the shortest BGP path from the source site to the other nodes [14]. The second lies with the rules of the BGP protocol, which tend not to advertise a backup path which is not in current use.[2]

In this work, we first suggest a paradigm for finding Internet connectivity under BGP policy routing based on existing business agreements. We discuss the different metrics suggested for measuring the resiliency of the network, and suggest our own. We find the resiliency of the Internet to attacks and random failures, and show that it is even more susceptible to attacks than previously found. We show that previous Internet models, which did not take into account the connectivity constraints imposed by policy-based routing, yielded too optimistic results for the case of a deliberate attack. In the case of random failures of nodes, the results show that the difference in resiliency is small.

Our testbed consists of partial Internet views obtained from the Oregon site [15] and from European exchange points [16]. We also obtained the very rich database collected by Chen et al. [14], who assembled 41 partial views along with added Looking Glass information and showed that the actual connectivity between ASs is higher than was previously known. Our results show that the added connectivity improves the resilience of the networks, and therefore results obtained on partial views are somewhat misleading. Moreover, hidden backup links which are used only in case of a disaster, would probably improve the resilience of the network even better. We made some first attempts to model how backup links may improve Internet reachability.

The paper is organized as follows. In Section 2 we give the background on autonomous systems connectivity and Internet topology. In Section 3 we discuss our model and present our reachability algorithm. We further discuss the metrics used for estimating resiliency and present the ones we use. Section 4 outlines our results on the resilience of the Internet. In Section 5 we discuss the added backup connectivity patterns and show how these patterns may influence the resilience of the Internet.

2. Background on AS connectivity and Internet topology

The Internet today consists of thousands of subnetworks, each with its own administrative management, called autonomous systems (ASs). Each such AS uses an interior routing protocol (such as OSPF, RIP) inside its managed network, and communicates with neighboring ASs using an exterior routing protocol, called BGP. The BGP protocol enables each administrative domain to decide which routes to accept and which to announce. Through the use of the protocol the autonomous systems

[2] BGP is a path vector protocol, that advertises preferred paths to a network prefix.

select the best route, and impose business relationships between them on top of the underlying connected topology. As a result, paths in the Internet are not necessarily the shortest possible, but rather the shortest that conform to the ASs' policies. Such routing is called policy-based routing.

The commercial agreements between the ASs create the following peering relationships: customer–provider and provider–customer, peer-to-peer, and siblings. A customer pays its provider for transit services, thus the provider transits all information to and from its customers. The customer, however, will not transit information for its provider. For example, a customer will not transit information between two of its providers, or between its provider and its peers. Peers are two ASs that agree to provide transit information between their respective customers. Such agreements are very common between ASs that connect at an exchange point (IX) and between smaller ISPs residing at the same geographical vicinity. In sibling relationships, the two ASs provide full transit services for each other. Such relationships are mainly due to financial acquisitions, mergers, or to a smaller degree, business transactions between smaller ISPs that maintain their own administration but unify their networking services.

In a pioneering work, Lixin Gao [17] suggested an algorithm for inferring the type of relationships between ASs through their advertised BGP paths. The algorithm assumes that the degree of connectivity of an autonomous system is an indication of its size, and infers the relationships between the ASs according to a set of rules obtained from the above description of commercial relationships. Gao has deduced, that a legal AS path may take one of the following forms:

1. *Up hill* path, followed by a *down hill* path.
2. *Up hill* path, followed by a peering link, followed by a *down hill* path.

Where an *up hill* path is a sequential set, possibly empty, of customer–provider links, and a *down hill* path is a sequential set, possible empty, of provider–customer links. Thus, a legal route between autonomous systems can be described as a *valley free* path. A peering link can be traversed only once in each such path, and if it exists in the path it marks the turning point down hill.

Further work on AS relationships [18] have characterized the Internet as hierarchical. They found that the top big American providers form a core with almost complete clique connectivity, and the second layer around this core consists of big providers from the USA and Europe, characterized mainly by their very rich connectivities to the core. The third layer consists of smaller providers, and forms the majority of the network. Recent works have investigated the relations between ASs, looking for anomalies and their possible solutions [19,12].

Inferring the AS relationships can be viewed as part of an ongoing effort to discover and map the exact topology of the Internet [7,10,9,14,20,8]. It is generally agreed today that the Internet, at the AS level, has a highly heterogeneous connectivity patterns, with a highly variable vertex degree distribution. Several works have also tried to characterize the growing mechanisms of the Internet and model it [21,22,11,12], and several network generators which rely on some of these algorithms exist [23–25] and evaluated [26–28,11].

In all previous works on the resilience of the Internet, it was assumed that the connectivity of the network is equivalent to its reachability. We show in this work that the two are not equivalent, and find the actual reachability of the network under different constraints.

3. Modeling reachability in a directed AS graph

In this section we characterize our graph model, and describe our reachability algorithm.

3.1. AS graph model

We model the AS graph as a *directed* graph, in which the set of nodes is the set of distinct autonomous systems and a link exists between two such nodes if the respective ASs have peering (business) relationship[3] and are BGP neighbors. For each link we maintain its direction and characteristics. For example, between two nodes that represent a provider and its customer, there will be an *uphill* link from the customer to the provider, and a *downhill* link from the provider to the customer. Between peers there is a directed peer edge in each direction, and between siblings there is an undirected link.

[3] Note that the derivatives of "peer" appears in two distinct meaning. We say that two ASs have "peering relationship" if they exchange BGP messages, and that two ASs are "peers" if they have peer-to-peer exchange agreement in BGP, namely if they are neither provider–customer nor siblings.

Connectivity in the AS graph with the valley free policy rules described in Section 2 maintains reflexivity, but does not maintain transitivity. For example, a small ISP with two providers reaches each of them on the directed link that connects the customer to its provider, but the providers cannot use the two link path through the customer to communicate. An algorithm for finding the shortest path under these restrictions was suggested in [29]. The algorithm uses an adaptation of the Dijkstra shortest path algorithm to the AS graph, for the problem of proxy and cache location.

3.2. AS reachability algorithm

The reachability algorithm we developed maintains a reachability map. It finds, for each node, the set of nodes that can be reached from it in the policy-constrained AS graph, regardless of the path taken. The algorithm does not look for the best path to a node, but rather, for each node, looks for all nodes reachable from that node through *some valid* AS path.

The algorithm is a free adaptation of a BFS algorithm to the AS graph. Instead of looking for a shortest path, though, the algorithm is looking for a valid AS path. Such a path is taken only if by taking it new parts of the graph can be discovered.

Starting from a source node, the algorithm looks first for valid uphill paths and only then for peer links and downhill paths. Each node, when reached for the first time, marks its state by the direction it was reached with. Thus, a node reached through a downhill path is marked as *down*, etc. Then, the node examines all of its neighbors. A link to a neighbor is taken only if it provides a valid AS path and the state of the neighbor is improved according to the following ascending order: NONE, DOWN, SIDE, UP. Here is a description of these possible states:

none: The node has not been traversed yet.
up: The node was in either *none*, *side* or *down* states, and there is an uphill path that can be traversed through it.
side: The node was in either *none* or *down* state and there exists a peer link that can be traversed through it.
down: The node was in *none* state, and there is a downhill path that can be traversed through it.

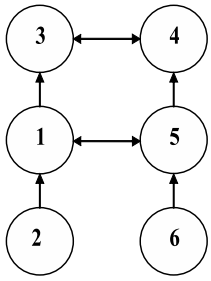


Fig. 1. Example for the reachability algorithm.

The algorithm gives the highest priority to an uphill path through a node, the next priority to traversing a peer-to-peer link from that node, and the lowest priority to a downhill path through the node. Each node, once reached, examines all of its links. A link is taken only if by taking it the state of the node reachable through it can be improved, according to the description above. The algorithm exploits the fact that uphill links reach higher providers in the Internet hierarchy, hence more nodes are reachable through them. For example, Fig. 1 describes a small network in which ASs 3 and 4 are top providers, each has one customer—1 and 5, respectively. ASs 1 and 5 are also peers, and also each of them has a customer—2 and 6, respectively. If we start the algorithm from AS 1, it will traverse its adjacent uphill path to its provider, AS 3, and thus will be able to first discover the largest part of the network, consisting of nodes 3, 4, 5, and 6. As the last step will discover its own customer, 2. Note, that although ASs 1 and 5 are peers, taking this route first will only enable AS 1 to reach AS's 5 customers, and therefore it should not be taken unless there is no valid AS path starting with an uphill path and reaching AS 5.

Fig. 2 presents a formal description of the algorithm. The following variables are used in the algorithm: V is the set of all nodes representing ASs in the graph; R_i is the reachability bitmap of node i, in which bit j is a set if there is a legal BGP path between node i and node j, and st_i is the state of node i. N_i is the set of immediate BGP neighbors of node i.

A proof of the correctness of the algorithm is given in Appendix A.

The algorithm time complexity is as follows. Each node starts in state *none*, and can change its state at most three times. Hence, each node is reached *at most* three times, giving a worst case time complexity of $O(|E|)$, where E is the number of links

Algorithm 1 *Reachability Algorithm*

```
∀v ∈ V do:
    set s ← v; direction ← up;
    set R_s ← ∅; st_s = up;
    inspect(s, direction)
    output R_s

function inspect(k,d)
set k in R_s
∀i ∈ N_k do:
    if < k, i > = sib_link then
        inspect(i, d)
        return
    switch d:/*Note the fall through*/
    case up:
        if < k, i > = customer_to_provider then
        if st_i = none or side or down then
            st_i ← up
            inspect(i,up);
    case side:
        if < k, i > = peer then
        if st_i = none or down then
            st_i ← side
            inspect(i,down);
    case down:
        if < k, i > = provider_to_customer then
        if st_i = none then
            st_i ← down
            inspect(i,down);
    return
```

Fig. 2. A formal description of the basic algorithm for the root node.

in the graph (in a worst case scenario, each link is examined three times).

3.3. Anomalies in the AS graph

The algorithm described in Fig. 2 is resilient to anomalies in the AS graph. However, there are two anomalies that need to be considered. The first, and more rare, is called a black hole, and is out of the scope of this paper. The second anomaly, and the more interesting for us, is inference mistakes.

Gao's inferring algorithm was shown to be 97% accurate on a test case database of AT&T, having inference problems only for links suspected as siblings. Out of the 3% inferred as sibling links, the actual relationships obtained from the AT&T data were almost half peering links, a quarter customer–provider links, and only the rest were actual sibling links. Battista et al. [19] have investigated the anomalies in AS graphs, showing that the problem of solving the AS relationships while minimizing the anomalies is NP-hard in the general case, and suggested heuristics for minimizing the number of anomalies. A recent work [30], that compares traceroutes to BGP AS paths, finds that much of the disparity results from ASs connected through exchange points, and by groups of ASs under the same ownership.

To obtain accurate results, we inferred manually through the use of WHOIS servers and Internet searches all of the automatically inferred sibling relationships in the databases obtained from [16]. For a combined view of the London and Zurich exchange points, gathered at the same time for the same set of ASs, we obtained the following results: Out of the 81 inferred sibling relations, only 32% (26) were actual siblings. 27% (22) were peers, 8% (7) were customer-to-provider links and 32% (26) were provider-to-customer links.

3.4. Metrics for defining resiliency

The problem of finding the right metric for evaluating the network resiliency was reduced in previous works to the problem of finding the connectivity of the graph [3,5,4,12]. Although the problem itself remains an open problem [7], the above mentioned works used some of the following metrics: Average diameter or average shortest path length $\bar{d}$; the giant component size S; the number of connected node pairs in the network, K; diameter-inverse-K, DIK.

The definition of $\bar{d}$ is as follows: let $d_{\min}(v,u)$ denote the minimal path between any connected pair of distinct nodes u and v, and Π the set of such distinct node pairs. Then: $\bar{d} = \frac{\sum d_{\min}(v,u)}{|\Pi|}$. According to [4] $\bar{d}$ can be used to asses when a network under attack reaches criticality. A measure of the size of the largest component, S, is the ratio between the number of nodes in the largest connected component and the number of nodes in the graph. The two metrics K and DIK, defined in [12], are as follows. K describes the whole network connectivity, by measuring all connected node pairs in a network: let Ψ be the set of all distinct node pairs, and Π defined as above, then: $K = \frac{|\Pi|}{|\Psi|}$. Park et al. [12] have suggested a different version of K, DIK, which measures both the expected distance between two nodes and the probability of a path existing between two arbitrary nodes: $DIK = \frac{\bar{d}}{K}$.

Table 1
Characteristics of data sets used

Name	source	Date	No. of ASs	No. of links	Average degree	Maximum degree
LZ	RIS	2002/07/03	13393	22001	3.28545	1958
OR1	Oregon	2003/03/01	14704	24020	3.26714	2330
OR2	Oregon	2003/04/01	15128	31426	4.15468	2503
UM	Umich	2001/05/26	11204	25980	4.63763	2417

We noted that the measures described above cannot be directly applied in our case, when reachability is *not* equivalent to connectivity, since the directed AS graph lacks transitivity. In this case, for example, the minimal distance between two nodes, $\bar{d}$, becomes the minimal BGP distance between two nodes, depending on policy constraints. Thus, we chose two different ratios, that capture best, in our understanding, the actual resilience of the Internet.

The first, denoted by R, captures the reachability of the Internet, and is defined as follows: let $r(v,u) \in 0,1$ denote the reachability between an arbitrary distinct pair of nodes v and u, $v,u \in V$, where V is the set of nodes describing ASs. Let Π_r denote the number of distinct node pairs in the graph, for which $r = 1$, and let O_r denote the theoretical limit of Π_r for the Internet (when there are no failures in the Internet we expect to have full reachability between all ASs). Then, we define R as the ratio:

$$R = \frac{\Pi_r}{O_r}.$$

The second metric quantifies the size of the strongly connected component in the directed AS graph, termed RS. We create a *reachability graph*, in which there exists an edge between two nodes v and u if and only if $r(v,u) = 1$.[4] Then, in order to find the largest strongly connected component in the original graph, we need to find the maximal clique in the reachability graph. The problem of finding the maximal clique in a graph is NP-complete [31]. The best known approximation for finding the maximal clique [32] gives an $O(n/(\log_n)^2)$ performance guarantee. Hence, for our topologies, we can expect a maximal mistake of 5.7–6% (see Table 1). We use a greedy heuristic for finding the maximal clique in the graph. Since we know which nodes still exist in the graph after the simulated failure or attack, and their respective degrees, we start with the one with the largest degree. Due to the hierarchical nature of the Internet [27], it is likely that such a node resides in the core, and therefore is used by many other nodes for reachability. We denote all of the nodes reachable from that node by C. Then, iteratively, we look for the maximal degree node in C, i, and extract from C all the nodes not reachable from i. We continue this process until all the nodes in the component are reachable from each other. The process is repeated several times with different starting nodes selected from the top connected ones. The size of the strongly connected largest component, S_d, is then divided by the number of nodes in the original graph, to obtain the ratio RS.

3.5. Critical point of failure (phase transition)

From a physics point of view, a phase transition occurs only when the network disintegrates [4]. The network is considered connected as long as RS, the ratio between the size of the largest component and the number of initial nodes in the graph, is a fraction of the number of nodes in the graph. For example, the removal of the top 20% of the nodes of a 100 nodes network, yields $RS = 0.2$. For a network with the same connectivity distribution, regardless of its size, any such removal of the top 20% of the nodes will yield a similar RS. Thus, as long as the size of the largest component is a fraction of the initial size of the network, the network is considered connected. The phase transition occurs when order $(RS) = 1$. Hence, physically speaking, the network is considered disintegrated only when the size of the largest component is one. The same discussion holds for the reachability function, R.

From a routing perspective, reachability is considered lost long before the Internet disintegrates. We arbitrarily assume here, that when $R < 0.5$, i.e., the overall reachability is less than 50% of the original reachability, or when $RS < 0.5$, i.e., the comparable size of the largest component is half the original network, the network is no longer considered connected.

[4] Note that while reachability is not transient, it is symmetric under the valley free rule.

4. Resiliency of the Internet

In this section we present our results for the resiliency of the Internet to random failures and attacks, given the policy routing constraints.

Table 1 describes the different data sets used in these tests and their characteristics. The topologies differ mainly in their connectivity. The LZ dataset, from the RIPE Routing Information Service [16], is the result of combining routing information from two exchange points, one in London and the other in Zurich. The data lacks most of the largest top US providers. The largest AS in this data set has a rather low degree of 1958, and the average degree in the set is also rather low. This implies that there are fewer alternative paths between the nodes in this topology, i.e., less redundancy, and therefore we expect it to be the most vulnerable to deliberate attacks. As discussed in Section 3.3 the topology is also siblings inference-anomaly free, as all automatically inferred sibling relations were manually checked using WHOIS databases and Internet searches. Datasets OR1 and OR2 are both partial views from the Oregon routeview project [15], collected March and April, 2003 respectively. The topologies differ greatly in the richness of the connectivity, as OR2 has 27% added connectivity compared to OR1. The last view, and the richest in connectivity, UM, is the enriched topology obtained by Chen et al. [14]. Although collected three years ago, the topology is the richest in connectivity, since it was collected from 41 BGP databases and augmented with summary data from different looking glass sites. The ongoing growth of the Internet, which increases its average degree, implies that such an enriched view of today's Internet will yield a much higher average degree than seen from the partial views OR1 and OR2. We examine the above four topologies, in an increasing order of their average degree and hence in their connectivity, in the intent to find a tendency, that may hint as to how a richer topology, as the Internet is thought to be, will actually behave.

All data sets presented here where first analyzed for their relationships with Gao's algorithm, then presented as directed topologies. We then ran for each such directed topology the reachability algorithm described in the previous section, as well as the algorithm for determining the largest clique.

In the graphs presented in this section, we compare the resiliency of the policy-constrained AS graph, referred to as the *directed* graph or the *reachability* graph to the resiliency of the graph used in previous works, referred to as *undirected* graph. For each topology, we present both the reachability R and the evaluation of the largest component, RS, as discussed in Section 3.4. Some of the partial views do not have 100% *reachability* to begin with, as can be seen in Fig. 3, for example.

Section 3.4 describes the different metrics we use here, and establishes that the largest component can be derived with a 6% error mistake upper bound. The monotonic increasing nature of our data suggests that in practice we may expect a much lower error mistake upper bound.

4.1. Resiliency of the Internet to deliberate attacks

We evaluate the resiliency of the Internet to deliberate attacks by targeting the topology's most connected ASs, dropping each time the next most connected node in the graph, and measuring both metrics R and RS for the directed and undirected graphs.

Figs. 3 and 4 show the resiliency of the LZ topology to deliberate attacks. Even before any node was dropped, the reachability is less than the connectivity, due to the partiality of the topology. The same can be seen in Figs. 5 and 6, representing the resiliency of topology OR1, also a rather sparse partial view. However, in the more connected views, OR2 and UM, the partial view gives a fully connected network, in which all nodes are reachable to begin with.

In the sparse topologies the overall reachability decreases very fast. Fig. 3, which starts from a

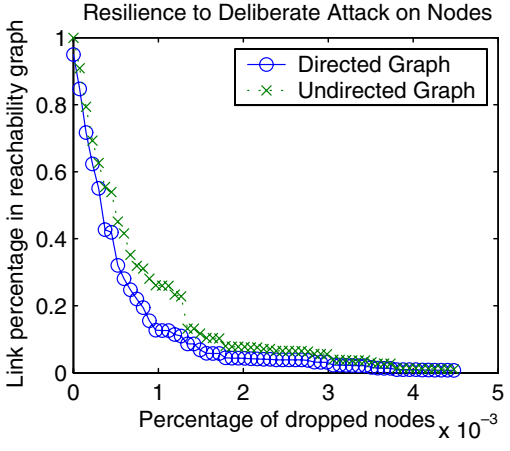


Fig. 3. Reachability under attacks in LZ.

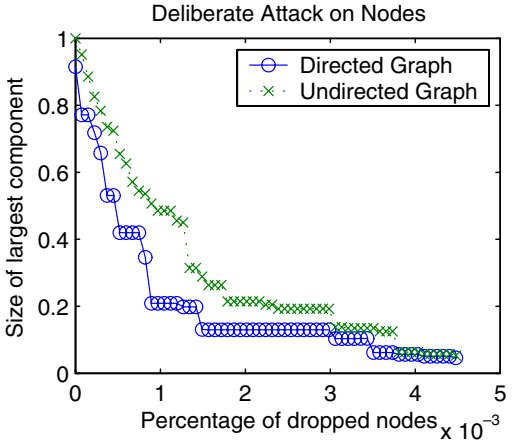


Fig. 4. Largest component size under attacks in LZ.

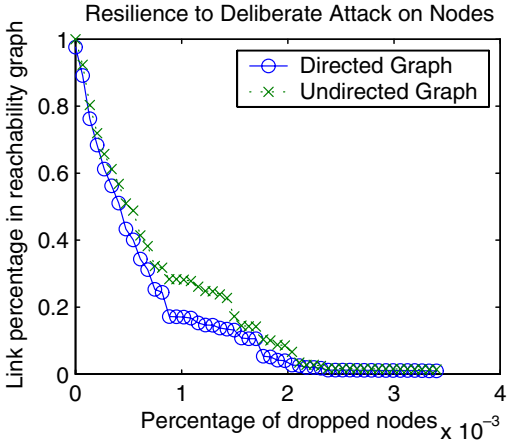


Fig. 5. Reachability under attacks in OR1.

95% reachability, shows that after dropping only the sixth most connected nodes, reachability drops to 42%, while the connectivity in the undirected graph is 12% higher, 54%. The gap is even larger when we check how much of this reachability is within the same component of nodes that communicate with each other (Fig. 4). After the removal of these six nodes, only 53% of the nodes are connected, while in the undirected graph, the largest component consists of 72% of the nodes, an evaluation error of 19%. The gap between reachability and connectivity increases as the network starts to break up—after dropping the 12th most connected nodes the largest component consists of only 20% of the nodes in the topology, while previously it was thought that it still consists of 50% of the nodes, as we can see from the results for the undirected graph.

Figs. 5 and 6 give similar results, namely, that the Internet is much more susceptible to deliberate attacks than previously thought. While the overall reachability drops at the same rate as the connectivity, it can be seen from Fig. 6 that the first node that was dropped was a large AS with a lot of customers, that lost reachability to the rest of the network. After the eighth most connected nodes were removed, the size of the largest component is less than 50% the size of the network, while in the case of the undirected graph it contains 69% of the nodes. We see here that after attacking only the 8th most connected nodes, the Internet's largest component contains less than 50% of the nodes.

Figs. 7–10 represent the resiliency of highly connected topologies (OR2 and UM), in which most nodes can be reached through several AS paths.

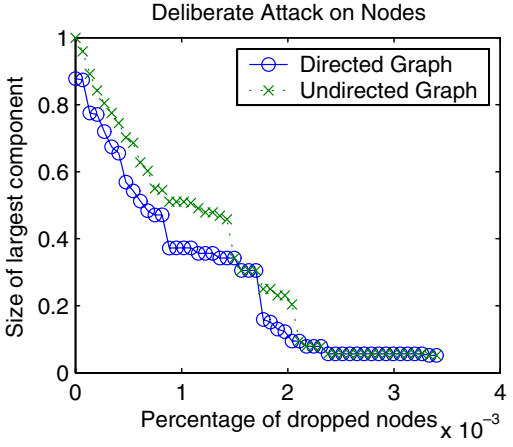


Fig. 6. Largest component size under attacks in OR1.

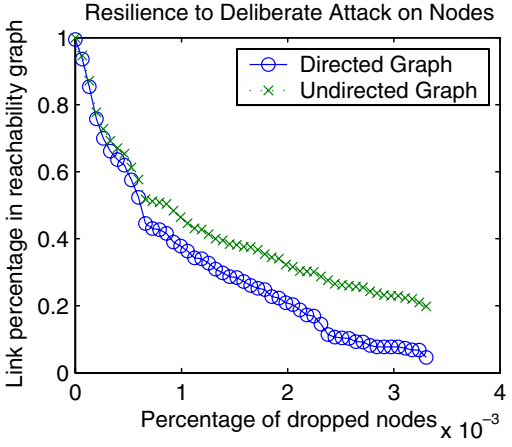


Fig. 7. Reachability under attacks in OR2.

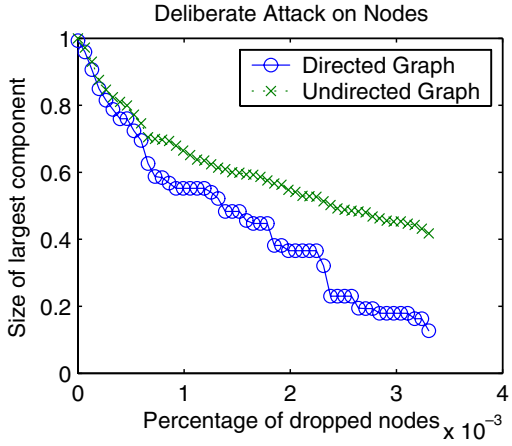


Fig. 8. Largest component size under attacks in OR2.

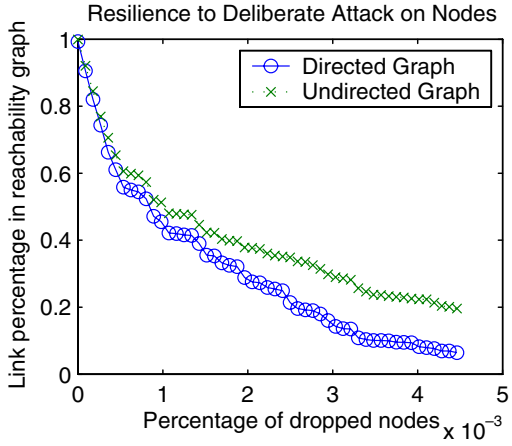


Fig. 9. Reachability under attacks in UM.

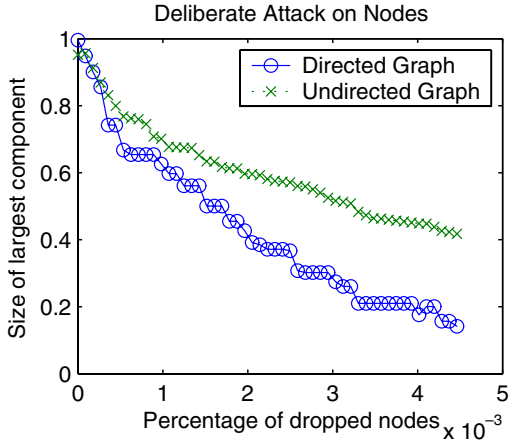


Fig. 10. Largest component size under attacks in UM.

Therefore, we expected that the resiliency of the directed AS graph would resemble the one of the undirected graph. Indeed, the reachability R is almost the same for both topologies when the five most connected nodes are dropped, since alternate paths are taken. The size of the largest component is also quite similar for both the directed and undirected graphs, although for the UM topology (Fig. 10) the gap between the component sizes reaches almost 7% after the removal of only five nodes. In all cases, the gap between the directed and undirected graphs increases after the ten most connected nodes are removed. After the removal of 28 nodes the gap in largest component size is over 15% (Figs. 7 and 10). After removing the 50 most connected nodes, for the highly connected OR2 topology, the network disintegrated to the point where the largest component holds only 4% of the nodes. In the unconnected case, the component holds more than 42% of the nodes, an order of magnitude difference.

The results on the UM topology, in Figs. 9 and 10 are similar, and show a constant rate of decrease in the largest component size, which reachability decreases fast at the beginning and then at a slower rate. decrease in reachability, followed by a rapid break down of the graph to small isolated islands. The above results may imply, that the medium-sized ASs tend to not rely only on one large provider, and mulithome to several top providers, to obtain maximal reachability in case the connection to one of the providers fails. Once these top providers were removed, the network became much more susceptible to attacks, and disintegrated very fast.

4.2. Resiliency to random failures of nodes

We checked the resiliency of the Internet to random failures by a random removal of 100 nodes at a time, until more than 95% of the nodes were removed.

Figs. 11 and 12 show the comparable resiliency of the Internet to random failures for the LZ topology. As previously found, the Internet is not susceptible to such random failures, and both R and RS do not fall below 0.8 even after the removal of 1000 nodes. The network starts to break down only after the removal of more than 2000 random nodes. The Internet disintegrates only after the removal of almost 95% of the nodes. The difference between the two graph models, the directed (policy-constrained) AS graph, and the undirected graph, is

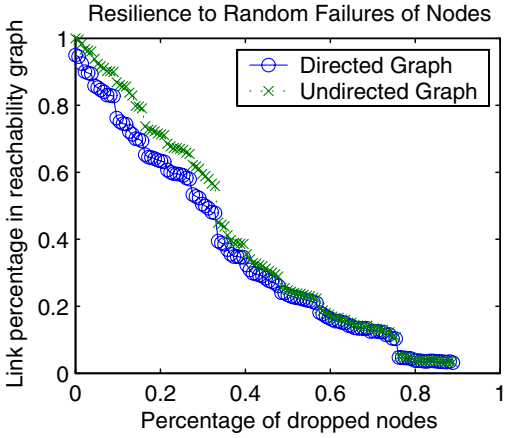


Fig. 11. Reachability under random failures in LZ.

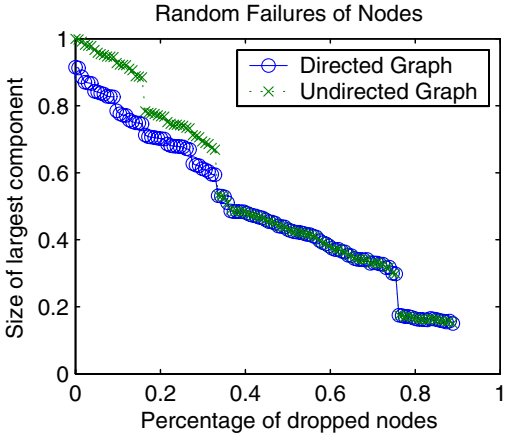


Fig. 12. Largest component size under random failures in LZ.

small. However, we found that the gap is larger for the sparse views than for the views richer in connectivity, where it is negligible. This result is somewhat surprising, indicating that the Internet maintains reachability of almost the same degree as its connectivity, under random failure of nodes.

Due to the high degree of the nodes in the core, and the fact that these nodes are rare in a scale free distribution, the statistical probability that they will be removed in a random failure scenario is low. However, it could be expected that the removal of small and medium-sized nodes will effect the reachability of the smaller ASs and therefore the size of the largest connected component. The surprising results, indicating that the reachability is very close to the possible limit, the undirected connectivity, prove differently. These results may indicate that most ASs use multihoming to several providers, and thus are less susceptible to these random failures.

5. A heuristic for added backup connectivity

Our results show that the Internet is significantly more susceptible to attacks than previously found. On the other hand, the Internet's resiliency to failures is higher than expected, and resembles the connectivity. These results lead us to point to the core of the Internet as the main transit point. It seems that the overall reachability, as well as the largest component size, depend on the level of connectivity of the nodes in the network to the top provider nodes in the core, and to a lesser degree on the connectivity to other non-top provider nodes. However, there is a lack of knowledge on existing backup links which many times are not advertised through BGP until used.

In this section, we make a first attempt to quantify the effect of existing backup links, which are usually not advertised through BGP until used, on the reachability and resiliency of the AS graph under attacks. We constructed a backup scenario, which relies on the existing connectivity, and provides alternate paths to small- and medium-sized ASs which connect only to one provider. These ASs, once their provider fails, use their peering links as backup links, effectively using them as customer-to-provider links. Thus, these ASs get connectivity to the network through their previous peer. Since we do not add links to the existing graph, the effect of such a backup scenario is only meaningful in the case of attacks. As we have shown in Section 4.2, Internet reachability under random failures is very close to its connectivity. Therefore the added paths gained from using the backup links can hardly improve the resiliency in this case. However, in the case of attacks, it might allow single-homed ASs to use alternate paths. If there are many such ASs, which do not rely on multihoming, we expect an increase in both the size of the largest component and the reachability.

Our backup scenario is as follows:

- AS x has one provider.
- link $\langle x, y \rangle$ is a peer link.
- if AS x disconnects from its provider, then link $\langle x, y \rangle$ becomes a customer to provider link.

We found that the added backup connectivity is more meaningful for the sparse topologies (LZ,

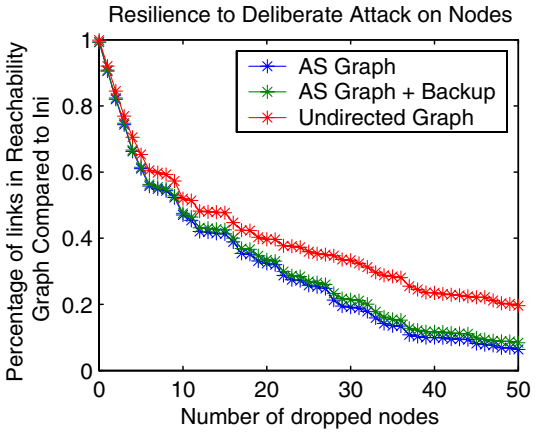


Fig. 13. Added backup: Reachability under attacks in UM.

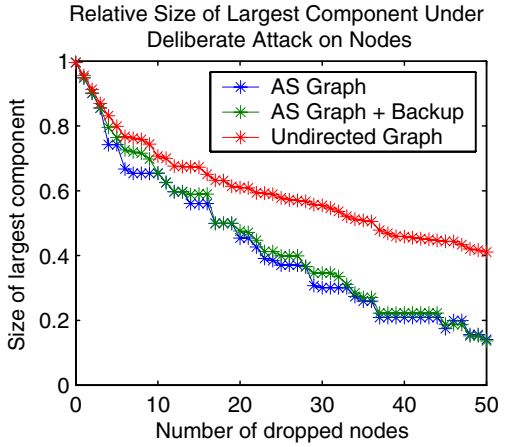


Fig. 14. Added backup: Largest component size under attacks in UM.

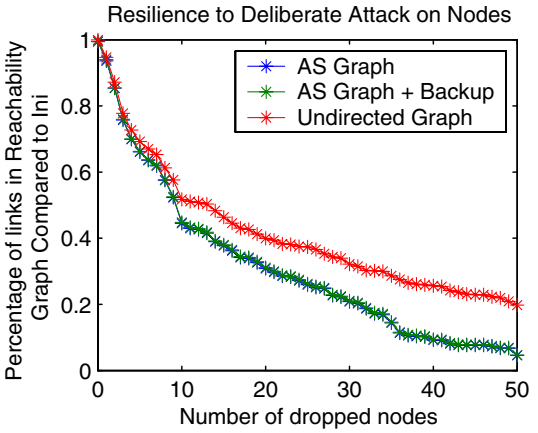


Fig. 15. Added backup: Reachability under attacks in OR2.

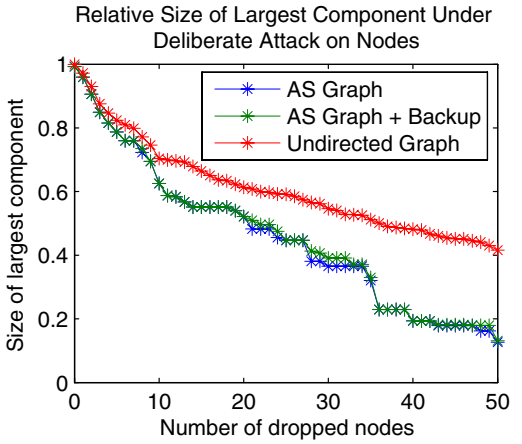


Fig. 16. Added backup: Largest component size under attacks in OR2.

OR1) than for the richer topologies (OR2, UM). We give here the results for the UM topology in Figs. 13 and 14. We see that the reachability is somewhat better with the added backup links, as is the size of the largest component, but the behavior of the topologies with and without the backup links is very similar. Figs. 15 and 16 show that for the newer partial-view OR2, which is also rich in connectivity, the reachability and size of the largest component are hardly effected by the added backup connectivities. These results suggest that there has been a vast increase in the number of ASs that use multihoming (as was also shown, for example, in [33]), and are therefore hardly affected if one of their providers fails. As a result, the number of singly homed ASs is rather small, and their affect on the Internet connectivity is rather small.

A more strict backup scenario, which enabled two ASs to use a peer link between them as backup only if both ASs have only one provider, yielded even smaller improvement in both reachability and the size of the largest component. Additionally, we examined the resiliency to a partial attack on the core, i.e., a few of the 30 most connected nodes were removed randomly. We examined the effect of the backup links in this scenario. The results showed no improvement in the size of the largest component, and a negligible improvement in reachability.

6. Conclusions

We examine the resiliency of the Internet to deliberate attack and random failures at the AS level, given that routing paths conform with the policy imposed by BGP. We compare our findings with

previous findings that did not consider these constraints, and evaluated reachability as connectivity. We suggest an efficient algorithm that determines reachability in such AS graphs, and discuss and suggest metrics for measuring the resiliency.

Our results show that the Internet is much more susceptible to deliberate attacks than previously found, and that reachability, as well as the size of the largest component, drop to less than half after the removal of the 25 most connected nodes—less than 0.2% of the nodes. The Internet also disintegrates much faster than previously found, under an attack that targets the top 0.5% ASs. We also found that the Internet is rather resilient to random failures, and its reachability is surprisingly close to the graph connectivity without policy constraints. These results can be attributed to that routing in the Internet is mainly through its core of highly connected ASs.

Our initial results on the effect of backup links suggests that they do not improve resiliency of the Internet by much. The decrease in the added resiliency of the partial views over the years suggest that ASs tend today to rely more on multihoming, and thus are less susceptible to a failure of one of their providers. We believe that further research to model backup connectivities at the AS level is important. In a future work, we plan to further validate our results using the DIMES project database [34]. The DIMES project contains a fuller view of the Internet's current topology.

Acknowledgements

This research was supported in part by a grant from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel; by a grant from the Israel Science Foundation (ISF) center of excellence program (grant number 8008/03); by a grant from the EU 6th FP, IST Priority, Proactive Initiative "Complex Systems Research", as part of the EVERGROW integrated project, and by a grant from the Israel Internet Association.

Appendix A

Theorem 1. *The reachability algorithm is correct.*

Proof. We shall first prove that the algorithm finds all the reachable nodes. A marking of a node as up, side, or down means it is reachable.

Select a node s. Lets first examine a node v for which there exists a path comprised only of up links. Clearly all the nodes in this path should be marked as reachable from s. Suppose for the contrary that v is not marked as reachable, and let u be the closest node to s on the path to v for which the state variable, st_u, is not up. By the assumption, the node before u is marked as up, and thus it is reachable. But by the inspect procedure the node must mark all its neighbors, with an up link connecting them, as up and inspect them, and thus it is impossible for u not to be in state up.

Now suppose that node v has a path with (possibly zero) up links and down links. Let u be the last node in the climbing part of the path. If the path has no up links $u = s$. As we proved above, u is bound to be in the up state. In case $u = s$, s is initialized to be in the up state. Let w be the first node on the down part of the path which is not marked as reachable. Examine the node before w on the path, which by the assumption is marked as reachable. Due to the fall through in the case statement of the inspect procedure, regardless of the state this node is in it will mark w in state down and activate inspect for it, contrary to the assumption, thus it is impossible for w not to be marked as reachable.

Finally, assume the path to v has a peer-to-peer link. Let this link be (u_1, u_2). We proved that u_1 will be marked as up. Based on the inspection procedure u_2 will be marked as side and be inspected in the down direction. Thus the downwards part of the path will be examined like proved above for the case of no side link and all nodes along it will be found reachable.

To complete the proof we must show that no node, v, which is not reachable from s will be marked erroneously as reachable. We will show that v's marking is correct, namely that it is marked as up only if there is a path leading to it comprised of only up links, as side if there exist a path leading to it comprised of only up links and the last link is side, and as down if the path leading to it contain a down link.

Let v thus be erroneously marked as up. Clearly, if no neighbor of v is marked as up this cannot happen since only nodes in state up can mark their neighbors as up. Let $p(v)$ be v's neighbor who marked it as up. Clearly $p(v)$ state must be up as well and there is an up link between $p(v)$ and v. Now examine the path $v, p(v), p(p(v)), \ldots$ If path reaches a node which is correctly marked as up, then all the

nodes in the path are correctly marked as up, which contradicts the assumption v is erroneously marked as up. Otherwise, either there must be a node that does not have a neighbor in the up state, or the path is cyclic. The first option is impossible since only nodes in the up state can mark their neighbor as up. The second option is impossible by the definition of $p(v)$ and the ordering of the marking times. Thus all the up markings are correct.

Clearly all the side markings are correct since only nodes whose neighbors are marked as up and have a peer-to-peer link to them can be marked as a side.

The nodes that need to be marked as down are correctly marked since we showed before that all the reachable nodes are marked as such, and we showed they cannot be erroneously marked as up or side. To show that no unreachable node is marked as down, we see that only nodes that have a down link can be marked as down by a reachable neighbor (at any state). As before we can look at a chain of nodes $v, p(v), p(p(v)), \ldots$ were $p(v)$ is the node that marked v as down first. The chain cannot exist using the same rational as before. □

References

[1] C. Labovitz, G.R.Melan, F. Jahanian, Internet routing instability, in: ACM SIGCOMM 1997, 1997.

[2] T. Griffin, G.T. Wilfong, An analysis of BGP convergence properties, in: ACM SIGCOMM 1999, 1999, pp. 277–288.

[3] R. Albert, H. Jeong, A.-L. Barabási, Attack and error tolerance of complex networks, Nature 406 (2000) 378.

[4] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Breakdown of the internet under intentional attack, Physical Review Letters 86 (2001) 3682.

[5] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Resilience of the internet to random breakdowns, Physical Review Letters 4626 (2000) 85–89.

[6] "WorldCom's IP Outages: Whodunnit?" April 25th, 2002; "WorldCom Outage Only the Start" October 14, 2002 ; "The Internet Has Broken" January 25, 2003. Available from: <www.lightreading.com>.

[7] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, in: ACM SIGCOMM 1999, Boston, MA, USA, 1999.

[8] R. Govindan, H. Tangmunarunki, Heuristics for internet map discovery, in: IEEE Infocom 2000, Tel-Aviv, Israel, 2000, pp. 1371–1380.

[9] A. Medina, I. Matta, J. Byers, On the origin of power laws in internet topologies, ACM Computer Communications Review 30 (2) (2000) 18–28.

[10] W. Cheswick, J. Nonnenmacher, C. Sahinalp, R. Sinha, K. Varadhan, Modeling internet topology, Tech. Rep. Technical Memorandum 113410-991116-18TM, Lucent Technologies, 1999.

[11] T. Bu, D. Towsley, On distinguishing between internet power law topology generators, in: IEEE Infocom 2002, New York, NY, USA, 2002.

[12] S. Park, A. Khrabrov, D. Pennock, S. Lawrence, C.L. Giles, L.H. Ungar, Static and dynamic analysis of the internet's susceptibility to faults and attacks, in: IEEE Infocom 2003, San-Francisco, CA, USA, 2003.

[13] J.W. Stewart III, first ed.BGP4 Inter-Domain Routing in the Internet, vol. 1, Addison-Wesley, 1999.

[14] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, The origin of power-laws in internet topologies revisited, in: IEEE Infocom 2002, New York, NY, USA, 2002.

[15] University of oregon route views project. Available from: <http://www.antc.uoregon.edu/route-views/>.

[16] Routing information service. Available from: <http://data.ris.ripe.net/>.

[17] L. Gao, On inferring autonomous system relationships in the internet, in: IEEE Global Internet, 2000.

[18] L. Subramanian, S. Agarwal, J. Rexford, R. Katz, Characterizing the internet hierarchy from multiple vantage points, in: IEEE Infocom 2002, New York, NY, USA, 2002.

[19] G. Battista, M. Patrignani, M. Pizzonia, Computing the types of relationships between autonomous systems, in: IEEE Infocom 2003, San-Francisco, CA, USA, 2003.

[20] H. Burch, B. Cheswick, Mapping the internet, IEEE Computer 32 (4) (1999) 97–98.

[21] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.

[22] R. Albert, A.-L. Barabási, Topology of evolving networks: local events and universality, Physical Review Letters 85 (24) (2000) 5234–5237.

[23] C. Jin, Q. chen, S. Jamin, Inet: Internet topology generator, in: Technical Report CSE-TR-433-00, University of Michigan, EECS Department. Available from: <http://topology.eecs.umich.edu>, 2000.

[24] A. Medina, A. Lakhina, I. Matta, J. Byers, Brite: an approach to universal topology generation, in: In Proceedings of MASCOTS 2001, IEEE Computer Society, 2001.

[25] D. Dolev, O. Mokryn, Y. Shavitt, On multicast trees: structure and size estimation, in: IEEE Infocom 2003, San-Francisco, CA, USA, 2003.

[26] P. Radoslavov, H. Tangmunarunkit, H. Yu, R. Govindan, S. Shenker, D. Estrin, On characterizing network topologies and analyzing their impact on protocol design, Technical Report 00-731, Department of CS, University of Southern California, 2000.

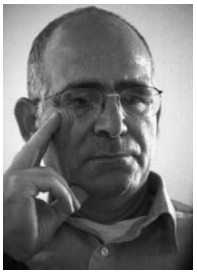
[27] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, W. Willinger, Network topology generators: degree based vs. structural, in: Proc. of ACM SIGCOMM 2002, Pittsburg, PA, USA, 2002.

[28] M. Mihail, A. Saberi, E. Zegura, Graph theoretic enhancements of internet topology generators, in: DIMACS Workshop on Internet and WWW Measurement, Mapping and Modeling, Piscataway, NJ, USA, 2002.

[29] K. Kamath, H. Bassali, R. Hosamani, L.Gao, Policy-aware algorithms for proxy placement in the internet, in: ITCOM 2001, Denver, CO, USA, 2001.

[30] Y. Hyun, A. Broido, K. Claffy, Traceroute and bgp as path incongruities, Cooperative Association for Internet Data Analysis—CAIDA, San Diego Supercomputer Center, University of California, San Diego, USA, 2003.

[31] R.M. Karp, Reducibility among combinatorial problems, in: R. Miller, J. Thatcher (Eds.), Complexity of Computer Computations, Plenum Press, 1972, pp. 85–103.
[32] R. Boppana, M.M. Halldórsson, Approximating maximum independent sets by excluding subgraphs, BIT 32 (1992) 180–196.
[33] G. Huston, The changing structure of the internet, in: APECTEL23, 2001.
[34] E. Shavitt, Y. Shir, Dimes: let the internet measure itself, ACM Computer Communication Review 35 (4) (2005).

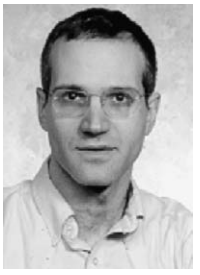


Danny Dolev (SM'89) received his B.Sc. degree in mathematics and physics from the Hebrew University, Jerusalem in 1971. His M.Sc. thesis in Applied Mathematics was completed in 1973, at the Weizmann Institute of Science, Israel. His Ph.D. thesis was on Synchronization of Parallel Processors (1979).

He was a Post-Doctoral fellow at Stanford University, 1979–1981, and IBM Research Fellow 1981–1982. He joined the Hebrew University in 1982. From 1987 to 1993 he held a joint appointment as a professor at the Hebrew University and as a research staff member at the IBM Almaden Research Center. He is currently a professor at the Hebrew University of Jerusalem. His research interests are all aspects of distributed computing, fault tolerance, and networking—theory and practice..

Sugih Jamin is an Associate Professor in the Department of Electrical Engineering and Computer Science at the University of Michigan. He received his Ph.D. in Computer Science from the University of Southern California, Los Angeles in 1996 for his work on measurement-based admission control algorithms. He spent parts of 1992 and 1993 at the Xerox Palo Alto Research Center, was a Visiting Scholar at the University of Cambridge for part of 2002, and a Visiting Associate Professor at the University of Tokyo for part of 2003. He received the ACM SIGCOMM Best Student Paper Award in 1995, the National Science Foundation (NSF) CAREER Award in 1998, the Presidential Early Career Award for Scientists and Engineers (PECASE) in 1999, and the Alfred P. Sloan Research Fellowship in 2001.

Osnat (Ossi) Mokryn Received the B.Sc. in Computer Engineering and M.Sc. in Electrical Engineering from the Technion—Israel Institute of Technology, Haifa in 1993 and 1998, respectively. Submitted the Ph.D. in Computer Science and Electrical Engineering to the Hebrew University of Jerusalem, Israel in December 2003. She currently holds a Post-Doctorate position at the department of Electrical Engineering at Tel-Aviv University. Her recent research focuses on Internet structure and topology; Complex systems; Multicast; Caching and content delivery.



Yuval Shavitt (s'88–M'97–SM'00) received the B.Sc. in Computer Engineering (cum laude), M.Sc. in Electrical Engineering and D.Sc. from the Technion—Israel Institute of Technology, Haifa in 1986, 1992, and 1996, respectively.

From 1986 to 1991, he served in the Israel Defense Forces first as a system engineer and the last two years as a software engineering team leader. After graduation he spent a year as a Post-Doctoral Fellow at the Department of Computer Science at Johns Hopkins University, Baltimore, MD. Between 1997 and 2001 he was a Member of Technical Stuff at the Networking Research Laboratory at Bell Labs, Lucent Technologies, Holmdel, NJ. Starting October 2000, he is a faculty member in the department of Electrical Engineering at Tel-Aviv University. His recent research focuses on Internet measurement, mapping, and characterization; QoS routing; and cache placement. He served as TPC member for INFOCOM 2000–2003, IWQoS 2001 and 2002, ICNP 2001, MMNS 2001, and IWAN 2003 and 2003, and on the executive committee of INFOCOM 2000, 2002, and 2003. He is an editor of Computer Networks, and served as a guest editor of IEEE JSAC and JWWW.