2ND QUARTER 2008, VOLUME 10, NO. 2

IEEE COMMUNICATIONS SURVEYS
The Electronic Magazine of Original Peer-Reviewed Survey Articles

www.comsoc.org/pubs/surveys

NETWORK TOPOLOGIES: INFERENCE, MODELING, AND GENERATION

HAMED HADDADI AND MIGUEL RIO, UNIVERSITY COLLEGE LONDON
GIANLUCA IANNACCONE, INTEL RESEARCH
ANDREW MOORE, UNIVERSITY OF CAMBRIDGE COMPUTER LABORATORY
RICHARD MORTIER, MICROSOFT RESEARCH CAMBRIDGE

ABSTRACT

Accurate measurement, inference and modeling techniques are fundamental to Internet topology research. Spatial analysis of the Internet is needed to develop network planning, optimal routing algorithms, and failure detection measures. A first step toward achieving such goals is the availability of network topologies at different levels of granularity, facilitating realistic simulations of new Internet systems. The main objective of this survey is to familiarize the reader with research on network topology over the past decade. We study techniques for inference, modeling, and generation of the Internet topology at both the router and administrative levels. We also compare the mathematical models assigned to various topologies and the generation tools based on them. We conclude with a look at emerging areas of research and potential future research directions.

The Internet connects millions of computers, sensors, monitoring devices, and Internet Protocol (IP) telephony devices together, offering many applications and services such as the World Wide Web, email, and content distribution networks. Hosts on the Internet are connected via thousands of Internet service providers (ISPs). An ISP contains one or more autonomous systems (ASs) depending on its size. An AS is a set of routers within a single administration domain, such as a university or corporate network.

By convention, the Internet is built on two domain categories, *transit* and *stub*. A transit AS usually carries traffic between other domains. A stub AS, such as a university network, is one that has connections to end hosts and relies on at least one transit AS for connectivity to the rest of the Internet. Stub ASs usually do not enable IP packets to transit their networks if they are not sent or received by an end host within the network. Figure 1 displays a simplified version of this structure.

In Fig. 1 transit domains carry traffic between customer ASs, ISPs or stub domains. The ISPs may have exchange (peering) relationships among themselves for resilience and cost saving purposes. Some ASs of ISPs are attached to more than one transit AS. This is a backup measure increasingly being taken by corporate networks and business customers in order to ensure the existence of alternative routes to the Internet should their main provider fail. It is also a technique for traffic engineering, allowing traffic to be sent over links of different levels of performance. This strategy is called multihoming and is displayed in Fig. 1.

The growth of the Internet and the overlay networks that rely on it has led to emerging applications and properties that have not been considered in the current topology inference and generation tools. Dynamic reconfiguration of routers and firewalls, changes in routing policies of ISPs, overlay networks, peer-to-peer networks, increasing use of virtual private networks (VPNs), protocols such as multiprotocol label switching (MPLS), and tunneling techniques, multihoming, on-demand circuit setup, and bandwidth allocation for home entertainment and videoconferencing, and the increased existence of mobile devices and laptops have caused the topology of networks to be in constant change.

There is no central node in the Internet. As a result of this, there are no coordinate systems embedded within the architecture of the Internet. However, an important result of the

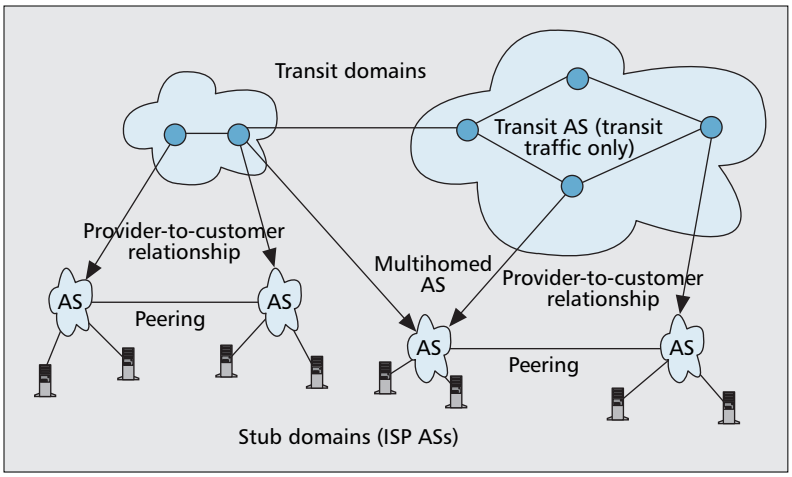

Figure 1. *Sample Internet transit-stub architecture.*

inference of the Internet topology is the ability to map nodes to geographic locations. Such a mapping will allow one to discover the collocation points of ISP points of presence (PoPs)[1] and provide for Internet maps. In reality, many ASs are formed of many customers that overall act as a major contributor to the traffic that is sunk or sourced by the large provider AS. An example is the U.K. academic network that is spread around the whole country and is only observed as a single AS number from the outside world.

In this section we introduce the basic concepts of the Internet's operation and the need for network topology inference, modeling, and generation. We describe commonly used metrics for topology characterization. We describe challenges of topology inference, modeling, generation, and validation. We describe the inference of router-level topologies of ISPs and the AS-level topology of the Internet, and the impact of geographical location of the nodes on inference techniques. We discuss the statistical and hierarchical models used to represent the topologies of the Internet at the AS and router levels. We describe the tools available for topology generation. Finally, we introduce possible future research directions and conclude the survey.

Topology Characterization

Many metrics are used to characterize Internet topologies. These metrics have mostly been adopted from graph theory. We discuss some of the most widely used metrics in this section.

Average degree: For an undirected graph with n nodes and m links, the average node degree is defined as $k = 2m/n$. Average degree is the most basic connectivity characteristic of a topology. Networks with higher k have higher connectivity on average and are thus considered to be more robust [1].

Degree distribution: The node degree distribution is the probability that a randomly selected node is k-degree: $P(k) = n(k) = n$, where $n(k)$ is the number of nodes of degree k. Degree distribution provides a more detailed view of the structure of a network and is one of the most frequently used topology characterization metrics [1].

Joint degree distribution (degree correlation): The joint degree distribution (JDD) is the probability that a randomly selected edge connects k- and k'-degree nodes: $P(k; k') \sim m(k; k0)/m$, where $m(k; k')$ is the total number of edges connecting nodes of degrees k and k'. JDD provides information about the node's proximity and neighborhood since it reveals both the degree distribution $P(k)$ and average degree $\bar{k}$, in addition to assortativity of a graph, which is a measure of the correlation between nodes. Assortative networks are ones in which nodes of similar degree are correlated. The Internet is considered to be a disassortative network [1].

Clustering: When $\overline{m}_{nn}(k)$ is the average number of links between the neighbors of k-degree nodes, local clustering is the ratio of this number to the maximum possible number of such links: $C(k) = 2\overline{m}_{nn}(k)/k(k-1)$. If two neighbors of a node are connected, these three nodes together form a triangle (3-cycle). By definition, local clustering is the average number of 3-cycles involving k-degree nodes. The two summary statistics associated with local clustering are *mean local clustering* $\overline{C} = \Sigma C(k)P(k)$ and the *clustering coefficient* C, which is the proportion of 3-cycles among all connected node triplets in the entire graph. Clustering provides a measure of how close a node's neighbors are to forming a clique. The larger the local clustering of a node, the more interconnected are its neighbors, increasing the path diversity locally around the node [1].

Rich-club connectivity: The rich club phenomenon is a property of the hierarchical nature of the Internet structure, and is the effect of well connected (*rich*) nodes being tightly coupled to form a group. Each node may have one or many links. For less connected nodes to connect together, they have to go through rich nodes. Rich club connectivity $\phi(\rho/n)$ is the ratio of the number of links in the subgraph induced by the ρ largest-degree nodes to the maximum possible links ($\rho(\rho - 1)/2$ where $\rho = 1 \ldots n$ are the first ρ nodes ordered by their nonincreasing degrees in a graph of size n [2].

Coreness: The k-core of a graph is a maximal subgraph in which each node has at least degree k. In other words, the k-core of a graph is defined as the unique subgraph obtained by recursively removing all nodes of degree less than k. A node has coreness k if it belongs to the k-core but not to the $(k + 1)$-core. Hence, the k-core layer is the collection of all nodes having coreness k. The core of a graph is the k-core such that the $(k + 1)$-core is empty [3]. The minimum node coreness in a given graph is $\kappa_{min} = k_{min} - 1$, where k_{min} is the lowest node degree present [1].

Shortest path length (distance): The shortest path length distribution is the distribution of the probability of two nodes being at minimum distance x hops from each other. The shortest path distribution is a strong indicator of network performance as it shows the reachability of nodes within each other, and for viruses and worms spreading over portions of the network [2].

Betweenness: Betweenness is a centrality measure of a node within a graph. Betweenness is a measure of the number of shortest paths passing through a node or link. It is formed by normalizing the shortest paths passing through a node by the number of pairs of links which do not include that node. Betweenness is important for traffic engineering applications as it allows the analysis of the load on various points in the network [2].

Spectrum: The spectrum of a graph is the set of eigenvalues of its adjacency matrix. Spectrum is an important measure of the overall characteristics of a network and its robustness [1].

[1] *A PoP is a physical location that houses servers, routers, and switches belonging to an ISP. A large ISP may have hundreds of PoPs.*

Topology Research Challenges

The Internet topology is usually investigated at two levels. The Internet AS-level topology is of interest for those looking at issues such as interdomain routing, quality of service (QoS) provisioning, and customer-provider and peering relationships between tier 1 ASs and lower-level ISPs. Within an AS, the router-level topology map of ISPs is needed for research on issues such as optimum network planning, and the ability to minimize the impact of router and link failures.

There are many challenges in inferring and generating realistic Internet topologies. Information on network topology, routing policies, peering relationships, resilience, and capacity planning are not usually available publicly as they are considered sensitive information that gives an ISP its competitive advantage. Researchers try to infer the required data by using passive and active measurement methods to produce snapshots of the global Internet or individual ISP topologies. The fundamental problem of these techniques is the lack of a complete image of the Internet topology. Such a map does not even exist, as the Internet is constantly evolving, and it is difficult to define the topology of the Internet at a given time. This in turn leads to poor perceptions and models, as the underlying data set is qualitatively poor. In this section we discuss these challenges in turn.

Inference of Topologies

At the AS level, it is not possible to obtain a consistent map of the actual AS-level topology of the Internet due to the constantly changing nature of the Internet topology. Operators are constantly reviewing their peering agreements, adding new links, withdrawing other links, and performing maintenance. AS operators will not reveal their peering relationships and traffic exchange policies with other ASs. Connectivity between ASs is made possible by use of interdomain routing protocols, primarily the Border Gateway Protocol (BGP) [4]. However, BGP data collected from various points on the Internet is not enough to provide a user with a *complete* map of the Internet at the AS level.

Challenges also exist when trying to get the router-level topology of a single AS. The router-level topologies of ISPs are also dynamic and constantly evolving due to failures, maintenance, and upgrades. Network operators are not willing to publicly release the maps of their network topology as they are considered commercially sensitive, potentially revealing their resilience planning, exposing potential vulnerability to attackers who will target bottlenecks in the network.

The most widely used tool for inference of router-level topologies is the *traceroute* tool [5], which is known to miss alternative links between routers. Also, routers have multiple interfaces with separate IP addresses. During the inference process, each of these interfaces may be reported as a different router. This problem is referred to as *aliasing*.

Aliasing leads to incorrect path prediction, missing routes, and measuring topologies which appear bigger than the actual topology, as a device may be listed more than once due to various physical and virtual interfaces with separate IP addresses. We discuss these issues in detail later.

Modeling the Internet

Simulation and modeling of the Internet is known to be a difficult task. Floyd and Paxson [6] discuss some of the challenges of simulation design. The constant growth of the Internet has made it difficult to develop a representative model for analysis of its topology.

Researchers have made significant efforts to model the characteristics of the Internet. The major problem currently in this field is the absence of detailed information about generated topologies. Many of these models are based on data sets that are known to be incomplete and prone to errors due to the nature of the collection process involved, discussed in detail later.

While it is extremely useful to have a model that characterizes the AS-level topology of the Internet as a whole, if the link bandwidths and routing policies between networks are unknown, it is still a difficult job to estimate the growth potential and characteristics of the traffic on addition of a new network. This is a vital stage for network traffic engineering purposes. Later we describe many of the widely used models.

Topology Generation

There are several topology generation tools available. Each generator is based on a specific model, which has been developed as the result of focusing on certain characteristics of interest to the designers of the generator. Some of the requirements for a network topology generator, also listed by Medina *et al.* [7], include:

Representativeness: The generated topologies must be accurate, based on the input arguments such as hierarchical structure and degree distribution characteristics.

Flexibility: In the absence of a universally accepted model, the generator should include different methods and models.

Extensibility: The tool should allow the user to extend the generator's capabilities by adding their own new generation models.

Efficiency: The tool should be efficient for generating large topologies while keeping the required statistical characteristics intact. This can make it possible to test real world scenarios

Designing a topology generator which satisfies the above requirements is a challenging task. For example, a generator must be able to fit thousands of nodes to represent a large AS, while keeping in mind all the graph characteristics of the topology such as power laws[2] or node degree distributions. The software designer has to make sure the generation process is completed in a reasonable amount of time, and within reasonable CPU and memory constraints.

The numerous challenges in achieving these objectives have led researchers to design topology generators that cover only some of them. We discuss topology generators in detail later.

Validation of Models

Validation of generated topologies can be done by comparison to real topologies. Another common method is to compare the statistical characteristics of a generated topology with input parameters and requirements such as certain node degree distributions or connectivity matrices. As there is no real *snapshot* of Internet traffic or its topology, it is difficult to devise a method to benchmark the success of a topology generator or the inference of a topology.

When inferring the router-level topology of a medium sized ISP, it may be possible to request the operator to verify the results, as done by Spring *et al.* [8]. However, as mentioned before, operators are unlikely to reveal such informa-

[2] *A power law of quantities* x *and* y *is one where the relationship can be written as* $y = ax^k$ *where* a *(the constant of proportionality) and* k *(the exponent of the power law) are constants.*

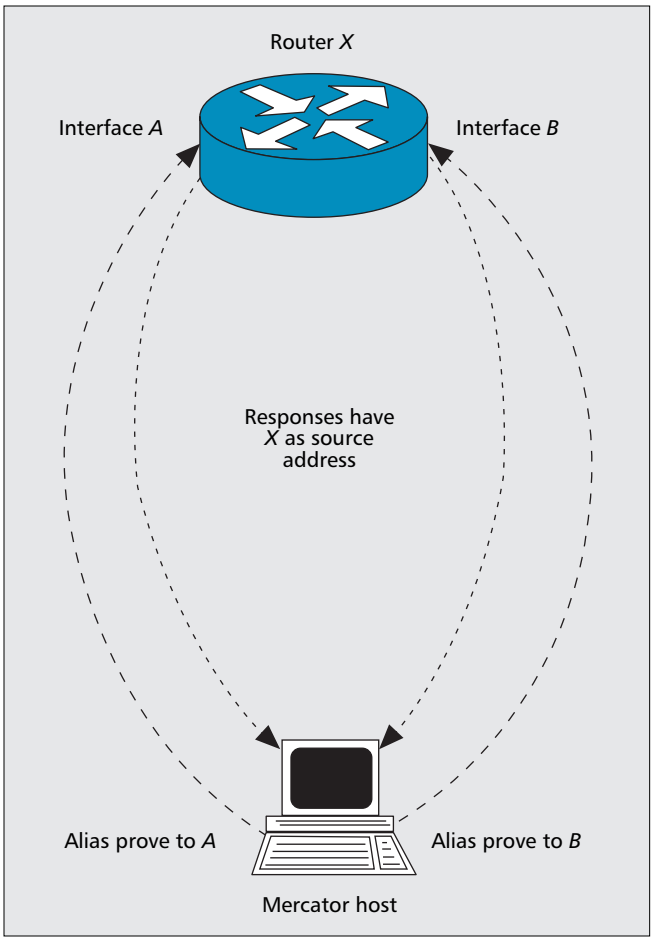

Figure 2. *Alias probing to all routers will help resolve router aliases.*

tion, although they may indicate the success level of an inference method as a percentage of routers or links discovered. BGP and AS ownership data can also be validated by relevant Internet domain registries, although the information held by such authorities is not continuously updated and is thus often inaccurate.

Topology Inference

In this section we discuss recent efforts on inference of the AS-level topology of the Internet and router-level topology of ISPs. It is essential to note the intersection of inference with measurement. Inference-based statistics are subject to the underlying measurement process, and the assumptions that have been made on the level of accuracy and detail of the measurement process. Thus, inaccurate inference methods lead to unrealistic models.

Topology inference work usually falls in two categories: router-level and AS-level. In related literature, Donnet and Friedman [9] also mention the IP interface and PoP-level maps. IP interface addresses are usually aliases for the same router, and we mention the problems associated with resolving such aliases in this section. Inferring PoP-level maps is a difficult task due to lack of publicly available data sets or tools. Hence, they are sometimes made available by network operators or inferred indirectly from IGP routing data.

ISP Router-Level Maps

Here we discuss the recent efforts and tools for discovering the Internet's router-level topology, also known as its *IP layer* or *layer 3* topology. These methods are usually based on the traceroute tool. Traceroute is the basic tool for discovering the paths packets take in the Internet. Nearly all attempts to extract routing and topology information of the Internet at the router layer use traceroute.

Traceroute works by sending multiple Internet Control Message Protocol (ICMP) [10] packets with increasing time to live (TTL) fields in the IP header. When a packet with a TTL of one reaches a host, the host discards the packet and sends an *ICMP time exceeded* packet to the sender. The traceroute tool uses the IP source address of these returning packets to produce a list of hosts the packets have traversed on their route to the destination. By incrementing the TTL value after each response, the overall path taken by the packets can be inferred.

Mercator — One of the first tools to rely on traceroute for mapping sections of an ISP is *Mercator*, introduced by Govindan *et al.* [11]. The aim of Mercator is to build a nearly complete map of the transit portion of the Internet from any location where Mercator is run, using hop-limited probing. Hop-limited probing differs from traceroute as it stops probing once a probe fails to elicit a response. This is appropriate for Mercator as it focuses on discovering router adjacencies. The technique used in Mercator is referred to as *informed random address probing*, in which a response from an IP address adds the /16 address prefix[3] to the Mercator list, and Mercator assumes that the neighboring prefixes are also addressable. Another assumption is sequential assignment of address space by the registries, such that, for example, `128.8/16` and `128.10/16` are the neighboring prefixes of `128.9/16`.

Using probing from source-route probe-capable routers, it is possible to find cross-links and avoid discovering only a tree-like structure. Mercator sends a UDP message to a high port number on the router and receives an ICMP reply back. If two source addresses of the reply message are the same, they are from the same router. This operation relies on the requirements for Internet hosts described in RFC 1122 [12]. This is a technique for *alias resolution* that identifies the interfaces belonging to the same router. Figure 2 displays a simple approach to resolving aliases on routers with multiple interfaces with different IP addresses.

A Mercator host sends packets to the router interfaces. If both interfaces reply with the same source address, they belong to the same router.

The challenges faced by Mercator are due to the fact that it does not attempt to cover the whole spectrum of a network due to its randomized process and the fact that many routers do not forward traceroutes for source-routing in the way Mercator requires. Only 8 percent of routers have actually responded to source-routing probes, which makes this approach to alias resolution inaccurate. However, it is stated that such a percentage is enough to discover more than 90 percent of the links. Mercator will take about three weeks to discover nearly 150,000 interfaces and 200,000 links on the Internet, although such an experiment has not actually been carried out [11]. Another drawback is the ubiquitous use of = 16 prefixes, which is no longer a valid assumption due to the

[3] *A prefix is a block of destination IP addresses to which an AS applies local policies to select the best route and decide whether to export this route to neighboring ASs.*

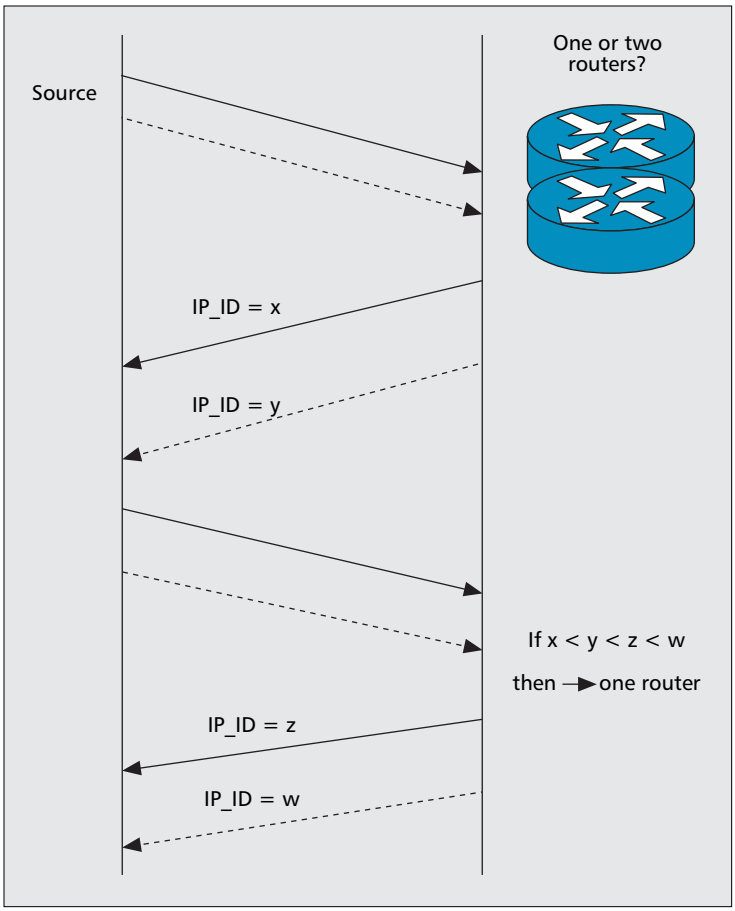

Figure 3. *Alias resolution using IP_ID field. Solid arrows and dotted arrows represent messages to and from two different IP addresses.*

use of Classless Inter Domain Routing (CIDR) [13].

Skitter — One of the most widely used data sets is that collected by the *Skitter* project.[4] Huffaker *et al.* [14] state the project focus as "active measurement of the topology and round trip time (RTT) information across a wide cross-section of the Internet." Even though the active probing process suffers the limitations of reachability in the huge address space of the Internet, a compromise is reached by probing frequently to a large number of destinations over many years.

Probing uses the traceroute tool. IP addresses are then mapped into their corresponding origin AS. The disadvantage of such a tool is the large amount of data it produces from a number of sources currently placed in 25 locations worldwide. This leads to the inherent problems of traceroute such as aliasing on a wider scale as multiple sources are involved. Skitter does not attempt to resolve aliases.

Visualizing tools have been developed by the Skitter project, which display nodes based on their connectivity degree, with highly connected ASs in the center of the produced diagrams.

As of March 2006, Skitter reports a total of 192,244 nodes, 636,643 directed links, and 609,066 undirected links. The average and maximum node degrees (undirected) are 6.34 and 1071, respectively.[5]

The Skitter project led to interesting discoveries about Internet topology when compared with those topologies based on BGP data or information from Internet Routing Registry (IRR) and WHOIS[6] servers. WHOIS servers provide a mechanism for finding contact and registration information for Internet domains, and they can contain information on peering relationships and routing policies. Mahadevan *et al.* [1] mention that the Skitter graph closely reflects the topology of Internet traffic flows (i.e., the data plane), while the BGP graph from RouteViews[7] reveals the topology seen by the routing system (i.e., the control plane). Distinct differences between the three topologies are shown by calculation of metrics such as average degree of nodes, node degree distribution, joint degree distribution, clustering coefficients, Rich-Club connectivity, node distance, and betweenness [1].

Rocketfuel — In an attempt similar to Mercator, Spring *et al.* in the Rocketfuel project [8] try to infer the maps of 10 ISPs, consisting of backbones, access routers, and directly connected neighboring domain routers. Validation is attempted by using some of the ISP's own topology data. Direct probing techniques are used to filter the traceroutes on the ISP of interest, using BGP table information from RouteViews. A BGP table maps destination IP address prefixes to a set of AS paths that can be used to reach the destination. Each AS-path represents the list of ASs that will be traversed to reach destinations within the prefix. Path reduction is also the method suggested in Rocketfuel for elimination of multiple paths by ignoring redundant traceroutes (those that take different paths within the same ISP). The traceroute path through an ISP usually depends only on the next-hop AS and hence the appropriate edge router, not on the specific destination AS. This means that only one trace from the ingress router to the next-hop AS is likely to be valuable; the rest can be eliminated by path reduction. Public traceroute servers are used as vantage points for the traceroutes.

Rocketfuel uses the direct probing method, as suggested by Govindan and Tangmunarunkit [11]. In order to ensure correct resolution of aliases, Rocketfuel also uses the `IP_ID` field of the router's responses to probe packets, which is incremented by the router. This method is illustrated in Fig. 3.

In Fig. 3 the source sends two probe packets to the two interfaces that are thought to be aliases of the same router. If consecutive responses from the interfaces increment the `IP_ID` by a small value, it indicates that the same IP stack is running on the same router with multiple interfaces; hence, the interfaces are believed to belong to the same router. Otherwise, the interfaces belong to two distinct routers.

Probing in Rocketfuel is performed by using 294 traceroute servers to query each /24 prefix, or 256 IP address blocks of the ISP of interest, to discover the egress routers. Rocketfuel finds very different backbone networks on the five different ISPs that have been willing to verify the inferred topologies. The aim of this practice has been to find the PoP sizes of ISPs, PoP out-degree, and router out-degree.

In the validation stage they discover that some parts of IP addresses were missed due to the randomized ping[8] process.

[4] *http://www.caida.org/tools/measurement/skitter*

[5] *http://www.caida.org/tools/measurement/skitter/router topology*

[6] *http://www.whois.net/*

[7] *http://www.RouteViews.org*

[8] *Ping is an IP network utility that allows a particular host to be tested for reachability.*

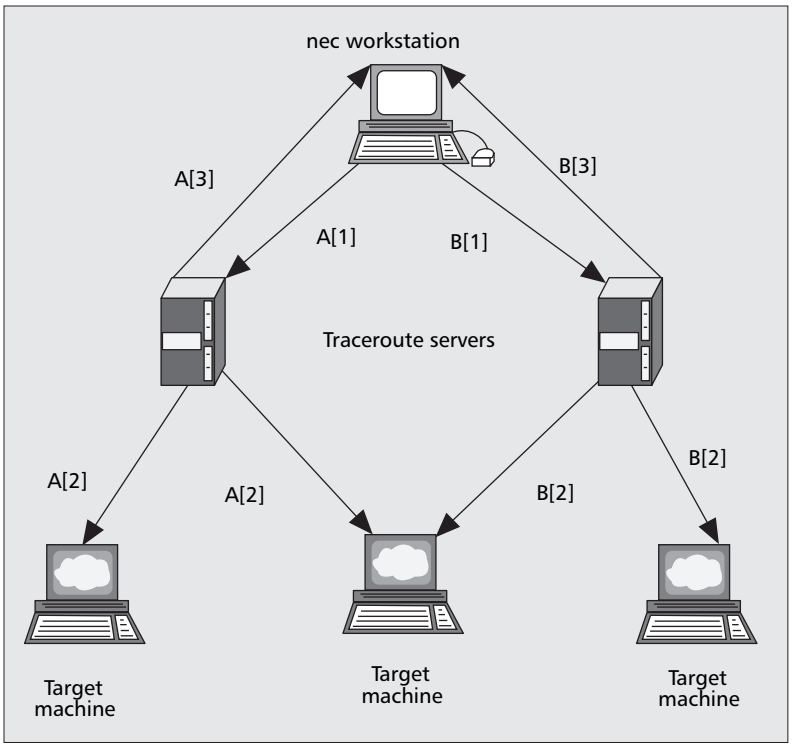

Figure 4. *nec mapping steps, figure courtesy of [16].*

Tool	Released	Alias resolution	Updated	Probes
Mercator	1999	Yes	No	Single
Skitter	1999	No	Yes	Multiple
Rocketfuel	2002	Yes	No	Single
nec	2003	No	No	Multiple
DIMES	2004	No	Yes	Multiple

Table 1. *Comparison of traceroute-based methods.*

Some routers do not respond to ping packets. Some do not follow the target ISP's naming conventions due to use of different IP address ranges. These lead to missing some nodes. Also, traceroute is blind to alternative unused links so some neighboring routers and links are missed.

Alias resolution is an important part of Rocket fuel's attempt to infer the correct topology of an ISP. Spring *et al.* [15] introduce two alias resolution approaches based on inference to handle addresses that cannot be resolved by ordinary DNS lookup methods. The first method decodes the DNS names assigned by the ISP to recognize the name fragments that identify a router. The DNS technique can only be as accurate as the ISP's database, which must be updated as addresses are re-assigned and ISPs are merged. It also relies on understanding the naming convention of the ISP, if such convention exists. The second method infers aliases from the graph of linked IP addresses and requires no additional measurement traffic. The proposed graph-based techniques are based on the following observations:

- Two addresses that directly precede a common successor are aliases, assuming point-to-point links are used.
- Addresses found in the same traceroute are not aliases, assuming there are no routing loops.

nec — Another tool for inference and mapping of a network topology is the *network cartographer* (nec) mapping software introduced by Magoni and Hoerdt [16]. The *nec* tool is a traceroute-based mapper from multiple traceroute servers, finding routers and links and producing a router-level connectivity graph. The major difference between nec and Rocketfuel [8] is that nec has wider scope while Rocketfuel focuses on a single ISP. Unlike Rocketfuel, where a few hosts target thousands of IP addresses, *nec* uses many traceroute Web servers to a limited set of chosen IP addresses. Figure 4 displays the steps involved in an nec mapping query, sent to two traceroute servers, A and B.

In the first stage the queries are sent from the workstations to the traceroute servers. In the second stage traceroute servers query the selected IP addresses. In the final stage the results of the traceroutes are sent back to the nec mapping workstations.

The described selective mapping technique allows nec to build an overlay map of the IP addresses corresponding to the relevant ASs. The destination addresses are chosen in such a way that they all belong to distinct ASs, in order to obtain an optimal topological distribution of the targets and build a valid AS-level map. For the target addresses nec takes random addresses created from all /16 and shorter prefixes of a RouteViews BGP dump, with lower prefixes truncated to /16. This yielded nearly 14,000 destination addresses. The nec maps and software are freely available.[9]

DIMES — The DIMES project [17] attempts to build a router-level map of the Internet. In this project, the DIMES agent, which can be installed on any computer connected to the Internet, performs Internet measurements such as traceroute and ping at a low rate, sending the results to a central collection station at regular intervals. The advantage of the DIMES approach over previous traceroute based mapping tools is that the probing process is done across many locations in the world, giving a more complete map of the Internet router-level topology. However, due to the large number of vantage points and collection of overlapping measurements, removing the redundancies in the data is a complicated process and DIMES also does not attempt to resolve router aliases.

Comparison of Traceroute-Based Methods

In this section we listed a number of methods for inferring router-level connectivity information. These methods have evolved over time from single source traceroute probes to universally distributed probing agents. Table 1 displays a summary of the characteristics of these methods.

It can be observed that the trend of inference tools has moved from single-source static maps to those spread across many sites and constantly updating their database. It is interesting to note that there are no maintained maps with alias resolution. Indeed, researchers realized the importance of

[9] *http://www-r2.u-strasbg.fr/magoni/nec/*

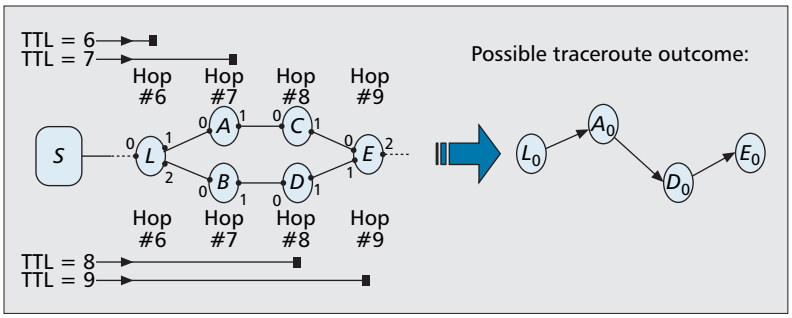

Figure 5. *Traceroute false reporting, figure provided by [27].*

looking at the growth of networks, and discovery of new links and routers that are added daily to the large networks.

Accuracy of Traceroute Maps

Most of the work in discovering the router-level topology of ISPs relies on the traceroute tool. Achlioptas *et al.* [18] discuss some of the problems associated with traceroute. They explore the mathematics of the sampling bias of traceroute, confirming that even when a given node degree distribution is Poisson, after traceroute sampling, the inferred node degree distribution exhibits power law properties. The presented theorem predicts the observed degree distribution after sampling, given a true degree distribution of a graph. It is difficult to remove this bias, as shown by Clauset and Moore [19], as the number of sources required to compensate for the bias in traceroute sampling grows linearly with the mean degree of the network.

Teixeira *et al.* [20] look at path diversity (number of available paths) in the Sprint network[10] and ISPs explored by Rocketfuel. Rocketfuel path diversity discovery is found to be at extremes, either overestimating or finding very little diversity, again due to the use of traceroute. The differences between the Sprint data and Rocketfuel inferred maps are due to nondiscovery of backup links, lack of vantage points, incomplete traceroute information, path changes in a traceroute, and incorrect DNS names.

Lakhina *et al.* [21] analyze the effects of such traceroute sampling techniques on random graphs and conclude that when graphs are sampled using traceroute-like methods, the resulting degree distribution can differ significantly from the underlying graph. For example, given a sparse Erdős-Rényi random graph, the subgraph formed by a collection of shortest paths from a small set of random sources to a larger set of random destinations can exhibit a degree distribution remarkably like a power law. The implementation of sampling in the article is performed on measurements from Skitter, Mercator, the data set used by Faloutsos *et al.* [22], and the Pansiot-Grad [23]. In studies of the four traces, the sampled subgraph shows differences in degree distribution and other characteristics from the original graph.

Deploying a large number of monitors usually results in having to process large data sets from each monitor. Donnet *et al.* [24] try to find out the amount of redundancy across data sets, focusing on the CAIDA Skitter data sets. They discover that around 86 percent of a given monitor's probes are redundant in a sense that they visit router interfaces which have already been visited by the monitor, especially those closer to the monitoring station. It is also observed that many of the probes are redundant in a monitor's data set as they have already been visited by the other monitors, particularly those at an intermediate distance (between 5 and 13 hops). This leads to intermonitor redundancy.

In an attempt to remove such redundancies, Donnet *et al.* [24] proposed the Double-tree algorithm, which relies on the tree-like structure of the routes as mostly observed at the edges of the networks. The Double-tree algorithm suggests that the probing process is started at a distance in the middle of the path between the monitor and the destinations. In order to avoid dealing with alias resolution, they consider metrics at the interface level, specifically host and router interfaces, suggesting that each interface can be thought of as a node with its own connection. The Double-tree algorithm assumes a tree-like structure of routes consisting of two types of trees. The first is the monitor-rooted tree, when all the traceroutes emanate from a single point toward a set of destinations. The second type is the destination- rooted tree, when all of the traceroutes converge from a set of monitors toward a given destination.

These trees are handled by Double-tree using a particular data structure named the *stop set*. As there are two trees, Double-tree considers two stop sets: a local stop set, which is a set of interfaces encountered by a given monitor, and a global stop set, which is a set of interface pairs encountered by any monitor in the system. This global stop set is shared by monitors. Furthermore, Double-tree uses forward probing (increasing TTL from the starting point in the network) until reaching the destination or finding a pair belonging to the global stop set and backward probing (decreasing TTL from the starting point in the network) until reaching the hop located at TTL = 1 or an interface belonging to the local stop set.

The conclusion drawn by Donnet *et al.* is that reduction in probing redundancy by placing the probing point at the center between the monitor and the destination can reduce the measurement load by 76 percent while still discovering 90 percent of the links and interfaces, compared to the simulation results of a Skitter-type network mapper.

In the next stage of this work Donnet *et al.* [25] build on the Double-tree topology inference algorithm by using CIDR address prefixes. Each monitor probes each destination and records it in the global stop set (interface, destination prefix) pairs instead of (interface, destination) pairs. This is based on the evidence that more than 80 percent of Skitter probes are redundant as they discover no new interfaces, based on statistics taken from a subset of the Skitter data. This figure is much higher at lower hop counts (closer to the monitor). It is also common between monitors to have large redundancy near destinations being probed as most monitors discover the same IP addresses. This method increases the accuracy comparable to classic Double-tree only when the prefix is as large as /24, due to destinations within subnetworks being missed entirely or probe packets being stopped at egress routers. Donnet *et al.* suggest the use of Bloom filters[11] for the global stop set membership and implement a prefix-based method to reduce the stop set that is used.

As a result of the traceroute sampling bias, there has been ongoing effort to modify traceroute behavior. Augustin *et al.* [27] propose *Paris traceroute*, which is a modified version of traceroute with the ability to discover redundant paths. One of the issues when using traceroute arises due to the equal cost multipath (ECMP) load balancing deployed by multihomed stubs and network operators. This leads to traceroute

[10] *http://www.sprintlink.net*

[11] *Bloom filters were proposed by Burton Bloom [26] as an efficient probabilistic data structure for approximate membership checking, with a controllable probability of giving false positives.*

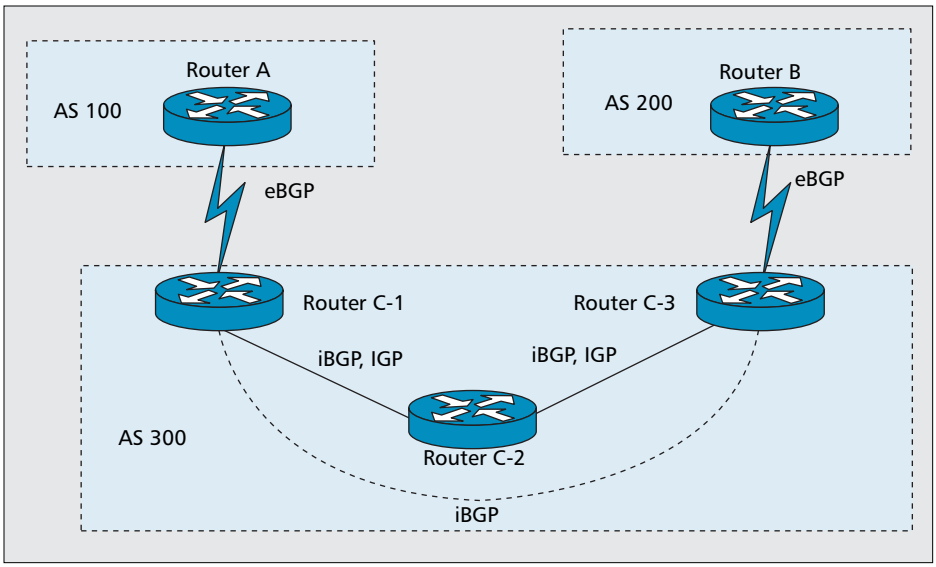

Figure 6. *Interaction of BGP, iBGP, and IGP in interdomain and intradomain routing.*

taking different paths on each occasion, as shown in Fig. 5. Paris traceroute looks into the effects of load balancing and its frequency on traceroute anomalies. Load balancing can be done per packet, per flow, or per destination IP address.

Augustin *et al.* show that by manipulating the ICMP sequence number and checksum in the ICMP packet header, it is possible to ensure that all the packets on traceroute take the same path. This leads to discovery of more possible routes. With this method it is also possible to report on the loops and cycles in ordinary traceroute reports. Paris traceroute is suggested as an alternative to the ordinary traceroute, rather than as a topology mapping tool; hence, it does not attempt to resolve any router aliases.

Dall'Asta *et al.* [28] find that the node and link detection probability depends on statistical properties of elements such as betweenness centrality. Hence, the shortest path routed sampling, or sampling the network from a limited set of sources as performed by traceroute, provides a better characterization of underlying graphs with broad distributions of connectivity, such as the Internet. The studied model analyzes the efficiency of sampling in graphs with heavytailed connectivity distributions and looks at metrics such as the node degree distribution. The conclusion drawn is that unlike homogeneous graphs, in those with heavytailed degree distribution such as the Internet, major topological features are easily captured though details such as the exponent of the power laws. However, this behavior appears to suffer from biases that result from the sampling process.

The studies in this section may imply that traceroute is not a suitable tool for detailed analysis of the Internet router-level maps. However, it is still widely used for topology measurement and is believed to be a reliable source, and in reality the only available tool, by many researchers.

AS-Level Internet Maps

The other important level of Internet topology is AS-level topology. The freedom of AS administrators to change their traffic exchange relationships with other providers has led to a constantly evolving AS-level map of the Internet. Obtaining such a map can enable better design of routing algorithms and traffic engineering between various ASs.

When BGP is used between ASs, the protocol is referred to as External BGP (eBGP). BGP information at routers is kept consistent by receiving BGP update messages from other ASs. BGP updates contain multiple route announcements and withdrawals. The announcements indicate that new network sections are available to the routers, or a policy change is enforced to prefer an alternative path over an existing one. Withdrawals occur when an existing route is replaced by a new route to a destination prefix by means of a withdrawal message. These messages report the withdrawal of links and addition of new links, and contain the AS-path traveled by the advertisement. Each router along the path prepends its own AS number to the AS-path in the BGP message. The AS-path is needed to avoid loops in the BGP route selection process.

The AS-paths, in conjunction with the AS prefix, are also used to decide on the best next hop to use for sending a packet to a destination. An edge router may not have a complete view of the BGP status of the Internet and may have a default path to a tier 1 provider. Tier 1 providers have default-free BGP information so that they can forward all the packets to the correct destination.

IP forwarding requires that all routers within an AS are aware of all the prefixes, which are learned by the edge routers from eBGP peering sessions. Interior BGP (iBGP) is the protocol used to advertise prefixes between the routers within an AS. The naive way to redistribute these prefixes is via a full mesh of iBGP sessions. However, Labovitz *et al.* [29] highlight the scaling problems with the advertisement of the routes on a regular basis in a large AS.

To avoid this scaling problem, there are two mechanisms used by large ASs. First is forming *AS confederations*, where a set of routers represent a single AS to the BGP peers external to the confederation, thereby relaxing the full-mesh requirements [30]. Second, *route reflectors* allow a BGP speaker to reflect learned routes to internal neighbors [31]. These methods reduce the distribution cost of the routing information within a network. Every router in such a network can then have either a full routing table or a default route.

While iBGP information in the routers allows them to identify the edge router for the packets to exit an AS, it does not allow them to calculate the route within the AS for reaching the edge routers. Within an AS, internal routing information is propagated using the Interior Gateway Protocol (IGP). Commonly used IGPs are the Intermediate-System-to-Intermediate-System (IS-IS) protocol [32], Routing Information Protocol (RIP), [33] and Open Shortest Path First (OSPF) protocol [34].

Figure 6 displays the interaction of the BGP, iBGP, and IGP within three distinct domains, *AS*100, *AS*200, and *AS*300. *AS*100 and *AS*200 each contain a single router, connected to a different edge router in *AS*300. *AS*300 contains three routers, only two of which are edge routers. Each of its edge routers will learn prefixes from a different AS using an eBGP session. To redistribute the prefixes so learned, each edge router will maintain an iBGP session with every other router within *AS*300. In order to forward packets within the AS, *AS*300 also runs an IGP, which ensures that all its routers can talk to each other.

Some attempts at AS-level topology discovery were based on using traceroute data. Inference of AS-level maps from

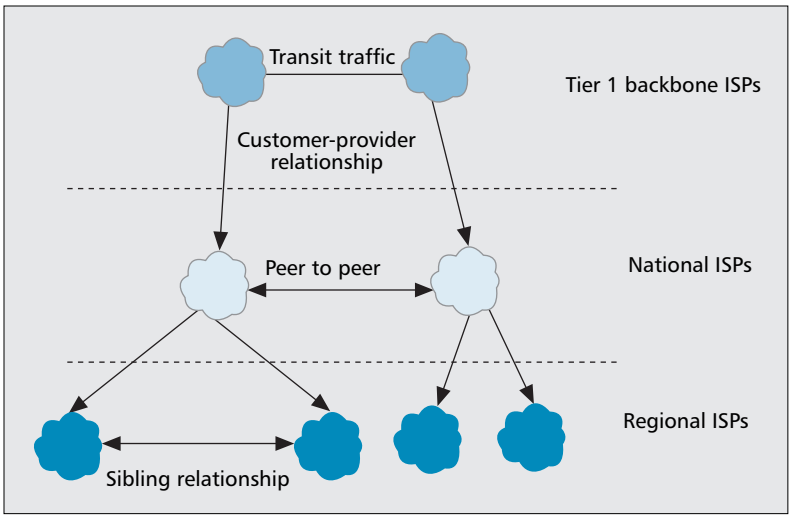

Figure 7. *Commercial relationships between ISPs.*

traceroute data includes problems not immediately noticed. Mapping of an IP address to the correct AS number incorporates challenges that are discussed by Mao *et al.* [35]. They propose techniques for improving mapping of IP addresses to corresponding ASs. These techniques rely on a measurement methodology for collecting both BGP and traceroute paths at multiple vantage points and using an initial IP-to-AS mapping derived from a large collection of BGP routing tables.

The difficulties arise due to the fact that the BGP table data and the actual path taken by packets can be inconsistent due to new route aggregation/filtering and routing anomalies [36]. The WHOIS data is also not always up to date due to company mergers, breakups, and IP address reallocations. An improvement can be made by collecting a large amount of information from BGP routing tables, BGP update messages and reverse DNS lookups in order to help traceroute build a more accurate AS-level map of the Internet.

The collection of traceroutes for measurement is done by sampling and picking two IP addresses for each prefix from a set of prefixes that covers the routable address space for a wide range of forwarding and signaling paths. The modification to traceroute is done by adding a second attempt UDP packet, with a waiting time of 5 s for an ICMP reply, as opposed to 2 s in the original first packet, in case of delays in the network. This allows for improvement in receiving ICMP replies. The corresponding AS path for the same prefix is also extracted from the BGP tables for each longest matching prefix. Having performed these steps, three sources of incompleteness in the paths are identified: unresolved hops within an AS, unmapped hops between ASs, and multiple origin AS mappings at the end of the paths. Most mismatches are due to the presence of Internet exchange points (IXPs), sibling ASs, and networks that do not announce routes for their infrastructure. Using the proposed alterations, it is possible to infer these mismatches and validate them using public WHOIS servers. However, the increased number of multihoming ISPs make the inference or definition of an origin AS in the path an increasingly difficult task [35].

Gao's seminal paper [37] is one of the first attempts to present an AS graph inferred from the Oregon RouteViews BGP data. The provision of such a map has enabled classification of AS relationships into customer-provider, peering, and sibling relationships. Figure 7 displays examples of the types of relationship between different ISPs.

A customer pays its provider for Internet connectivity and does not transit any traffic between its providers. A pair of peers agree to exchange traffic between their customers by sharing the cost of the peering links and eliminating traffic charges between each other. A pair of small ISPs may provide additional connectivity or backup connectivity to the Internet to each other in the form of a sibling relationship.

Despite the presence of such contractual agreements, there is little publicly available information about inter-AS relationships. The Routing Policy Specification Language [38] can be used to register information about peering relationships, but this information is not always accurately published due to its sensitive business nature. However, it is possible to infer such information from the BGP routing tables. Gao proposed heuristic algorithms for such discovery, and then validated some of the results by using a Tier 1 ISP's internal information. The discovery of the relationships is based on the BGP routing update export rules that are different for the individual relationships. The proposed solution by Gao is based on forming annotated graphs of the network and making sure the AS paths are *Valley-Free*; that is, after traversing a provider-to-customer or peer-to-peer edge (link), the AS path cannot traverse a customer-to-provider or peer-to-peer edge. The Valley-Free criteria holds only when the following conditions are met:

- A provider-to-customer edge can be followed only by provider-to-customer or sibling-to-sibling edges.
- A peer-to-peer edge can be followed by only provider-to-customer or sibling-to-sibling edges.

The proposed *basic algorithm* goes through the AS path of each routing table entry, finds the highest degree AS, and marks it as the top provider AS. It is then possible to go through the other ASs and set them as having customer-to-provider or sibling-to-sibling edges. This is assuming that all the BGP configurations are correct. However, in the *refined* version of the algorithm this assumption is relaxed by counting the number of routing table entries that infer an AS pair having a transit relationship by assigning consecutive AS pairs before the top provider with a transit relationship and consecutive AS pairs after the top provider with a transit relationship as well. This in turn finds any mismatches between the entries of the routing table.

The proposed algorithms are applied to the BGP data available from RouteViews and then verified against AT&T internal information. It is inferred that AT&T has one provider while in reality AT&T has no provider. 77.4 percent of inferred peers and 20 percent of inferred siblings are confirmed by AT&T. These siblings relationships are verified against the public WHOIS lookup service, which provides the name and address of the company that owns an AS. It is then possible to confirm the relationship between two ASs or two merged companies. The WHOIS data has confirmed more than 54 percent of inferred siblings.

Subramanian *et al.* [39] focused on peering relationships between ASs from a commercial relationship point of view. They combined BGP data from multiple vantage points to construct a view of the Internet topology, using BGP routing tables from 10 *Telnet Looking Glass* servers[12] using the *show ip bgp* command. The proposed algorithm ranks each AS from each of the vantage points based on the number of *up-hill* and *down-hill* portions. The results suggest the design of a topology generator based on directed graphs, as opposed to degree-based methods, as the directed graphs make distinction

[12] *http://www.traceroute.org/#LookingGlass*

between edge ASs, connecting to several transit core ASs.

This work led to many other interesting findings about AS-level relationships. Batista *et al.* [40] took this approach further by proving that identifying AS relationships from BGP data, especially when measured from multiple sources, is an NP-complete problem. The suggested solution is a linear time algorithm for determining the AS relationships in the case in which the problem admits a solution without anomalies for large portions of the Internet (e.g., data obtained from single points of view). The solution is performed by starting from a set of AS paths so that the number of invalid paths is kept small. This method can be applied on the address prefix of the hosts within an AS.

When looking at the path taken between ASs, direct access to endpoints is not always possible. The approach of using multiple sources of data is an extremely useful method in such scenarios. It enables a more detailed analysis of the possible paths between two end nodes (ASs in this case). Mao *et al.* [41] explored the feasibility of inferring AS paths by using BGP tables from multiple vantage points, router-level paths from traceroute servers, and AS-level paths from Looking Glass sites.

One of the inherent issues of inference of AS-level topology of the Internet by use of mapping node IP addresses to registered AS numbers is that sibling relationships are missed. Dimitropoulos *et al.* [42] proposed an alternative solution to AS-level map inference that attempts to find sibling-to-sibling (s2s) relationships, as well as customer-to-provider (c2p) or provider-to-customer (p2c). The proposed inference model avoids the mistake of considering siblings as customers or peers, which in turn may result in wrong inference of a provider as a customer, or the other way around, while still rendering a path as valid. The inference of s2s links plays an important role when looking at corporate networks, where multiple ASs belong to the same organization.

In order to look at the s2s relationships, the IRR databases are consulted and dictionary of synonymous organizations is manually created. Although a disadvantage of this approach is the fact that the IRR are not always up-to-date.

In the proposed inference model, weights are added to edges. These added weights are proportional to degrees of nodes connected to those edges. In these heuristics the weight on the edge is large when there is a significant degree difference between the neighbor ASs. When an edge is directed from a small-degree AS to a large-degree AS, it earns a bonus. Based on the bonuses, the edges of the graph are directed. The objective is to maximize the inference of valid paths and the number of paths. In summary, the inference heuristics take as input a set of BGP paths P and a corresponding graph $G(V, E)$ and perform the following three steps:

- Use IRRs to infer s2s relationships based on organizational ownership details and create set S of s2s links.
- Remove the subset S from consideration and apply the heuristics, assigning c2p/p2c relationships to the links remaining in E.
- Use P and G to infer p2p relationships and to create set F of p2p links.

The final result is set S of s2s links, set F of p2p links, and set E of c2p links.

These steps have been examined using the collected BGP tables from Route Views at 8-h intervals, over a period of four months in 2005. Invalid paths caused by BGP misconfigurations occur quite often and affect 200 to 1,200 prefixes each day and hence they are manually removed. To infer s2s relationships in graphs, IRR from RIPE,[13] ARIN,[14] and APNIC Whois databases are used. The authors then conducted a survey by contacting the AS engineers, and nearly a third of the providers responded with information that has helped them verify some of the inferred relationships.

The main conclusion is that with BGP derived inference, it is possible to identify less than 50 percent of peer-to-peer links. Another conclusion is that nearly all relationships are p2p and c2p, as confirmed by the conducted survey.

Techniques known as *tomography* have also been tried in discovery of topology of networks. Duffield and Presti [43] evaluate the use of multicast probing and end-to-end delay measurement for topology and bottleneck discovery. Multicast traffic is suitable for this since a given packet only appears once on a link in the multicast tree. End-to-end characteristics seen at different endpoints are then highly correlated. The collection of multicast tree delays and their corresponding estimators are then used to infer the tree topology.

Focusing on loss rate as the performance metric, they evaluate two algorithms. The first, the minimum variance weighted average (MVWA) algorithm, performs inference on each tree separately, returning a weighted average of the estimates taken from the different trees for each link. However, this procedure may not always be able to infer the behavior of links even when their loss rates are identifiable.

The second algorithm, the expectation-maximization (EM) algorithm, applies the standard expectation-maximization technique [44] to the measurement data taken from all of the trees. It returns estimates of the loss rates of all identifiable links. This work is done simulation using the Network Simulator 2 (NS2).[15]

When focusing on AS-level graphs of the Internet, peering relationships play an important role in providing alternative routing and resilience. Muhlbauer *et al.* [45] focus on the connections between the ASs within the Internet, due to the importance of the inter-AS relationships. Peering relationships are difficult to infer due to the business nature of this information and the limited ability of methods to correctly identify such peering relationships. However, their importance is significant as they affect interdomain routing policies. They build a simple model that captures such relationships by using BGP data from observation points such as Routeviews and RIPE. They then use simulations to provide an AS-level map that they compare with the BGP data from other vantage points.

In a view inspired by the business relationships of providers, Chang *et al.* [46] present a model of the economic decisions an ISP or AS has to make in order to peer with other ASs with transit tier 1 ASs. The economic decisions that have to be considered by an ISP are of three types: peering, provider, and customer. In each case the cost-centric multilateral decision, as referred to by the providers, has to bring mutual benefits for both parties. The gravity model [47] has been used to describe decisions on traffic demand and exchange. The distance of ASs from each other plays a critical role in the decision made by an AS to peer with another. They use BGP data to form node degree distributions to infer peering relationships. An important result of their work is an analysis of changes in the topology of a network, by introduction of new peering relationships and updates to the current ones.

Muhlbauer *et al.* [48] investigated the role and limitations

[13] *http://www.ripe.net/db/irr.html*

[14] *ftp://ftp.arin.net/pub/rr/arin.db*

[15] *http://www.isi.edu/nsnam/ns*

of business relationships as a model for routing policies. They observe that popular locations for filtering correspond to *valleys* where no path should be propagated according to inferred business relationships. This result reinforces the validity of the valley-free property used for business relationship inference. This work reveals two dimensions to policies:

- Which routes are allowed to propagate across interdomain links (route filtering)
- Which routes among the most preferred ones are actually chosen (route choice) and thus observed by BGP monitors

They use BGP data from more than 1300 BGP observation points, including Routeviews. The observation points are connected to more than 700 ASs with some feeds from multiple locations. They provide a model of ASs and have identified sets of per-prefix policies in order to obtain agreement between the routes selected in their model and those observed in the BGP data.

They report that the business relationships do not contain enough information about the path choices made by ASs. They introduced a new abstraction: *next-hop atoms*. Next-hop atoms capture the different sets of neighboring ASs an AS uses for its best routes. They show that a large fraction of next-hop atoms correspond to per-neighbor path choices. Some path choices, however, do not correspond to simple per-neighbor preferences, but hot-potato routing and tie-breaking within the BGP decision process, more detailed aspects of the Internet routing process.

Although most AS-level inference efforts are based on BGP data, they may be subject to errors in the BGP data itself. Feamster and Balakrishnan [49] tried to detect faults in BGP configurations. They crawl through BGP data to detect two classes of faults: route validity faults, where routers may learn routes that do not correspond to usable paths, and path visibility faults, where routers may fail to learn routes for paths that exist in the network. They have analyzed deployed configurations from 17 different ASs and detected more than 1,000 BGP configuration faults. This adversely impacts the quality of the inferred AS graphs.

Routescope [50] is a tool developed for inference of AS-level forwarding paths between two endpoints without direct access to them, using the shortest policy paths in an AS graph obtained from the BGP tables. The types of relationships considered in this case are peers, customers and providers. It is assumed that the shortest AS paths are always preferred, and that routing is uniform within an AS and for any destination AS, i.e., the hop count is always the sxame from all sources to the same destination AS.

Based on these assumptions, RouteScope uses existing algorithms [37, 39, 40] to infer AS relationships and categorize the types of relationships. Comparing the inferred AS paths with those of the BGP tables from three networks, they find the existing algorithms to be of low accuracy. The challenges faced here are due to asymmetric routing between pairs of nodes, policy routing between providers, multihoming of ISPs and misconfigured BGP table entries.

The proposed inference method of Mao *et al.* [50] starts by initializing all the links as down-links i.e., provider to customer. They then use a random walk to infer the AS path between the nodes. This approach is shown to have an accuracy of more than 60 percent when compared with BGP tables. There have also been algorithms suggested for inferring the first-hop AS from a source S to destination D, where there is no direct access to D.

Most of the inference methods listed focus on the mathematical view of a network. However, alongside statistical graph properties such as node degree distribution, it is also worth considering features such as link capacities and latencies. Adding such metrics enables the provisioning of accurate Internet maps at various levels, from an ISP network inter-AS connectivity.

Geographic Location of Nodes

Connectivity of one node to another is a measure of many factors such as hop count, delay, and available bandwidth. Another issue is that there is no location information associated with any node on the Internet. This has led to consideration of the *relative distance* between nodes. In order to create a map of the Internet from any point, researchers have used coordinate systems such as landmarking. Some of these techniques are discussed in this section.

An important area where location of nodes plays a significant role is visualization, the representation of the discovered nodes on a geographical map. One of these attempts was in the Mapnet project.[16] Mapnet is a tool for visualizing the infrastructure of certain international backbone providers. Each backbone infrastructure is divided into a group of nodes (PoPs) and links between these nodes, plotted based on the geographical location of the PoPs on a map of the world.

Landmarks are needed in the Internet in order to enable the nodes to measure distance between themselves accurately. In one of the first efforts to identify such distances, Francis *et al.* [51] introduce IDMaps, which aims to provide an estimate of the distance between any two valid IP addresses on the Internet. The authors have listed various methods of finding node locations on the Internet by use of a distributed set of measurement points and triangulation schemes as a method to estimate distances on the Internet. In this method the distance between the collection points, called *Tracers*, is a known measure. The distance between any two nodes is their distance from the nearest Tracers, plus the distance between the two Tracers.

One of the challenges in such a scheme is to find the minimum number of monitoring points, also known as *centers*, in order to minimize the maximum distance between the nodes and the nearest center. They use topologies generated by the Waxman random model [52], the Tiers generator [53], and the Inet topology generator tool [54] to evaluate their algorithms. They find that mirror selection for clients using IDMaps gives noticeable improvement over random selection. Their study provided positive results to demonstrate that a scalable Internet distance map service can indeed be built.

The use of coordinates on the Internet was explored by Ng and Zhang [55] in an attempt to create a Global Network Positioning (GNP) system. GNP was based on absolute coordinates computed from modeling the Internet as a geometric space. As opposed to the IDMaps where a client-server model was adopted, GNP is based on a peer-to-peer system. which gives it a more scalable architecture. In this approach each node has been assigned a set of coordinates that allows it to compute its distance relative to the other nodes. In this scheme a small number of distributed base nodes are deployed over the Internet. Each host on the Internet measures the round-trip time between itself and the base nodes using ping messages, and takes the minimum of several measurements as the distance. These distances are used as the end host's coordinates.

The strategic choice of landmarks for an Internet coordinate system is discussed by Tang and Crovella [56], looking at the challenges involved due to the size of the Internet and

[16] *http://www.caida.org/tools/visualization/mapnet/*

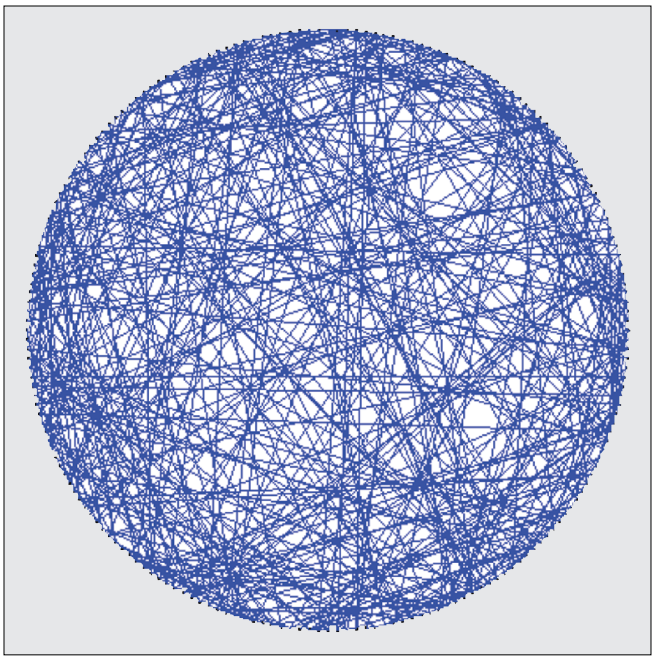

Figure 8. *A random network of 200 connected nodes, with two incoming connections per node.*

lack of end-to-end path visibility, and hence knowledge of properties of perfect landmark sets. The authors look at clustering methods in order to analyze different methods of selecting landmark sets. They use Lipschitz coordinate systems, using three distinct data sets to test their four algorithms for selection of landmarks: greedy, *k*-means, minimum distance, and random. The greedy method is found to be the most efficient method for a small topology. However, random landmark selection, for more than 10 landmarks, is the best method, in terms of comparative accuracy and simplicity.

Lua *et al.* [57] discuss the accuracy of use of RTT for geometric spacing of nodes. These estimates are sometimes inaccurate, and it is difficult to identify the quality of the distance estimates. The embedding of nodes varies greatly with different numbers of landmarks even if the topology of the network is fixed. They propose two new metrics of quality, node distance from others (its rank or proximity) and closest neighbor loss, which defines how many mistakes are made in identifying closest neighbors. They perform their embedding using nodes on Planetlab,[17] the Waxman topology generator data, and RouteViews data. They use Dijkstra [58] to find shortest path distance matrices, and use their coordinate and embedding method to test the metrics. In conclusion, the optimal choice of metrics and coordinate placement remains an open problem.

Geographical location of landmarks plays a critical role in their usefulness. Lakhina *et al.* [59] focus on relationships between router location and population density, geographic location and link density, and size and geographic extent of ASs. They disprove the Waxman topology generator's assumption of uniform router distribution but find that connectivity degree exponentially decreases with distance of routers from each other, as modeled by Waxman. They use Skitter data from CAIDA, without alias resolution so that each interface is a node, and also the Mercator data set, focusing on routers and the links between them. They then use tools that infer geographic locations of IP addresses, using DNS LOC records, and WHOIS records and host-name mapping. A comparison is made between the population of people in continents and number of interfaces, taking into consideration the economical strengths of countries. They show these to be similar in economically homogeneous regions.

Despite the efforts of researchers in the field of coordinate systems, there is still no accurate system in place. Zheng *et al.* [60] show that triangle inequality violations (TIVs) might be exploited by overlay routing if an end-to-end forwarding path can be stitched together with paths routed at layer 3. However, TIVs pose a problem for Internet coordinate systems that attempt to associate Internet hosts with points in Euclidean space so that RTTs between hosts are accurately captured by distances between their associated points. Three points having RTTs that violate the triangle inequality cannot be embedded into Euclidean space without some level of inaccuracy. This is a constraint that is put on the Internet as a result of various routing policies and traffic engineering. They have used router-level topology and IS-IS weights from the GEANT backbone, a multigigabit pan-European research network. They also measured the minimum RTT values between all pairs of GEANT backbone routers using their looking glass interface. They show that typical interdomain and intradomain routing policies almost guarantee that RTTs are not a metric space (as they violate the triangle inequality), so embeddings of coordinate systems face mathematical problems.

Models of Internet Topology

Mathematical modeling of the characteristics of the Internet is a key stage for successful generation of realistic topologies. These mathematical models can range from geographical distance and clusters to distribution of nodes with different degrees of connectivity. In reality, the constant change in the Internet topology makes it difficult to obtain a single topology of the Internet and instead it is more appropriate to refer to the obtained maps as Internet *topologies*.

In this section we present some of the models of Internet topologies. The objective of this section is to familiarize the reader with the common methods of characterizing the topology of a network and provide a basic understanding of the most common terms used in this context.

Random Graphs

Complex networks such as the Internet have traditionally been described using the random graph theory of Erdős and Rényi [61]. In a simple model, for a given number of nodes n, edges m, and the average degree $k = 2m/n$, one can construct the class of random graphs having the same average degree k by connecting every pair of nodes with probability $p = k/n$.

Figure 8 displays a random graph of 200 nodes, with 2 incoming connections per node. It can be observed that there are no dense cores as all the connections are spread equally between the nodes. This observation is in contrast with the Internet, where some routers have many connections and hosts, and some have as few as one host connecting to them.

Despite the ease of use of the random network model and their ability to produce some of the required metrics for a generator such as average node degree, they were abandoned in favor of models that capture the statistical characteristics of the Internet, as discussed in the next section.

Power Laws in Topologies

Power laws are one of the most widely used parameters in the context of topology analysis of the Internet. Power laws are

[17] *http://www.planet-lab.org/*

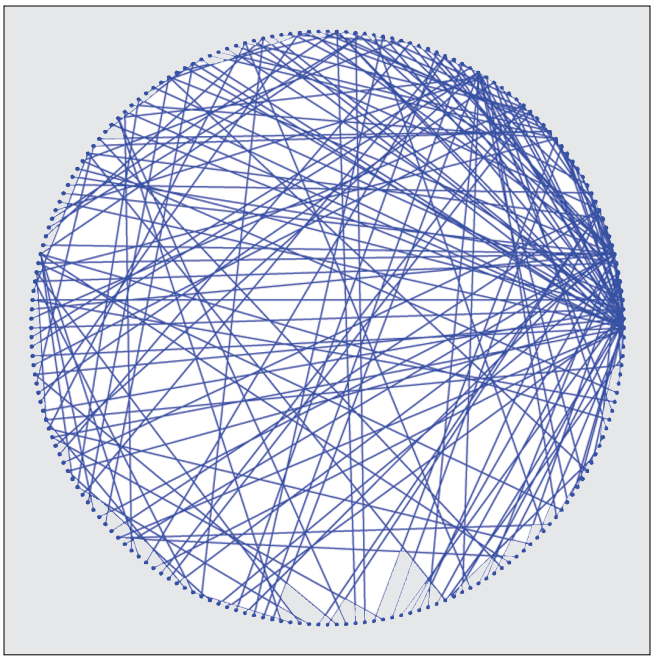

Figure 9. *A power law network of 200 nodes.*

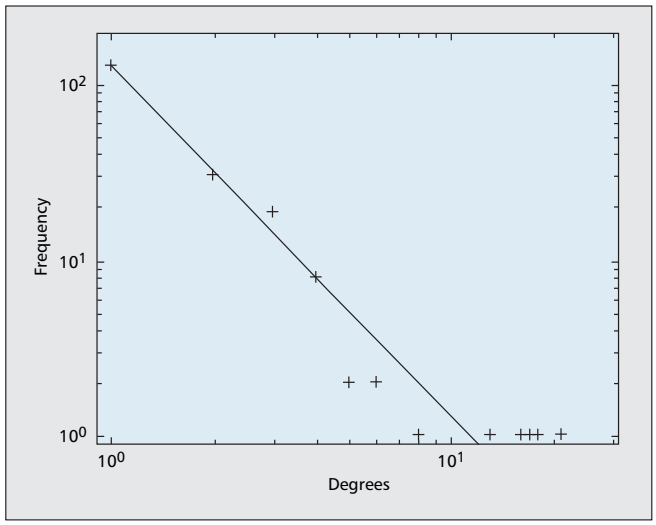

Figure 10. *Power law node degree distribution.*

seen in phenomena where there is no concept of scale variance; that is, a property, such as a distribution of nodes in a network, follows the same rules at different scales or resolutions. In a seminal paper Faloutsos *et al.* [22] stated that certain properties of the AS-level Internet topology are well described by power laws. In this work the authors use three Internet instances (topologies inferred from BGP tables). Three specific power laws were observed, and these were believed to hold for the Internet:

- *Rank exponent*: Out-degree of a node is proportional to its rank to the power of a constant.
- *Out-degree exponent*: The frequency of an out-degree is proportional to the out-degree to a constant power.
- *Eigen-exponent*: The eigenvalues of the adjacency graph are proportional to the order *i* to a constant power.

Such a graph will have power law characteristics, and it has a tree-like structure. If one relies on the traceroute tool, it is difficult to infer the cross-links between the nodes. A scale-free network is not a homogeneous network as the nodes have a very heavytailed distribution of the number of connections. Despite the small size of the Internet at the time of observations, these observations were believed to hold in future growth stages of the Internet. This hypothesis intrigued Siganos *et al.* to repeat the above analysis [62].

They prove the existence of power laws in the Internet at the AS level, looking at two topology measurements and a few snapshots over five years, one from Oregon RouteViews and another the dataset used by Chen *et al.* [63]. The test for the existence of power laws is carried on metrics such as rank exponent, degree exponent, and eigenvalues. The conclusions are that the power laws exist over a five-year period, and they are an efficient way to describe metrics of topology graphs.

One of the classic models used in this context is the *BA* model, introduced first by Barabási and Albert [64]. This model is based on the incremental growth of networks, by addition of new nodes and preferential attaching of nodes to well connected ones. They also reported that the Internet has power law characteristics, alongside the findings of Faloutsos *et al.* Barabási and Albert focus on Web pages and links between them as an alternative measurement of the Internet. Figure 9 shows a network of 200 nodes connected based on the BA model.

Figure 10 displays the node degree distribution of the power law network in Fig. 9, plotted on a log-log graph. Existence of a straight line indicates the existence of a power law distribution of node degrees.

The existence of power laws in the Internet is interesting as the Internet is formed from smaller networks that are self-managed. Medina *et al.* [65] look at four factors in formation of Internet topologies that may cause various power laws inferred on the Internet:

- Preferential connectivity of nodes to nodes with more connections
- Incremental growth of the networks
- Distribution of nodes in space (random or heavytailed)
- Locality of edge connections (preference to connect to nearby nodes)

The BRITE topology generator [7] was used by Medina *et al.* to test these hypotheses. Topologies of between 500 and 15,000 nodes were considered, with and without incremental growth and preferential connectivity.

The final conclusions are that the rank and out-degree power laws are more effective in distinguishing topologies than the number of hops between nodes and eigenvalue power laws that are observed similarly in all topologies. Preferential connectivity and incremental growth are found to be the main causes for all power laws in the simulations. They establish that for best correlation coefficients (approaching 1) and slope of linear fits for rank exponents (approaching 0.5 observed by Faloutsos *et al.* [22]), both preferential connectivity and incremental growth must be present. This methodology can be extended by grouping nodes into administrative domains.

The findings in this section indicate the existence of power laws in various statistics extracted from the Internet. However, the inferred statistics are not always perfect as one cannot obtain a single snapshot of the Internet topology and must rely on various measurement techniques. We now present results which indicate that the existence of power laws are merely a side-effect of poor inference techniques.

Arguments against Power Laws

The inherent biases of traceroute sampling and collection of BGP data from limited vantage points made researchers question the true existence of power laws in the Internet AS-level topology. Chen *et al.* [63] state that BGP data represents a partial view of the Internet; hence, power laws may not exist in the strict form suggested by Faloutsos *et al.* [22] for AS

connectivity degree distribution. This argument is based on their findings that BGP AS paths do not completely capture the topology, and the data from Routeviews suggest that the node degree distribution is perhaps heavytailed (close to Weibull distribution), and perhaps only the tail exhibits power laws. The authors use BGP routing tables of 41 ASs and information from Looking Glass Websites to infer the local AS connectivity map and compare it to the one achieved by Routeviews. Data from the European Internet routing registry (RIPE), which has the peering relationships of most European ASs, is used in order to find relationships that are not seen from BGP inference, such as siblings [66]. Another observation in conflict with the existence of power laws is the important observation made by Mahadevan *et al.* [1]. For a comparative study, three distinct data sources are used:

- Traceroute data from the CAIDA Skitter project, using the 31 daily graphs for the month of March 2004
- Routeviews BGP data for March 2004, including static table and updates
- RIPE WHOIS database dump for April 7, 2004

For the traceroute data, both multi-origin ASs and AS-sets create ambiguous mappings between IP addresses and ASs, while private ASs create false links. Hence, AS-sets are filtered to multi-origin ASs and private ASs. Indirect links are discarded. All daily graphs are then merged in order to form one graph. Similar tasks are performed on the BGP data, one from the static tables and one from the updates. In both cases, AS-sets and private ASs are filtered, and the 31 daily graphs are merged into one. When using the WHOIS data, the records of interest are:

```
aut-num: ASx
import: from ASy
export: to ASz
```

which indicate links *ASx – ASy* and *ASx – ASz*. They then construct an AS-level graph (referred to as a WHOIS graph) from these data and exclude ASs that did not appear in the `aut-num` lines.

The findings confirm that the Skitter data displays power law characteristics [22]; however, the WHOIS graph has an excess of medium degree nodes, and hence its node degree distribution does not follow power laws. They also compared many metrics of the Skitter and RouteViews graphs to those graphs generated based on power law random graphs (PLRG) [67], and it is observed that the PLRG model fails to accurately capture the important properties of the Skitter or RouteViews BGP graphs. Similarly, the PLRG model fails to recreate the WHOIS graph since its node degree distribution does not follow a power law at all.

Alternative Topology Models

Power laws were not the only point of interest for network researchers who used data sets from various inference projects. For example, the graphs produced by Rocketfuel and Skitter consist of physical connectivity of Internet routers for an ISP or a section of the Internet. However, for an improved understanding of the physical infrastructure of the Internet, it is essential to have more information about the common characteristics of links such as the link bandwidth and router capacities. These concerns were first raised by Alderson *et al.* [68], where they focus on annotated graphs of the Internet at the IP layer with addition of bandwidth and buffer sizes. The Abilene[18] and Rocketfuel maps are used to look at various differences between network models, by use of a metric proposed as *network performance*, defined as the maximum throughput of a network under a gravity model of end-user traffic demands. Hence, their proposed design for an ISP network graph is referred to as *Heuristically Optimal Topology* which is based on having sparsely connected high speed routers at the core of the network, supported by hierarchical tree-like structure at the edges. This is similar to the proposed *Highly Optimized Tolerance* approach suggested by Carlson and Doyle *et al.* [69] and *Heuristically Optimized Trade-offs* considered by Fabrikant *et al.* [70]. The main contribution of this work is in comparing the first principles [71] approach on five toy networks:

- A graph constructed from a preferential attachment model (BA model) of Barabasi and Albert [64], where nodes are added and connected with a probability proportional to an existing node's current degree
- A construction based on random graphs explained By Waxman [52]
- A construction based on the proposal by Alderson *et al.* [72]
- Abilene-inspired topology by extracting a graph of the Abilene core network
- Suboptimal topology, intentionally designed for poor performance

The authors propose that detailed study of the technological and economic forces shaping the router-level topology of a single ISP provides convincing evidence that the Internet is not necessarily formed of highly connected routers in the core of the network. They expect border routers again to have a few relatively high-bandwidth physical connections supporting large amounts of aggregated traffic. In turn, high physical connectivity at the router level is again expected to be confined to the network edge. They also note that modeling router-level robustness requires at a minimum adding some link redundancy (e.g., multihoming) and incorporating a simple abstraction of IP routing which accounts for the feedback mechanisms that react to the loss or failure of a network component.

Magoni and Pansiot [73] use BGP data from the Routeviews project to look at AS relationships. Alongside traditional AS types, there are other types of ASs that are used for classification purposes by the authors [73]:

- Cycle AS: an AS belonging to a cycle, being on a closed path of disjoint Aes
- Bridge AS: an AS connecting two cycle ASs only
- In-mesh AS: a cycle AS or bridge AS
- In-tree AS: an AS belonging to a tree; the opposite of an in-mesh AS
- Branch AS: an in-tree AS with a connectivity degree of 2 or more
- Leaf AS: also referred to as a stub AS
- Root AS: an in-mesh AS that is the root of a tree (i.e., it is adjacent to two or more in-mesh ASs and to one or more in-tree ASs)
- Relay AS: an AS with two connections
- Border AS: an AS located on the diameter of the network
- Center AS: an AS located on the core of the network

They report on various statistics such as connectivity and shortest paths, and confirm the observed power laws at the AS level. They also validate some of the power laws in metrics such as number of shortest paths exponent, and compare Mercator and nec maps. The variation in power law exponents suggest that the power laws and average values are changing and may not hold true in the future.

[18] *http://abilene.internet2.edu*

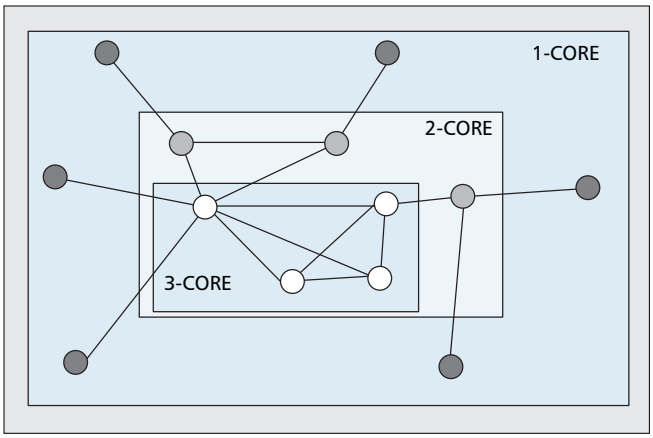

Figure 11. *Example of* k*-core decomposition [77].*

Structural Models of the Internet

Alongside power laws, other metrics of network topologies have been studied extensively in the literature. One of the most important factors that has already been explained in this section is the clustering of nodes. Clustering has been widely studied using techniques to find the clustering coefficient of the nodes in a network. An alternative to this method is *spectral filtering*. Gkantsidis *et al.* [74] perform a comparison of clustering coefficients, by using eigenvalues of adjacency matrices from various BGP data of networks, and also on methods of topology generation such as BRITE. This work identifies a global problem with topology generators, which is the lack of data to generate the large topologies required for a realistic simulation. Use of a small topology leads to concentrating only on the AS and router-level geographic topologies, as opposed to looking into the peering relationships, clustering, and amount of traffic on the links. They have introduced the basics of degree-based graph generation and conditions to which the links and nodes are attached to ensure existence of spanning trees using a Markov-chain-based algorithm.

They believe that degree sequence is not sufficient for topology generation that matches the real data. They use clustering methods and eigenvalues to analyze the generated topologies and compare with real data from NLANR.[19] The generation methods that meet a degree-sequence while incorporating clustering are suggested by researchers. Good clustering methods are also needed in topology generators, as both the degree-sequence and clustering are found in real networks [74].

Li *et al.* [71] discuss the need for topology inference and generation at different levels. For congestion control protocols, IP-level connectivity with bandwidth and buffer sizes is needed, while for attack assessment and network planning, a detailed map of node and router capacities is required. For routing protocols one needs a graph of AS-level connectivity and peering information. The authors focus on node degree distribution and their heavytailed characteristics, and whether the node degree distribution is the most important objective of a topology. They discourage the use of random generators as they do not produce power laws in node degrees, so they have been replaced by degree-based generators. The proposed *first principles approach* focuses more on the physical layer, router, and links. In the context of network engineering for an ISP, physical metrics such as performance and likelihood are used for graph generations. The first principles method is presented and compared against a few toy models. They observe that simple heuristically designed and optimized models that reconcile the trade-offs between link costs, router constraints, and user traffic demand result in configurations that have high performance and efficiency.

The Internet has a hierarchical structure in the form of different tiers. Jaiswal *et al.* [75] look at comparing the structure of power law graph generators and that of the Internet AS graph. This is an important step in proving the existence of power laws. By decomposing graphs of the Internet at different levels, the authors establish the properties of power law graphs and the Internet graph, and find skewed distributions in degree connectivity; that is, a large number of less connected nodes connect to well connected ones, and well connected ones tend to interconnect more closely. The decomposition procedure is as follows:

- Given an input graph, G, select a set of nodes to be removed based on defined criteria to identify the root-level transit nodes.
- Compute the connected components (CCs) of the graph G obtained after removing the selected nodes.
- Repeat the procedure recursively on the resulting CCs until the number of nodes that can be removed is less than 1.

The criteria for removing nodes are based on *node degree*, *node rank*, and *node stress*. They order the nodes in each connected component in descending order of degree and choose a fixed fraction of the highest degree nodes to remove. Node rank is based on the steady-state probability of visiting a node while performing a random walk in the graph where each outgoing link from a node is equally likely to be chosen. The stress of a node in a graph is a measure of the number of node-pair shortest paths that pass through it.

Carmi *et al.* [76] use the data from the DIMES project, combined with AS-level maps from the RouteViews project, to form a map of the Internet. The map formation method is based on *k-shell decomposition*, which involves removing nodes in groups based on the number of connections they have to form shells of nodes. In the first step the *k-pruning* technique is performed by removing all the nodes with only one neighbor recursively, as well as removing the link to that neighbor along with the node. The nodes removed in this step are called the 1-shell. This process carries on with index k to form shells of higher connectivity degree. The last nonempty k-core will be, by definition, the backbone of a network such as the Internet. Figure 11 displays a sketch of the k-core decomposition for a small graph from Alvarez-Hamelin *et al.* [77]. Each closed line contains the set of vertices belonging to a given k-core, while colors on the vertices distinguish different k-shells.

Carmi *et al.* found that for the DIMES data used, the size of each k-shell decreases with a power law distribution, $n(k) \propto k^{-\delta}$, where the exponent δ is about 2.7.

In the topology modeling area, as described by Chang *et al.* [78], a challenge faced by researchers is the *fitting* problem where a trend in networking is forced to fit a known mathematical model. Such methods are only accurate enough to be justifiable when they are supported by evidence gained from identification of complementary measurements that complete the picture. By looking at the Abilene network traffic matrices, they analyze the distributions on that using cumulative distribution functions that illustrate an 80–20 power law,[20] and go on to compare it to the GEANT network[21] traffic matrices; these traffic matrices fail to fit the given gravity models. The

[19] *http://www.nlanr.net*

[20] *That is, 20 percent of the nodes own 80 percent of the traffic.*

[21] *http://www.geant.net*

gravity model corresponds to an assumption of independence between source and destination of the traffic, and can be written as a matrix formed from the product of two vectors. The gravity model is shown to be able to replicate some statistics of the actual traffic matrices very well; hence, by characterizing the traffic vectors, it is possible to obtain a reasonable characterization of the matrix. The difficulties faced by the authors in obtaining AS-wide traffic volumes and topologies from ISPs highlight the challenges faced by researchers in this field.

Clustering techniques for nodes has also been a method of characterizing topologies in the Internet. Wool and Sagie [79] propose a clustering method that enables the view of Internet topology as AS-graphs in different granularity levels. They find a few main dense cores, which interconnect the regional cores. They compare various degree-based generators, and state the need to consider power laws and clustering coefficients when generating topologies in BRITE and Inet. They use the dense k-subgraph approach for clustering in different levels.

Yook *et al.* at [80] propose a model of networks based on fractals. They find that the physical layout of nodes form a fractal set, determined by population density patterns around the globe. The placement of links is driven by competition between two models: preferential attachment and linear distance dependence. Preferential attachment assumes that the probability that a new node will link to an existing node with k links depends linearly on k. The nodes with higher connectivity degree are more desirable for attachment by new nodes. Preferential attachment is believed to be one of the main reasons for power-law properties of the Internet. Linear distance dependence is due to the fact that the further the nodes are from each other, the less likely it is for them to have a direct connection.

Zhou and Mondragon [2] propose various mathematical models, such as rich-club phenomena, the interactive growth model, and betweenness centrality. In a follow-up [81], they use the CAIDA data to look at degree-degree correlation and rich-club connectivity. They show that for these data, rich-club connectivity and degree-degree correlation for a network with a given degree distribution are closely linked. This leads to a proposed model, Positive Feedback Preference (PFP).

The PFP model starts with a random network of size n. At each time step:

- With probability p, a new node is attached to a host node, and at the same time a new internal link appears between the host node and a peer node.
- With probability $q \in [0, 1 - p]$, a new node is attached to a host node, and at the same time two new internal links appear between the host node and two peer nodes.
- With probability $1 - p - q$, a new node is attached to two host nodes, and at the same time a new internal link appears between one of the host nodes and a peer node.

Zhou and Mondragon demonstrate that the PFP model produces graphs that closely match the degree distribution, rich club connectivity, and maximum degree of the AS graphs of a given network. The model is validated against BGP derived data.

The Internet, like many complex networks, is believed to have small world characteristics. Such characteristics are important for delivery of messages and content on networks. Jin and Bestavros [82] consider the small world characteristics when generating topologies at the router and AS levels. At the AS level, the high variability in node degree, and at the router level the preference for local connectivity result in this phenomena. They use simulation of multicast trees on different models. They also use AS graphs of the University of Michigan AS graph data set (RouteViews plus Looking Glass), and various router-level graphs including Skitter. They use these to get the statistics such as node degree and local connectivity in order to evaluate their model. They suggest simulators taking into consideration vertex degree distributions as well as preference for local connectivity, and suggest improvement by considering scale-free characteristics as well.

Some of the power law traits suggested by Faloutsos *et al.* are also seen in other telecommunications networks. Spencer *et al.* [83] showed that a national synchronous digital hierarchy (SDH) transport network also exhibits such traits, even though it has an explicit hierarchy and strict technological structure, and is much more closely coupled to the topology of the physical layer. They showed that the topology formed by SDH circuits followed similar power laws in degree ranking, degree frequency distribution, largest eigenvalues, and hop-plot approximation; they also demonstrated a power law in the clustering of nodes and that the degree frequency distribution is very pervasive, appearing at a range of geographic scales. Since their topology was captured directly by the network operator, it does not suffer from the possible sampling bias that may be seen in the Internet topology measurements by traceroute.

The Internet architecture and structure is constantly evolving. Pastor-Satoras and Vespignani [84] highlight the self-organizing nature of the Internet and its evolution since birth from a statistical and physical viewpoint. Their conclusion is that the Internet can be modeled as a network of nodes and links growing in a scale-free manner. However, the growth and death rates of ISPs and ASs, and predictions of future trends on the Internet remain open issues.

This section has gathered various models that are presented for the Internet at physical and routing levels. The variety of models is an indication of the complex structure of the Internet, which makes it difficult to capture all the characteristics with a simple model. Based on these models, researchers develop topology generators, discussed in the next section.

Comparison of Topology Generation Models

Despite the availability of many topology models, there has not yet been an agreement between researchers on a single standard method of modeling and generation of Internet or ISP network topologies. This problem is due to the many aspects one has to consider when studying a topology.

Tangmunarunkit *et al.* [85] analyze the differences between different classes of topology generators, The focus of degree-based generators is the local degree distribution, while the old-style structural generators focus on hierarchical properties of networks. The authors assume that correct large-scale hierarchical generation of topology is more important than retaining local properties such as degree distribution. However, they reach the conclusion that degree-based generators are better at representing such large-scale properties. Degree-based generators represent the power law properties better. This hypothesis has been verified using two representations of the Internet, one at the AS level using BGP routing tables, and the other at the router level using IP next-hop connectivity measured by traceroute.

Three types of generators are used by the authors: random (Waxman), hierarchical (Tiers [53] and Stub), and degree-based (PLRG and BA model). They choose three metrics: expansion (nodes reachable in h hops), resilience (number of alternative paths between nodes), and distortion (a spanning tree of the graph that has the least total cost) for evaluation of generators. Based on the metrics and those measured from Internet BGP tables, they conclude that PLRG has a better qualitative match to the Internet, with high expansion and resilience and low distortion, indicating that based on link

value distribution (a measure of resilience), PLRG graphs qualitatively model the hierarchy present in AS- and router-level graphs better than structural generators due to their long-tailed degree distribution.

Chang *et al.* [86] look at the problem of generating AS-level topology of the Internet. They discuss the weakness of current power-law-based generators and BGP-inferred AS topologies in detecting AS peering and business relationships. The authors focus on the optimization of a topology based on AS geography, business model, and evolution in time using the RouteViews data plus inferred information from Looking Glass sites to form two data sets. For simplicity, all multihoming and multiple connections of ASs are removed by choosing just one link based on criteria such as lowest average hop distance. The final graph is 50 percent of the size of the original data set, with similar node degree distribution.

Alderson *et al.* [72] discuss generating topologies using the Highly Optimized Tolerance concept. In this strategy the focus of the generator is the economic trade-offs, such as cost and performance, and technical barriers faced by an ISP when designing its own network. This would allow for a focus on the economical challenges faced by network operators. These issues are important for backbone service providers, which must ensure optimized use of the network capacity.

Bu and Towsley [87] focus on generating similar power laws to those observed in measurement data from networks, focusing on the "power law generator" class of topology generation tools. They encourage the use of empirical complementary distribution (ECD) in generation of nodes as opposed to histograms, and show (using characteristic path length and clustering coefficients) the variability in graphs of different generators using the same heuristics. They use a probabilistic method to generate topology by adding nodes and links one by one based on the probability within the ECD. This model is an advancement in the direction of meeting metrics of power law graphs. This model has been incorporated in the BRITE tool.

Mahadevan *et al.* [88] discuss the lack of analysis and topology generation tools that can focus on specific requirements of metrics of a graph, focusing on degree correlations of subgraphs of a graph that represents a network or Internet. However this method becomes extremely complex as the number of correlated nodes increases. In a basic model, a set of subgraphs are defined with various distributions and are used to define a topology. The metrics considered for analysis are spectrum, distance distribution, betweenness, node degree distribution, likelihood (sum of products of degrees of adjacent nodes), and clustering. However, in practice, the focus has been put on connectivity as the other metrics are hard to compare and classify. They focus on reproducing a given network topology, and compare their results with the Skitter data set and BGP data from RouteViews.

Mahadevan *et al.* believe an improvement in topology generation can be achieved by focusing on peering relationships and graph annotations such as bandwidth and latency. In Orbis [89] the aim is to produce a random graph of desired size while keeping the characteristics of the input graph. They allow a user to feed in average degree, node degree, and joint degree distributions from a measured topology, and the tool should also annotate the routers with AS memberships and annotate the AS links with type of relationship between them.

They have verified their algorithms against router-level topologies observed from Skitter traceroute data in September 2006 and AS-level statistics from five years of RouteViews data. The longest matching prefixes of the IP addresses of the router-level topologies are used to find the AS membership data. This allows AS membership information to be available from the router-level topology data. This information is then used with logarithmic binning to group ASs into categories with equal numbers of routers within an AS.

They observe that AS-level topologies can be approximated by power laws. However the router-level topology does not fit strict power laws. The observed maximum degree at the router level does not increase significantly by increasing the size of the graph. In 1*k*-rescaling, they attempt to preserve the shape of the probability distribution function (PDF) of the graph's degree distribution. In 2*k*-rescaling, they try to preserve the degree correlation profile. They encourage the addition of latency and bandwidth distribution as another metric for rescaling for realistic topology generation.

A problem usually faced in generation of a topology is generation of realistic traffic matrices. While it is important to have a network topology, for it to be useful, related traffic matrices must be assigned to the links. Nucci *et al.* [90] discuss two issues with traffic matrix generation for a given topology: the issue of generating traffic matrices that look like a realistic network, and that of assigning metrics such as traffic volume to nodes. They focus on synthesis of stationary traffic matrices. They propose a technique based on distribution fitting.

On the placement of values (metrics), they argue why the values should not be arbitrarily added to links and nodes, as the link capacities may be exceeded, or the links that are there just for resilience are incorporated into carrying traffic, which changes the temporal topology of the network. Any uniform distribution in traffic generation is thus considered inappropriate, and they suggest a lognormal distribution, using measurements from the Sprint and Abilene to test the proposed hypothesis. In conclusion, the proposed load minimization solution (for matching the metrics assigned to links to a given topology) is considered too computationally expensive to be implementable.

However, there are still arguments that modeling and generation of graphs at large scales remains an unsolved problem. Krishnamurthy *et al.* [91] introduce graph sampling in order to reduce the size of inferred topologies for analysis while preserving metrics, in this case power laws and slope of graphs. They model the network as an undirected graph at the AS level. They propose sampling the graph by deleting nodes and links probabilistically, contracting the graph at steps, or generating a subset of graphs from traceroute paths. They perform probabilistic deletion of nodes and edges, and can reduce the graphs by about 50–70 percent while keeping metrics such as power laws within an acceptable range.

Simulation of enterprise networks is a difficult task as there are no benchmarks present for validation of the simulators. Mizuta and Nakamura [92] conduct a survey on business relationships in the IBM consultancy network, simulating network nodes as agents. They use the transactions obtained from email send/receive logs, together with the organization network structure diagrams to model and evaluate input/output degree, distance, and closeness. The presented model is probabilistic with random distances and random communication probability distribution.

One of the objectives of generation of topologies that closely map those of the Internet is to arm network researchers with tools with which they can analyze various issues in and around the Internet, such as congestion, optimal routing, and fault finding. Spring *et al.* [93] look at traceroute measurements, using *scriptroute*, from around 40 vantage points on planetlab to look at topology and routing policies, internal and between ISPs, to analyze the causes of path inflation, and find that interdomain routing and peering policies have significant effects on the inflation. They suggest improvements to BGP policies to look after routing across ISPs, as the

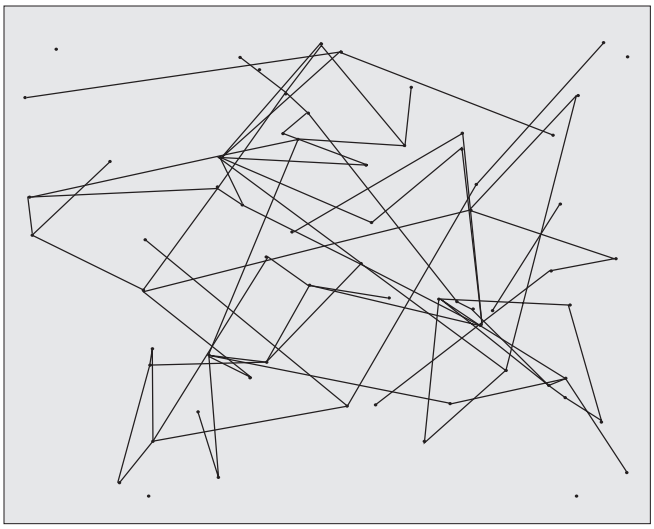

Figure 12. *A topology generated by the Waxman model.*

ISPs have to use minimum AS hop count, which may take longer sometimes. They compare the taken routes to the topology they inferred using Rocketfuel.

Topology Generation

For successful simulations of traffic and network events, any generated network model must be topology-aware. Topology generation is an area researchers have been actively working on in the last decades. The first generated topologies were randomly generated by selecting a certain number of nodes and randomly assigning links between them. This was due to the lack of understanding of the architecture of the Internet and the lack of validation tools. In this section some of the popular network topology generators are discussed.

Waxman

The Waxman generator is based on the random network model [61] where the nodes of the network are uniformly distributed in the plane, and edges are added according to probabilities proportional to the distances between the nodes. It was soon found that such generators fail to represent the important metrics present in the Internet topology; hence, they were abandoned in favor of hierarchical topology generators. Figure 12 displays a topology generated by the Waxman model. It can be seen that some nodes are not connected to others. Nodes are placed in the plane by a Poisson distribution process and connected with probability proportional to the Euclidean distance between them.

GT-ITM

With the explosion of the Internet, researchers realized that they needed to capture the structural properties and attempted to model the design of the Internet. The hierarchical modeling of the Internet topology was originally done by transit-stub models. Calvert *et al.* [94] presented one of the first results in this field by focusing on graph-based models to represent the topology. The parameters used include the number of transit and stub domains, number of local area networks (LANs) per stub domain, and number of edges (links) between transit and stub domains to initialize the topology generator. Then the transit domains, transit nodes, and their interconnecting edges are placed; similarly, the stub domains. The transit-stub model produces connected subgraphs by repeatedly generating a graph according to the edge count and checking the graph for connectivity. Unconnected graphs are discarded. This method ensures that the resulting subgraph is taken at random from all possible (connected) graphs; however, it may take a long time to generate a connected graph if the edge count is relatively small compared to the number of nodes. Extra edges from stub domains to transit nodes are added by random selection of domains and nodes.

Georgia Tech Internetwork Topology Models (GT-ITM), also known as the *transit-stub* generator, is capable of producing several forms of network topologies:

- *Flat random graphs*: GT-ITM has five models of topology embedded within it, including a pure random model and varieties of the Waxman [52] model. These are not hierarchical models.
- N-*level model*: The *N*-level model constructs a topology recursively. In this method a graph is made by dividing the Euclidean plane into equal-sized square sectors; then each sector is divided into smaller sectors in the same manner, so the scale of the final graph is equivalent to that of the individual levels.
- *Transit-stub model*: This model produces interconnected transit and stub domains. This model is controlled by number of domains, average node per transit domain, average stub domains per transit domain, and average nodes per stub domain.

In the transit-stub domain care has been taken to ensure that the paths are similar to those of the Internet; for example, the path between two stub domain goes through one or more transit domains and not any stub domains, and not the other way round. This is done by assigning appropriate weights to the interdomain edges.

The transit-stub model is comparable to the *tiers* model [53], in which the three levels of hierarchy, or tiers, are referred to as wide area network (WAN), metropolitan area network (MAN), and LAN levels, corresponding to the transit domains, stub domains, and LANs of the transit-stub method. The tiers model produces connected subgraphs by joining all the nodes in a single domain using a minimum spanning tree algorithm, a popular method used as the basis for laying out large networks. This generation method has been tried in two implementations of the transit-stub (TS) model, part of GT-ITM.

Inet — Inet is an AS-level Internet topology generator. Winick and Jamin [54] compare the Inet generated topologies to those obtained form RouteViews and NLANR BGP table data. They extract connectivity information into a simple adjacency list of ASs. The data set consists of 51 Internet topologies from November 1997 to February 2002. By default, Inet produces random networks with characteristics comparable to the Internet, as the minimum number of nodes is 3037 (the number of ASs on the Internet at the time of development), and the fraction of degree 1 nodes set to 0.3 based on measurements from the BGP data.

There are modifications made to the Inet model that differentiate the generated topology from a random network. For example, a preferential linear weight is factored into the connection probability, and the core of the generated topology is modeled as a full mesh network. However, Winick and Jamin believe Inet-3.0's topologies still do not represent the Internet well in terms of maximum clique size and clustering coefficient. These related problems stress a need for better understanding of Internet connectivity.

BRITE — One of the more popular network topology generators has been the BRITE generator [7]. BRITE produces

Tool	Year	AS-level	Router-level	Hierarchy
GT-ITM	1996	Yes	No	Yes
Inet	2000	Yes	No	No
BRITE	2001	Yes	Yes	No
PFP	2004	Yes	No	No
IGen	2006	No	Yes	Yes

Table 2. *A comparison of network topology generation tools.*

power law graphs on a large scale and is not suitable for smaller networks. BRITE provides an easy-to-use tool for generating a basic network model. BRITE can take as input maps from some other generators and NLANR topology formats, and, based on them, place nodes in planes, interconnect the nodes, assign delay and bandwidth and AS IDs to nodes, and output the topology. BRITE is capable of producing router-level topologies by placing nodes, assigning bandwidth based on distributions (either constant, uniform, exponential, or heavytailed), or based on the Waxman or BA model. BRITE can also create an AS-level model in a hierarchical topology model by assigning nodes and bandwidths, deciding hierarchy, and generating graphs.

The BRITE implementation has a two-level hierarchy model. In the top-down method it generates an AS-level topology (using the Waxman model, flat models, or imported data sets); then it generates a router-level topology for each AS using the same models.

In the bottom-up approach BRITE first generates the router-level topology and then assigns to each AS node a number of routers (using either constant, uniform, exponential, or heavytailed distributions). Then for each individual AS, a number of router nodes are assigned, either randomly or by performing a random walk on the topology and assigning routers to ASs.

BRITE has the following models embedded for AS-level and router-level topologies:

- Waxman: Uses the Waxman probability model [52] for generating random topologies.
- BA and BA-2: These models are inspired by the Barabasi and Albert [64] model of networks, and incorporate preferential attachment and incremental growth factors.
- GLP: Based on the generalized linear preference model proposed by Bu and Towsley [87].

BRITE is the first topology generator that makes an attempt to assign bandwidth to links. In an attempt to make AS-level and router-level topologies within them for hierarchical topologies, BRITE has incorporated the capability to produce two-layer topologies.

Positive Feedback Preference — PFP is an AS-level topology generator based on the model proposed by Zhou and Mondragon [2]. In this model the AS-level topology of the Internet is considered to be growing by the interactive growth of new nodes and links, and a nonlinear preferential attachment. The PFP model is described by the authors to represent many topological properties of the Internet such as degree distribution, rich-club connectivity, shortest path lengths, and betweenness centrality.

IGen — Another generator that aims to generate topologies which have the geographical problems associated with network design is the Igen generator. Quoitin [95] explains why it is difficult to infer topologies and thus proposes the generation of topologies based on network design parameters. He argues why pure degree-based generators such as BRITE or GT-ITM fail to capture real optimization challenges faced by network designers. Metrics such as latency minimization, dimensioning, and redundancy are discussed. IGen first creates PoPs to look like the Sprint network, then makes connected trees based on the Highly Optimized Tolerance methodology [72].

Table 2 compares the most basic capabilities of the tools discussed in this section and currently available to the community.

The development of the above work suggests that realistic topology generators must take link bandwidth and geographic distribution of nodes into consideration. It is becoming increasingly important for network researchers to take into consideration the evolution and structure of networks and the Internet as a whole over time, and the presence of annotated links plays an important role in this context.

Future Directions and Conclusions

Today the Internet's complex architecture and organizational structure has made it a challenging task for engineers and network researchers to provide a concrete map of the network, and for statisticians to propose extensive mathematical models. At a lower level, defining the physical interconnection of the nodes is essential for routing and resilience purposes. At the higher layers, the virtual types of connectivity structures are very different when studied from different sources of data, and a correct understanding of the nature of these connections is essential for traffic engineering and economic modeling of the network.

In the inference research field the focus is on trying to get a map of the Internet at the router or AS level. Researchers try to understand routing policies and provide connectivity maps by focusing on the router- and AS-level graphs.

In the modeling area the success of a model of Internet topology improves by annotating the nodes and edges of the router and AS graphs with information that will bring the models closer to the reality of the network.

In the topology generation literature, the important factor present is the focus on use of distribution-derived methods, which rules out randomly generated graphs and puts the focus on attaching *meta-data* (metrics) on any link and router generated in a graph.

Future research in topology generation needs an extensive comparative study to compare different topology generators. The main issue is lack of a model that is realistic, without trying to simulate the whole of the Internet or being an abstract mathematical model with just AS-level details. Such a task calls for a generation model that achieves a good balance between keeping the structural and degree-related properties intact. Such a tool is yet to be developed. Another unresolved issue is meta-data inference, such as bandwidth and delay, and more important, the ability to associate them with the inferred topologies.

An area of importance for future work is tools that produce a complete view of a given network, form the annotated link and router level to the PoP level and eventually to the AS level. An example of such a situation would be enterprise networks spanning across the globe, with nodes and routers being part of multiple AS sets. Focusing on enterprise or corporate networks is a challenging task as there is little data available for benchmarking due to the unwillingness to share such information. There has been no work to model discrete events such as on-demand circuit setups, VLAN setups, multiproto-

col label switching and virtual private networks, and content distribution on-demand operators, which will most probably be of interest in a corporate environment. Such research is becoming increasingly important for enterprise network operators who try to achieve security and resilience by segmentation of networks into various operational domains using VLANs, private AS numbers, global routers, and firewalls.

Another important step toward realistic modeling of the Internet and similar networks is characterizing the evolution of the topology of such networks, and their effect on the performance of various protocols and traffic models. Such analysis is only possible if a topology generator provides scenarios of failure in links and nodes that are similar to those in the Internet. An obstacle to such research has been the lack of a detailed study of failure models in large area networks and the AS-level Internet. Oliveira *et al.* [96] look at the evolution of the Internet at the AS level as observed from some data collection points, and based on the model, they propose a model to distinguish real topology changes from transient routing changes with a given confidence level. Availability of such models will help future topology generators take into consideration such changes in the topology.

Acknowledgments

The authors would like to thank Richard Clegg, Christos Gkantsidis, and Timothy Griffin for their constructive comments and feeback. Eng Keong Lua and Jon Crowcroft provided insight on landmarking techniques. Christian Schwarzer and Mathew George provided the code for random and power law networks. We also appreciate the detailed comments of the anonymous reviewers. This survey is conducted as part of the EPSRC UKLIGHT/MASTS 22 project (Grants GR/T10503/01 and GR/T10510/03).

References

[1] P. Mahadevan *et al.*, "Lessons from Three Views of the Internet Topology," CAIDA, tech. rep., 2005.
[2] S. Zhou and R. Mondragon, "Accurately Modeling the Internet Topology," *Physical Rev. E*, vol. 70, no. 066108, 2004.
[3] M. Baur *et al.*, "Drawing the AS Graph in 2.5 Dimensions," *Graph Drawing*, J. Pach, Ed., Springer, 2004, pp. 43–48.
[4] Y. Rekhter and T. Li, "A Border Gateway Protocol 4 (BGP-4)," IETF RFC 1771, Mar. 1995.
[5] G. Malkin, "Traceroute Using an IP Option," IETF RFC 1393, Jan. 1993.
[6] S. Floyd and V. Paxson, "Difficulties in Simulating the Internet," *IEEE/ACM Trans. Net.*, vol. 9, no. 4, 2001, pp. 392–403.
[7] A. Medina *et al.*, "BRITE: An Approach to Universal Topology Generation," *9th Int'l. Symp. Modeling, Analysis and Simulation of Comp. and Telecommun. Sys.*, 2001, pp. 346–53.
[8] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP Topologies with Rocketfuel," *Proc. ACM SIGCOMM 2002*, ACM Press, pp. 133–45.
[9] B. Donnet and T. Friedman, "Internet Topology Discovery: A Survey," *IEEE Commun. Surveys and Tutorials*, 2007.
[10] J. Postel, "Internet Control Message Protocol," IETF RFC 0792, Sept. 1981.
[11] R. Govindan and H. Tangmunarunkit, "Heuristics for Internet Map Discovery," *Proc. IEEE INFOCOM 2000*, Tel Aviv, Israel, 2000, pp. 1371–80.
[12] R. Braden, "Requirements for Internet Hosts — Communication Layers," IETF RFC 1122, Oct. 1989.
[13] V. Fuller and T. Li, "Classless Inter-Domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan," IETF RFC 4632, Aug. 2006.
[14] B. Huffaker *et al.*, "Topology Discovery by Active Probing," *Proc. 2002 Symp. Apps. and the Internet Wksps.*, IEEE Comp. Soc., p. 90.
[15] N. Spring *et al.*, "How to Resolve IP Aliases," Tech. rep., 2004; http://citeseer.ist.psu.edu/article/spring04how.html
[16] D. Magoni and M. Hoerdt, "Internet Core Topology Mapping and Analysis," *Comp. Commun.*, vol. 28, no. 5, 2005, pp. 494–506.
[17] Y. Shavitt and E. Shir, "DIMES: Let the Internet Measure Itself," *SIGCOMM Comp. Commun. Rev.*, vol. 35, no. 5, 2005, pp. 71–74.
[18] D. Achlioptas *et al.*, "On the Bias of Traceroute Sampling: or, Power-Law Degree Distributions in Regular Graphs," *Proc. 37th Annual ACM Symp. Theory of Comp.*, ACM Press, 2005, http://masts.uklight.ac.uk, pp. 694–703.
[19] A. Clauset and C. Moore, "Accuracy and Scaling Phenomena in Internet Mapping," *Physical Rev. Lett.*, 94:018701, 2005.
[20] R. Teixeira *et al.*, "In Search of Path Diversity in ISP Networks," *Proc. 3rd ACM SIGCOMM Conf. Internet Measurement*, 003, pp. 313–18.
[21] A. Lakhina *et al.*, "Sampling Biases in IP Topology Measurements," *Proc. IEEE INFOCOM '03*, San Francisco, CA.
[22] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *Proc. ACM SIGCOMM '99*, pp. 251–62.
[23] J.-J. Pansiot and D. Grad, "On Routes and Multicast Trees in the Internet," *SIGCOMM Comp. Commun. Rev.*, vol. 28, no. 1, 1998, pp. 41–50.
[24] B. Donnet *et al.*, "Deployment of an Algorithm for Large-Scale Topology Discovery," *IEEE JSAC*, vol. 24, no. 12, 2006, pp. 2210–20.
[25] B. Donnet and T. Friedman, "Topology Discovery Using an Address Prefix Based Stopping Rule," *IFIP*, vol. 196, 2006, pp. 119–30.
[26] B. H. Bloom, "Space/Time Trade-Offs in Hash Coding with Allowable Errors," *Commun. ACM*, vol. 13, no. 7, 1970, pp. 422–26.
[27] B. Augustin *et al.*, "Avoiding Traceroute Anomalies with Paris Traceroute," *Proc. ACM/Usenix Internet Measurement Conf. '06*, pp. 153–58.
[28] L. Dall'Asta *et al.*, "Exploring Networks with Traceroute-Like Probes: Theory and Simulations," *Theoretical Comp. Sci.*, vol. 355, no. 1, 2006, pp. 6–24.
[29] C. Labovitz, G. R. Malan, and F. Jahanian, "Internet Routing Instability," *Proc. ACM SIGCOMM 1997*, pp. 115–26.
[30] P. Traina, D. McPherson, and J. Scudder, "Autonomous System Confederations for BGP," IETF RFC 3065, Feb. 2001.
[31] T. Bates and R. Chandra, "BGP Route Reflection An Alternative to Full Mesh IBGP," IETF RFC 1966, June 1996.
[32] D. Oran, "OSI IS-IS Intra-Domain Routing Protocol," IETF RFC 1142, Feb. 1990.
[33] C. Hedrick, "Routing Information Protocol," IETF RFC 1058, June 1988.
[34] J. Moy, "OSPF Version 2," IETF RFC 2328, Apr. 1998.
[35] Z. M. Mao *et al.*, "Towards an Accurate AS-level Traceroute Tool," *Proc. ACM SIGCOMM '03*, pp. 365–78.
[36] T. G. Griffin and G. Wilfong, "On the Correctness of iBGP Configuration," *SIGCOMM Comp. Commun. Rev.*, vol. 32, no. 4, 2002, pp. 17–29.
[37] L. Gao, "On Inferring Autonomous System Relationships in the Internet," *IEEE/ACM Trans. Net.*, vol. 9, no. 6, 2001, pp. 733–45.
[38] C. Alaettinoglu *et al.*, "Routing Policy Specification Language (RPSL)," IETF RFC 2622, June 1999.
[39] L. Subramanian *et al.*, "Characterizing the Internet Hierarchy from Multiple Vantage Points," *Proc. IEEE INFOCOM '02*, June 2002.
[40] G. D. Battista *et al.*, "Computing the Types of the Relationships Between Autonomous Systems," *IEEE/ACM Trans. Net.*, vol. 15, no. 2, 2007, pp. 267–80.
[41] Z. M. Mao *et al.*, "On AS-Level Path Inference," *SIGMETRICS '05: Proc. 2005 ACM SIGMETRICS Int'l. Conf. Measurement and Modeling of Comp. Sys.*, 2005, pp. 339–49.
[42] X. Dimitropoulos *et al.*, "AS Relationships: Inference and Validation," *SIGCOMM Comp. Commun. Rev.*, vol. 37, no. 1, 2007, pp. 29–40.
[43] N. G. Duffield and F. L. Presti, "Network Tomography from Measured End-to-End Delay Covariance," *IEEE/ACM Trans. Net.*, vol. 12, no. 6, 2004, pp. 978–92.
[44] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, 1997.

[45] W. Muhlbauer *et al.*, "Building an AS-Topology Model That Captures Route Diversity," *SIGCOMM Comp. Commun. Rev.*, vol. 36, no. 4, 2006, pp. 195–206.
[46] H. Chang, S. Jamin, and Willinger, "To Peer or Not to Peer: Modeling the Evolution of the Internet's AS-Level Topology," Barcelona, Spain.
[47] P. Poyhonen, "A Tentative Model for the Volume of Trade Between Countries," *Weltwirtschaftliches Archive*, vol. 90, 1963, pp. 93–100.
[48] W. Muhlbauer *et al.*, "In Search for an Appropriate Granularity to Model Routing Policy," *Proc. ACM SIGCOMM '07*, Kyoto, Japan, Aug. 2007.
[49] N. Feamster and H. Balakrishnan, "Detecting BGP Configuration Faults with Static Analysis," *Proc. 2nd Conf. Symp. Networked Sys. Design & Implementation*, Berkeley, CA, 2005, pp. 43–56.
[50] Z. Mao *et al.*, "Inferring AS-level paths with RouteScope," AT&T Labs-Research, tech. rep. TD-5T3RRP, Nov. 2003.
[51] P. Francis *et al.*, "Idmaps: A Global Internet Host Distance Estimation Service," *IEEE/ACM Trans. Net.*, vol. 9, no. 5, 2001, pp. 525–40.
[52] B. M. Waxman, "Routing of Multipoint Connections," *Broadband Switching: Architectures, Protocols, Design, and Analysis*, IEEE Comp. Soc. Press, 1991, pp. 347–52.
[53] M. B. Doar, "A Better Model for Generating Test Networks," *IEEE GLOBECOM '96*, 1996.
[54] J. Winick and S. Jamin, "Inet-3.0: Internet Topology Generator," Univ. of MI tech. rep. CSE-TR-456-02, 2002.
[55] T. Ng and H. Zhang, "Predicting Internet Network Distance With Coordinates-Based Approaches," *Proc. IEEE INFOCOM '02*, New York, NY.
[56] L. Tang and M. Crovella, "Geometric Exploration of the Landmark Selection Problem," LNCS 3015, *Proc. Passive and Active Measurement Wksp.*, 2004, pp. 63–72.
[57] E. K. Lua *et al.*, "On the accuracy of Embeddings for Internet Coordinate Systems," *Proc. Internet Measurement Conf.*, Berkeley, CA, 2005, p. 11.
[58] E. W. Dijkstra, "A Note on Two Problems in Connexion With Graphs," *Numerische Mathematik. Mathematisch Centrum*, Amsterdam, The Netherlands, 1959, vol. 1, pp. 269–71.
[59] A. Lakhina *et al.*, "On the Geographic Location of Internet Resources," *IEEE JSAC*, vol. 21, no. 6, 2003, pp. 934–48.
[60] H. Zheng *et al.*, "Internet Routing Policies and Round-Trip-Times," Passive and Active Measurement, 2005.
[61] P. Erdos and A. Renyi, "On Random Graphs," *Mathematical Inst. Hungarian Academy*, no. 196, London: Academic Press, 1985.
[62] G. Siganos *et al.*, "Power Laws and the AS-level Internet Topology," *IEEE/ACM Trans. Net.*, vol. 11, no. 4, 2003, pp. 514–24.
[63] Q. Chen *et al.*, "The Origin of Power Laws in Internet Topologies Revisited," *Proc. IEEE INFOCOM '02,* New York, NY.
[64] A. L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, 1999, pp. 509–12.
[65] A. Medina, I. Matta, and J. Byers, "On the Origin of Power Laws in Internet Topologies," *SIGCOMM Comp. Commun. Rev.*, vol. 30, no. 2, 2000, pp. 18–28.
[66] H. Chang *et al.*, "Towards Capturing Representative AS-level Internet Topologies," *Comp. Networks*, vol. 44, no. 6, 2004, pp. 737–55.
[67] W. Aiello, F. Chung, and L. Lu, "A Random Graph Model for Massive Graphs," *Proc. 32nd Annual ACM Symp. Theory of Comp.*, New York, NY, 2000, pp. 171–80.
[68] D. Alderson *et al.*, "Understanding Internet Topology: Principles, Models, and Validation," *IEEE/ACM Trans. Net.*, vol. 13, no. 6, 2005, pp. 1205–18.
[69] J. M. Carlson and J. Doyle, "Highly Optimized Tolerance: Robustness and Design in Complex Systems," *Physical Rev. Lett.*, vol. 84, no. 11, 2000, pp. 2529–32.
[70] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, "Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet," *Proc. 29th Int'l. Colloq. Automata, Languages and Programming*, 2002, pp. 110–22.
[71] L. Li *et al.*, "A First-Principles Approach To Understanding the Internet's Router-Level Topology," *Proc. ACM SIGCOMM '04*, pp. 3–14.
[72] D. Alderson *et al.*, "Toward an Optimization-Driven Framework for Designing and Generating Realistic Internet Topologies," *SIGCOMM Comp. Commun. Rev.*, vol. 33, no. 1, 2003, pp. 41–46.
[73] D. Magoni and J. J. Pansiot, "Analysis of the Autonomous System Network Topology," *SIGCOMM Comp. Commun. Rev.*, vol. 31, no. 3, 2001, pp. 26–37.
[74] M. Mihail, C. Gkantsidis, and E. Zegura, "Spectral Analysis of Internet Topologies," *Proc. IEEE INFOCOM '03*, San Francisco, CA.
[75] S. Jaiswal, A. L. Rosenberg, and D. Towsley, "Comparing the Structure of Power-Law Graphs and the Internet AS Graph," *Proc. 12th IEEE Int'l. Conf. Network Protocols*, IEEE Comp. Soc., 2004, pp. 294-303.
[76] S. Carmi *et al.*, "MEDUSA -New Model of Internet Topology Using k-shell Decomposition," *ArXiv Condensed Matter e-prints*, 2006.
[77] J. I. Alvarez-Hamelin *et al.*, "k-Core Decomposition: A Tool for the Visualization of Large Scale Networks," *Advances in Neural Info. Processing Sys. 18*, Canada, 2006, p. 41.
[78] H. Chang *et al.*, "The Many Facets of Internet Topology and Traffic," vol. 1, no. 4, *Amer. Inst. for Mathematical Sci.*, 2006, pp. 569–600.
[79] G. Sagie and A. Wool, "A Clustering Approach for Exploring the Internet Structure," *23rd IEEE Convention of Elec. and Elect. Engs. in Israel*, 2004, pp. 149–52.
[80] S.-H. Yook, H. Jeong, and A.-L. Barabasi, "Modeling the Internet's Large-Scale Topology," Applied Physical Sci., vol. 99, no. 21, 2002, Pp. 13, 382–86.
[81] S. Zhou and R. J. Mondragon, "Structural Constraints in Complex Networks," *New J. Physics*, vol. 9, no. 6, June 2007, pp. 173+.
[82] S. Jin and A. Bestavros, "Small-World Characteristics of Internet Topologies and Implications on Multicast Scaling," *Comp. Networks*, vol. 50, no. 5, 2006, pp. 648–66.
[83] J. Spencer *et al.*, "Emergent Properties of the BT SDH Network," *BT Technology Journal*, vol. 21, no. 2, 2003, pp. 28–36.
[84] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge Univ. Press, 2004.
[85] H. Tangmunarunkit *et al.*, "Network Topology Generators: Degree-Based vs. Structural," *Proc. ACM SIGCOMM '02*, New York, NY, pp. 147–59.
[86] H. Chang, S. Jamin, and W. Willinger, "Internet Connectivity at the AS-level: An Optimization-Driven Modeling Approach," *Proc. ACM SIGCOMM Wksp. Models, Methods and Tools for Reproducible Network Research*, 2003, pp. 33–46.
[87] T. Bu and D. Towsley, "On Distinguishing between Internet Power Law Topology Generators," *Proc. IEEE INFOCOM '02*, New York, NY.
[88] P. Mahadevan *et al.*, "Systematic Topology Analysis and Generation Using Degree Correlations," *Proc. ACM SIGCOMM '06*, 2006, pp. 135–46.
[89] P. Mahadevan *et al.*, "Orbis: Rescaling Degree Correlations to Generate Annotated Internet Topologies," *SIGCOMM Comp. Commun. Rev.*, vol. 37, no. 4, 2007, pp. 325–36.
[90] A. Nucci, A. Sridharan, and N. Taft, "The Problem of Synthetically Generating IP Traffic Matrices: Initial Recommendations," *SIGCOMM Comp. Commun. Rev.*, vol. 35, no. 3, 2005, pp. 19–32.
[91] Krishnamurthy *et al.*, "Reducing Large Internet Topologies for Faster Simulations," *NETWORKING '05*, pp. 328–41.
[92] H. Mizuta and F. Nakamura, "Agent-Based Simulation of Enterprise Communication Networks," *37th Conf. Winter Simulation*, Orlando, FL, 2005.
[93] N. Spring, R. Mahajan, and T. Anderson, "The Causes of Path Inflation," *Proc. ACM SIGCOMM '03*, 2003, pp. 113–24.
[94] K. L. Calvert, M. B. Doar, and E. W. Zegura, "Modeling Internet Topology," *IEEE Commun. Mag.*, vol. 35, no. 6, 1997, pp. 160–63.
[95] B. Quoitin, "Topology Generation Based on Network Design Heuristics," *Proc. 2005 ACM Conf. Emerging Network Experiment and Tech.*, pp. 278–79.
[96] R. Oliveira, B. Zhang, and L. Zhang, "Observing the Evolution of Internet AS Topology," *Proc. ACM SIGCOMM '07*, Kyoto, Japan, Aug.

BIOGRAPHIES

HAMED HADDADI (hamed@ee.ucl.ac.uk) received B.Eng. (Hons.) and M.Sc. degrees from University College London, United Kingdom. He is currently working toward a Ph.D. degree at University College London. He spent his internship working at Intel Research, Cambridge, United Kingdom. His research interests include network measurement and monitoring, distributed sampling techniques, network topology characterization, and fault finding.

GIANLUCA IANNACONE (gianluca.iannaccone@intel.com) received his B.S. and M.S. degrees in computer engineering from the University of Pisa, Italy, in 1998. He received a Ph.D. degree in computer engineering from the University of Pisa in 2002. He joined Sprint as a research scientist in October 2001 working on network performance measurements, loss inference methods, and survivability of IP networks. In September 2003 he joined Intel Research, Cambridge, and recently moved to Intel Research, Berkeley, California. His current interests include network measurements, router architecture design, and network security and troubleshooting.

ANDREW MOORE is a lecturer at the University of Cambridge Computer Laboratory. Previously as an EPSRC Academic Fellow he founded the Network Research Group at Queen Mary University of London, and he was an Intel Research Fellow with the University of Cambridge Computer Laboratory. His research theme is to address the scalability, usability, and reliability of the Internet. He obtained a Ph.D. at the Computer Laboratory of the University of Cambridge.

RICHARD MORTIER is a researcher at Microsoft Research Cambridge in networked systems, focusing particularly on their measurement, management, and resource control. Most recently he has worked on novel approaches to network management. Prior to working for Microsoft he worked for Sprint at their Advanced Technology Laboratory in California, collecting and analyzing routing data from their IP backbone (Sprintlink). He completed his Ph.D. with the University of Cambridge Computer Laboratory in 2001 on approaches for using pricing in Internet traffic engineering. He also has a B.A. in mathematics from the University of Cambridge.

MIGUEL RIO is a lecturer at the Department of Electronic and Electrical Engineering of University College London, specialized in Internet technology. He completed his Ph.D. at the Computer Laboratory of the University of Kent at Canterbury. His research interests include Internet congestion control, multimedia distribution using peer-to-peer networks, QoS routing, and traffic measurement and analysis.