An Analysis of Internet Inter-Domain Topology and Route Stability

Ramesh Govindan
USC/Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292-6695
Phone: +1 310–822–1511 (ext. 103)
Fax: +1 310–823–6714
`govindan@isi.edu`

Anoop Reddy
USC/Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292-6695
Phone: +1 310–822–1511 (ext. 742)
Fax: +1 310–823–6714
`areddy@isi.edu`

Abstract

The Internet routing fabric is partitioned into several *domains*. Each domain represents a region of the fabric administered by a single commercial entity. Over the past two years, the routing fabric has experienced significant growth. From more than a year's worth of inter-domain routing traces, we analyze the Internet inter-domain topology, its route stability behavior, and the effect of growth on these characteristics. Such an analysis is important because inter-domain routing provides the foundation for Internet wide-area communication.

Despite growth, the degree distribution and the diameter of the inter-domain topology have remained relatively unchanged. Furthermore, there exists a four-level hierarchy of Internet domains classified by degree. However, connectivity between domains is significantly non-hierarchical. Despite increased connectivity at higher levels in the topology, the distribution of paths to prefixes from the backbone remained relatively unchanged. There is evidence that both route availability and the mean time to unreachability have degraded with Internet growth.

Keywords: Internet, Inter-domain Routing, BGP, Topology, Route Stability.

1 Introduction

In the past two years, we have witnessed the emergence of a commercial Internet infrastructure. Today, the Internet routing fabric is partitioned into different administrative *domains*. A campus or internal corporate network is an example of a domain. Data exchange between campus or corporate domains is facilitated by one or more *provider* domains; these domains offer, as a service, transmission and switching facilities for data exchange between their customers. Providers usually interconnect at Internet *exchange points*, and can vary in the geographical scope of their operations from regional, to national and international.

Domains may need to exercize traffic access restrictions and express transit preferences. Currently, domains realize these *policies* by selectively disseminating routing information. The Border Gateway Protocol (BGP) version 4 [16], an *inter-domain* hop-by-hop routing protocol, is used for this purpose. In BGP, a domain may originate one or more *routes*. Routes advertise reachability to IP *address prefixes* within a domain. A domain realizes its traffic policies by independently selecting, and selectively propagating routes heard from its neighbors or *peers*. Associated with each route is a list of identifiers for domains traversed by the route. This list is the route's `AS_PATH`.

The collection of domains, their policies and peering relationships, and the address prefixes they advertise defines the Internet *inter-domain routing system*. The inter-domain routing system is large. In December 1995, there were more than 900 domains, and nearly 31,000 address prefixes. Paralleling the growth of the Internet, the routing system has also rapidly increased in size. Between mid 1994 and late 1995, domains and address prefixes have each nearly doubled.

The inter-domain routing system provides the foundation for Internet wide-area communication. Two characteristics of the routing system can impact the performance of such communication:

Inter-domain topology: This is the graph of domains and the inter-domain peering relationships. A link in this graph signifies route exchange—and presumably IP traffic exchange—between the corresponding domains.

Route stability: The routing system experiences transient changes in routes, caused by router and link failures or router misconfiguration.

With the growth of the routing system, it is reasonable to expect significant changes in these characteristics.

With the emergence of the World-Wide Web and Mbone-based applications, wide-area communication is becoming more prevalent [14]. Understanding the topology and route stability behavior of the routing system is important for end-to-end protocol design and evaluation. However, the geographic extent of the routing system precludes an accurate analysis of these characteristics.

From traces of routing updates seen at a large Internet service provider and at a popular Internet exchange point, this paper *approximately* characterizes the routing system and its growth. This approximation is possible because most large service providers carry routes to almost all IP-connected destinations and BGP's `AS_PATH`s contain some inter-domain topology information.

Our analysis reveals several interesting results about the inter-domain topology. Despite the growth of the routing system, the diameter and degree distribution in the inter-domain topology have remained relatively constant. The path "redundancy"—the number of distinct paths to a prefix—has also not changed significantly, despite increased connectivity. A closer analysis of the degree distribution reveals the existence of a four-level hierarchy of domains. Connectivity between domains, however, appears to be significantly non-hierarchical.

While routes to prefixes were highly available (more than 90% of the address prefixes were available for more than 95% of the time), there was some evidence of degradation of availability with growth. Not surprisingly, more than 80% of the prefixes were reachable for more than 95% of the time by a single *primary* path. The mean reachability duration for a prefix shows wide variation, but a noticeable degradation with growth. This reachability duration, however, appears to be correlated with the length of the primary path.

Section 2 motivates the need for a study of routing system growth. Section 3 describes some simple measures of growth of the inter-domain routing system. Section 4 studies the inter-domain topology, and the effect of growth on the topology. Section 5 analyzes route stability and its variation with growth. Finally, Section 6 describes related work and Section 7 presents our conclusions.

2 Background and Motivation

In this section, we briefly describe Internet inter-domain routing. We also motivate the need for an analysis of the inter-domain routing system.

The Internet Inter-Domain Topology

A domain is an autonomously administered portion of the Internet's routing fabric. Domains vary in size, geographic extent, and function. *Backbones* provide national or inter-continental transit, *regionals* serve a metropolitan area or a collection of such areas, and *stubs* provide campus level IP connectivity. Domains interconnect in several ways. The larger providers usually exchange traffic at neutral *exchange points*, such as MAE-East in Washington DC and the NSF-sponsored NAPs [10]. Stubs usually connect to their providers using direct leased lines.

By abstracting out the internal topology of domains, we can describe the Internet as an undirected graph whose nodes are the domains and whose links are the inter-domain interconnects. In the early Internet, this *inter-domain topology* was approximately tree-structured. A single national *backbone* provided cross-country transit to several regional providers. Campus networks obtained global IP connectivity by attaching to their respective regionals. With the advent of commercial IP service, the structure of this graph has become more complex. Several backbones now interconnect in a partial mesh. A regional provider may interconnect with other regionals serving the same or adjacent geographical area, as well as with one or more backbones. Finally, a stub may connect to one or more regionals or backbones.

The Border Gateway Protocol

The Border Gateway Protocol (BGP) is used to achieve inter-domain routing information exchange. The unit of information exchanged by BGP-speakers is an *update*. An update contains a collection of address prefixes. Address prefix X represents the topological region of the Internet in which hosts and routers' network-layer addresses have X as a prefix; a route from a peer containing prefix X advertises that peer's reachability to some entity within that topological region. Address prefixes are usually represented by an IP address and a prefix length.

For each address prefix it contains, an update defines a *routing transition*. We distinguish between two kinds of transitions: *unreachables*, and routes. An unreachable for prefix X advertises the sender's inability to reach destinations in X. Conversely, a route for prefix X advertises the sender's reachability to destinations in X. Associated with each route is an `AS_PATH`. The `AS_PATH` for prefix X lists the domain-level path to destinations in X.

In today's Internet, each domain originates routes to all address prefixes within its borders. Usually, the domain's border routers learn of these address prefixes through a separate *intra-domain* routing protocol. BGP allows domains to realize access restrictions and transit preferences—domain traffic *policy*—by selectively propagating routing information. A domain that receives two or more routes to an address prefix can independently select one of the routes, thereby exercising its transit preference. A domain can also independently restrict the advertisement of a selected route to a neighbor, thereby realizing its access restrictions.

Characterizing the Inter-Domain Routing System

The inter-domain routing system determines the perceived quality of wide-area communication in the Internet. Such wide-area communication is becoming increasingly prevalent; studies at one campus indicate that wide-area TCP connections originating at that campus domain grew exponentially [14]. Several factors, including the rapid growth of the World-Wide Web, and the increasing use of the Mbone, account for this increase. Another measure of the growth in wide-area communication is the growth in aggregated backbone traffic; the utilization of the MAE-East Gigaswitch nearly doubled in the five months between May and September 1995 [8].

To understand the impact of the routing system on wide-area communication, we focus on two characteristics of the routing system: the inter-domain topology and route stability. What factors influence inter-domain topology and route stability? In today's Internet, the inter-domain topology is determined by bilateral transit agreements between providers. The presence of several exchange points enables increased domain connectivity. Router failure or temporary router overload is one reason for route instability. Loss of link connectivity either within a domain or between domain border routers is another. Routing misconfiguration is a third—currently, most inter-domain route exchange is effected by manual policy configuration.

When analyzing topology and route stability, it is important to understand how these characteristics change with *growth*. A unique feature of the Internet is the relatively rapid growth in the topology and the number of routed address prefixes, caused by the rapid increase in the number of connected hosts [13] and the deployment of a commercial Internet infrastructure. In June 1994, the routing system consisted of about 400 domains and 18,000 prefixes. By November 1995, it had grown to 900 domains and 31,000 prefixes.

Why might an analysis of routing system topology and route stability be useful? First, such an analysis enables a better understanding of how these characteristics impact end-to-end communication performance. Greater domain connectivity, caused by growth, can increase the redundancy of end-to-end connectivity. However, a larger Internet might increase the operational range of delays and throughput encountered by distributed applications. Changes in prefix reachability can affect the delay, loss, packet re-ordering and throughput characteristics observed by long-lived connections. Such variations can affect protocol behavior and application architectures. Path delay variation can affect TCP round-trip estimates, and adversely impact adaptive audio and video applications. Increased packet losses or packet re-ordering may reduce the efficiency of the network as a whole. Further, route changes can also reduce the efficacy of "flow" caching schemes.

Second, such an analysis can impact the design of routing protocols and route distribution mechanisms, as well as the implementation and provisioning of routers. For example, frequent changes in prefix reachability can increase route computation overhead on routers, and have anecdotally been known reduce the efficacy of forwarding table caching schemes. Growth affects the amount of information routing protocols need to propagate and the storage and computational requirements of routers. Greater domain-level connectivity, and the availability of alternate paths, may argue for the deployment of explicit routing mechanisms [9].

Finally, a better understanding of the routing system can improve modeling of internetwork topologies [17], and of end-to-end delay and loss characteristics [5].

Accurately analyzing the routing system and its growth is extremely difficult. Because the Internet is geographically distributed, no *complete* map of the inter-domain topology exists today. Efforts are underway to establish Internet policy registries containing such information [2, 4]. Because the Internet is administratively decentralized, it is difficult to obtain a composite picture of the inter-domain topology and route stability.

In this paper, we *approximately* characterize routing system topology and route stability, and the impact of growth on the routing system, using two chronological traces of routing transitions:

- $\mathcal{D}_A$ contains nearly 6.6 million BGP updates heard from a backbone router of a major IP service provider between June 7 1994 and June 26 1995[1].
- $\mathcal{D}_M$ contains nearly 5.1 million BGP updates heard from more than twenty large and small IP service providers attached to a large Internet exchange point (MAE-East in Washington DC) between August 1995 to November 1995. $\mathcal{D}_M$ was collected at an experimental facility, the MAE-East Route Server.

These traces are incomplete in parts, caused by failures at the trace collection points. In our analysis, we have been careful to eliminate the effect of these trace flaws.

To simplify our analysis of topology and route stability, we focused on three different *snapshots* (Section 4) of these traces, separated by nearly six months. Each snapshot contains some inter-domain connectivity information, namely the `AS_PATH`s associated with updates. Using this, we derive an approximate characterization of the inter-domain topology. This enables us to quantitatively analyze the growth of several characteristics of this topology: the number and size of domains, the extent of inter-domain connectivity, the number of different domain-level paths to address prefixes, and so on. From the traces, we also obtain distributions for two measures of route stability: availability of a path to a prefix, and the mean time to unreachability for a prefix.

3 Routing System Growth

Before understanding the effect of growth on topology and route stability, it is important to quantify this growth. A simple measure of the growth of the routing system is the arrival pattern of new address prefixes, domains and links.

---

[1] Advanced Networks and Services, who operated the now defunct NSFNET backbone service

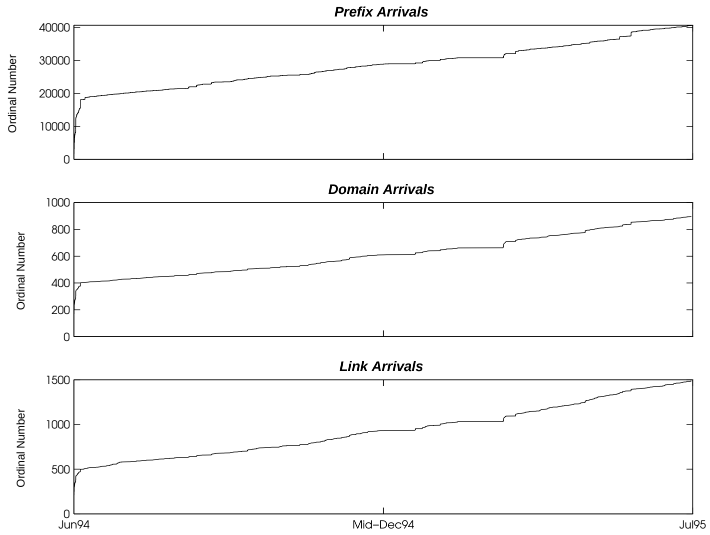


Figure 1: *Arrivals:* In $\mathcal{D}_A$, the arrivals of prefixes, domains, and links is approximately linear. In each graph, the horizontal segments correspond to gaps in $\mathcal{D}_A$.

We define the arrival time of a new address prefix as the first instant at which an update containing that prefix is seen in our traces. Figure 1 plots successive new prefix arrivals in $\mathcal{D}_A$. Over the duration of $\mathcal{D}_A$, the arrivals of new prefixes is approximately linear (the initial step in the figure corresponds to prefixes present in the routing system prior to June 1994, and the horizontal segments represent gaps in $\mathcal{D}_A$). The slope of the curve corresponds to one new prefix *every 25 minutes.* Considering that a prefix usually represents an entity the size of a campus departmental network, this represents a significant rate of address prefix growth. The linearity of the growth curve offers some evidence of the ability of CIDR [11] to curtail the growth of the backbone routing table in the face of exponential Internet growth. The arrival patterns in $\mathcal{D}_M$ are similar.

There is some evidence in $\mathcal{D}_A$ of two or more new prefixes arriving together. We see nearly 1700 such prefix *arrival clusters.* The median cluster size is 3, and nearly 90% of the clusters contain less than 18 prefixes. This corresponds to domains “turning on” connectivity to small numbers of their customers at a time. Some clusters of hundreds of prefixes are also visible, but these may be an artifact of the gaps in $\mathcal{D}_A$.

Even though the traces in $\mathcal{D}_A$ were collected at a single backbone router, we expect the prefix arrival patterns to be representative of the Internet as a whole. This is because backbone routers today contain reachability information to a significant fraction of the Internet; reachability to a new prefix is likely to be heard at a backbone router.

We define the arrival time of a new domain as the first instant at which that domain's identifier was seen in an update's `AS_PATH`. Figure 1 also shows the arrival pattern of new domains. Over the duration of $\mathcal{D}_A$, the arrival of new domains is also approximately linear. The slope of the arrival curve corresponds to one new domain *every 12 hours*. Considering that a domain corresponds to an IP service provider, or a campus network, this statistic offers more evidence of the significant recent growth of the Internet. Domain arrivals in $\mathcal{D}_M$ show a similar trend. Our traces may underestimate the arrivals of new domains, particularly when the policies at our trace collection location exclude routes from a particular domain. However, we believe that the extent of this underestimation is small.

We define the arrival time of a new inter-domain link as the first instant at which the endpoints of the link were observed as successive elements of an update's `AS_PATH`. Figure 1 shows the arrival pattern of new links. Over the duration of $\mathcal{D}_A$, the arrival of new links is approximately linear. The slope of this curve corresponds to a new inter-domain link every *8 hours*. However, our technique of counting new links can can significantly underestimate their actual arrival pattern. If, for policy reasons, routes with a particular `AS_PATH` were never visible at our trace collection location, our technique would not record the corresponding links. For example, we might miss "backdoor" links between service providers.

4 Topology

The arrival patterns of new domains and inter-domain links do not provide information about the inter-domain topology. In this section, we obtain approximate characterizations, called *snapshots* of the inter-domain topology from three different segments of our traces. Using these snapshots, we study the following properties of the topology, and their variation over time: domain degree distribution, diameter, and connectivity.

Snapshots

Backbone routers (from which $\mathcal{D}_A$ and $\mathcal{D}_M$ were obtained) carry routing information about most of the routed address prefixes in the Internet. It is generally believed that backbone routers for larger providers contain upwards of 95% of the entire routed IP address space. Thus, from $\mathcal{D}_A$ and $\mathcal{D}_M$, we can draw fairly reliable conclusions about macroscopic trends in arrival patterns of domains, inter-domain links, and prefixes.

Of interest to us, however, are more detailed characteristics of the routing system: its topology, and its reliability and availability. Two artifacts of the traces preclude our using the entire traces for analyzing these characteristics. First, the traces contain several gaps, primarily caused by outages at the trace collection

	Domains	Links	Address Prefixes
$\mathcal{S}_a$	531	709	21524
$\mathcal{S}_b$	746	1000	26945
$\mathcal{S}_c$	909	1369	31470

Table 1: *Snapshots:* We obtain an approximate representation of the routing system topology from 21-day segments of our traces. This table shows the number of domains, links, and address prefixes in the topologies obtained from our three snapshots.

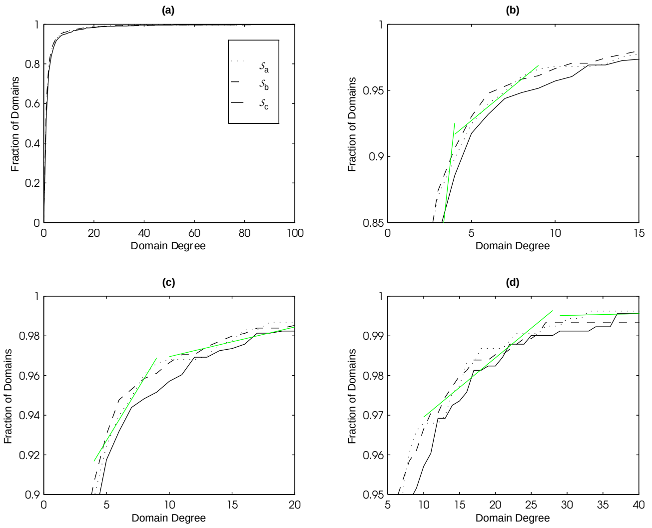
routers. Second, $\mathcal{D}_M$ was collected at an experimental facility (the Mae-East Route Server) and is incomplete in parts.

Nevertheless, there exist segments of these traces that are devoid of these artifacts. From these segments, we can derive an approximate "instantaneous" characterizations (*snapshots*) of the routing system. By comparing widely separated snapshots, we can observe the variation of routing system characteristics with growth. This is the methodology we use in this and the next section. For our analysis, we chose three snapshots: $\mathcal{S}_a$, corresponding to the routing system in November 1994; $\mathcal{S}_b$, corresponding to the routing system in May 1995; and, $\mathcal{S}_c$, corresponding to the routing system in November 1995. Table 1 gives the sizes of the topology and address prefixes obtained from these snapshots.

From the `AS_PATH`s in a snapshot, we can derive the inter-domain topology. Each element of an update's `AS_PATH` defines a node in the topology and each successive pair of domains in the `AS_PATH` represents the end-points of a link. Furthermore, the collection of unique address prefixes seen in the snapshot approximately represents the routed destinations at that instant.

The longer the segment of the trace chosen, the more complete our snapshot is likely to be. However, a longer duration trace segment may represent significant growth and thereby skew our snapshot. We empirically chose 21-day segments of our traces. In each 21-day segment, the number of domains, links, and prefixes grows by no more than 10%. Further, the routing activity in the Internet is such that in each segment, at least two transitions are seen for more than 99% of all prefixes seen in the window. For this reason, we believe that our choice of segment length is likely to capture most of the routed address prefixes in the routing system at the time. We validated this by comparing the address prefix count in our snapshots with an independently collected history of the size of backbone routing tables [6]. In each case, our approximation was within 5% of the size of the routing table at the corresponding instant.

This technique provides a richer view of the topology than an instantaneous dump of the routing table. This is because we may see more than one `AS_PATH` to an address prefix over the 21-day trace segment.



(e)

	Average Degree
$\mathcal{S}_a$	2.67
$\mathcal{S}_b$	2.68
$\mathcal{S}_c$	2.99

Figure 2: *Degree Distribution:* The cumulative fraction of domains plotted against degree. (a) The distributions are nearly identical for the three snapshots. A closer look at the distribution ((b), (c), and (d)) shows that domains can be divided into different classes based on frequency of occurrence. The gray lines in each of (b), (c) and (d) demarcate different classes of domains (see text). The average degree has remained relatively constant.

However, there is still a likelihood of "missing" some of the inter-domain links, and a smaller likelihood of "missing" some domains as well (Section 3). In general, we expect that this technique gives a fairly good picture of the topology closer to the trace collection locations (i.e. in the North American portion of the Internet). The "fuzziness" of our snapshots is likely to increase with with increasing topological distance from the trace collection locations.

We now describe the diameter, degree distribution, and connectivity characteristics of the inter-domain topology.

Degree Distribution

A link in the inter-domain topology represents BGP peering between the end-points of the link. Usually, such peering is based on bilateral data traffic exchange agreements between domains. The number of such peerings, the domain's degree, is therefore an approximate measure of its size. The distribution of domain degree illustrates the range of domain sizes in the Internet.

Class	Degree Range	Approximate Fraction of Domains in Class	Types of domains
$\mathcal{C}_1$	>= 28	0.9%	National or international backbones
$\mathcal{C}_2$	10-27	3.1%	Large North American regional providers and European national networks
$\mathcal{C}_3$	4-9	9%	Smaller regional providers, and large metropolitan area providers
$\mathcal{C}_4$	1-3	87%	Smaller metropolitan area providers and multi-campus corporate or academic networks

Table 2: *Degree Classification:* The degree distribution can be well approximated by a piecewise linear function with four components. This suggests a classification of domains by degree range. Within a degree range, the frequency of occurrence of domains of a certain degree is similar.

Figure 2(e) shows that the average degree of domains in the topology is relatively small. This indicates that the inter-domain topology is sparse. This observation is illustrated better by plotting the cumulative fraction of domains against node degree (Figure 2(a)). Nearly 75% of the domains have a degree of 1 or 2.

Figure 2(e) also shows that the average degree has increased only slightly across the different snapshots, even though the number of domains and links has doubled between $\mathcal{S}_a$ and $\mathcal{S}_c$. This may indicate that the rate of growth has been approximately *uniform* across all sizes of domains. The degree distribution confirms this—the distributions for the three snapshots appear *identical* (Figure 2(a)). This suggests that as the Internet has grown, the fraction of domains having a given degree has remained approximately constant.

The degree distribution is well approximated by a piecewise linear function with four components (Figure 2(b), (c), and (d)). Each component defines a degree range. Within a degree range, the fraction of nodes with a given degree is similar. We believe that these four degree ranges represent a "natural" hierarchy of domains classified by function. Table 2 describes this hierarchy. The constancy of the degree distribution implies that the Internet has been growing "laterally", adding domains to the different levels of the hierarchy in proportion to growth.

Diameter

Table 3 shows the diameter of the inter-domain topology. From this we see that any domain is at most 10 domain-level hops away (in a shortest-path sense) from any other domain. The path taken by packets between two domains is a function of domain policies, and can exceed this figure. This figure is consistent with our domain classification above. Between two stub domains, there are at most 3 or 4 domain-level hops each to $\mathcal{C}_1$ providers. As we show below, connectivity within $\mathcal{C}_1$ is rather high, so the shortest path between

	Diameter
$\mathcal{S}_a$	9
$\mathcal{S}_b$	10
$\mathcal{S}_c$	10

Table 3: *Diameter of Inter-Domain Topology:* Despite the rapid growth of the Internet, the diameter of the inter-domain topology has remained relatively constant.

$\mathcal{C}_1$ providers should be between 2 and 4.

More importantly, despite the near doubling in the number of domains and links between $\mathcal{S}_a$ and $\mathcal{S}_c$, the domain-level *diameter of the Internet has remained relatively constant.* Despite this, however, it is likely that the *router-level* diameter has increased with growth; our traces do not contain the information necessary to confirm this.

Connectivity Between Domain Classes

The degree distribution revealed a hierarchy of domains by size and function. Are inter-domain links also largely hierarchical? That is, is a domain primarily connected to domains in classes immediately above and below its own? In this section, we investigate domain connectivity within and between classes.

Figure 3 depicts the sub-graph of the inter-domain topology containing domains in $\mathcal{C}_1$ and $\mathcal{C}_2$, for the three snapshots. This figure reinforces the uniform growth in the number of domains in $\mathcal{C}_1$ and $\mathcal{C}_2$. In addition, this visual representation of the topology reveals several important features of domain-level connectivity in $\mathcal{C}_1$ and $\mathcal{C}_2$. Between $\mathcal{S}_a$ and $\mathcal{S}_c$, the connectivity within $\mathcal{C}_1$ has increased significantly. A similar observation may be made about the connectivity within $\mathcal{C}_2$. Between $\mathcal{S}_a$ and $\mathcal{S}_c$, the connectivity between $\mathcal{C}_1$ and $\mathcal{C}_2$ has also increased. Each $\mathcal{C}_2$ node appears to peer directly with more $\mathcal{C}_1$ nodes on average. We attribute these trends to the establishment of several tens of Internet exchange points which facilitate inter-provider connectivity. These trends are quantified in the bar graph shown in Figure 3. The fraction of links within $\mathcal{C}_1$ has nearly quadrupled, the fraction of links between $\mathcal{C}_1$ and $\mathcal{C}_2$ has nearly doubled, and the fraction of links within $\mathcal{C}_2$ has increased by nearly a half.

The fraction of links within and between other classes remain approximately constant. However, it is interesting to note that connectivity between domains is significantly non-hierarchical. Table 4 shows the fraction of links between different classes in $\mathcal{S}_c$. Nearly a quarter of the total number of links is between $\mathcal{C}_4$ domains and $\mathcal{C}_1$ domains. There appear to also be some evidence of “backdoor” links between domains in

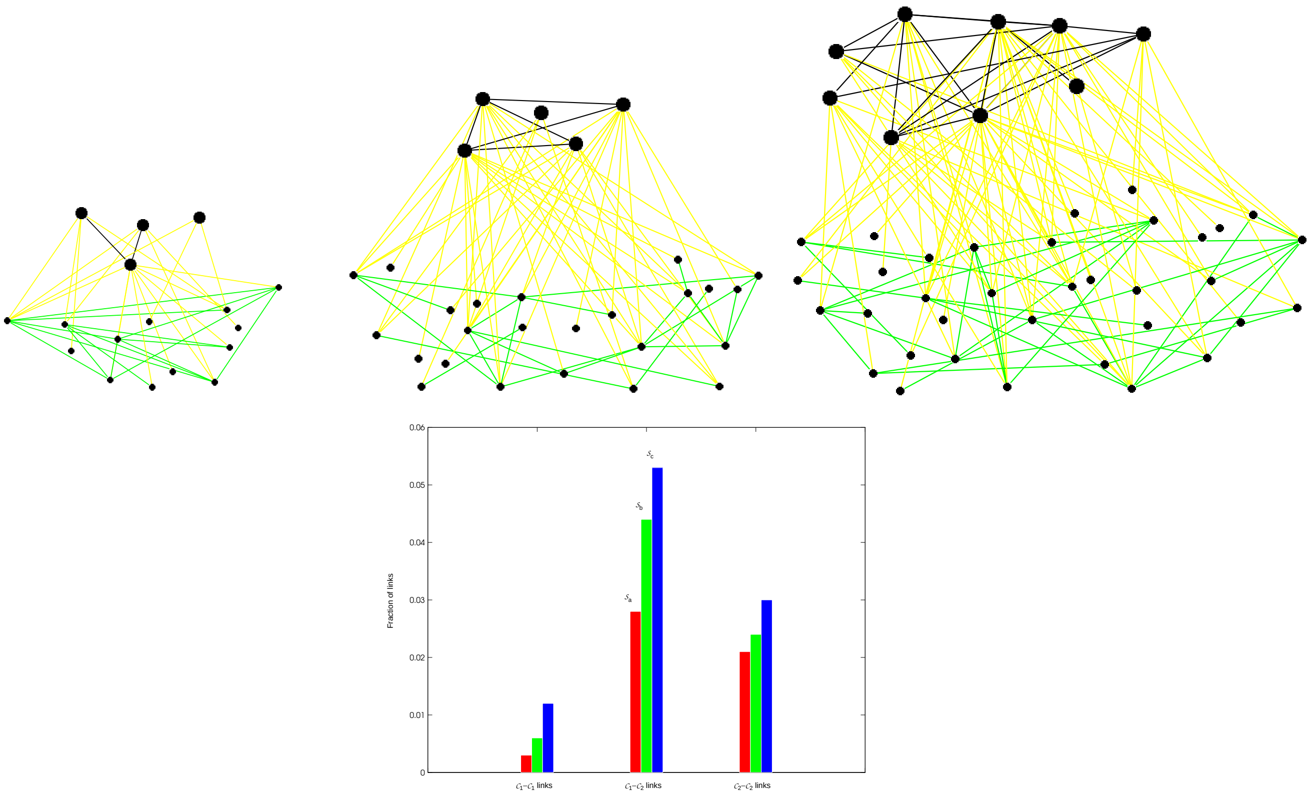


Figure 3: *$\mathcal{C}_1$ and $\mathcal{C}_2$ Topology:* The three graphs describe the subgraph of the inter-domain topology containing $\mathcal{C}_1$ and $\mathcal{C}_2$ nodes, for $\mathcal{S}_a$, $\mathcal{S}_b$, and $\mathcal{S}_c$ respectively, from left to right. The larger circles represent $\mathcal{C}_1$ nodes and the smaller circles $\mathcal{C}_2$ nodes. The solid black lines indicate links between $\mathcal{C}_1$ nodes, the faint lines indicate links between $\mathcal{C}_1$ and $\mathcal{C}_2$ nodes, and the gray lines links between $\mathcal{C}_2$ nodes. The bar graph plots links within $\mathcal{C}_1$, between $\mathcal{C}_1$ and $\mathcal{C}_2$, and within $\mathcal{C}_2$ as a fraction of the total number of links for the three snapshots.

the same class; about 3% of all links are each within $\mathcal{C}_2$ and $\mathcal{C}_3$.

Redundancy in Domain-Level Paths

If the inter-domain connectivity has become increasingly meshed, the number of domain-level paths between any two domains must also have increased. There are several possible measures of this *path redundancy* in general graphs. In the inter-domain topology, however, not all feasible domain-level paths are *useful.* We define a useful path as one whose length differs from the shortest domain-level path by at most 1.

A simple measure of path redundancy in the inter-domain topology is the average number of useful paths from a fixed point in the backbone mesh (Table 5). This average indicates a considerable path redundancy

	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$
$\mathcal{C}_1$	0.012	0.053	0.064	0.250
$\mathcal{C}_2$		0.030	0.059	0.236
$\mathcal{C}_3$			0.034	0.164
$\mathcal{C}_4$				0.098

Table 4: *Connectivity between classes:* An element of this matrix indicates the fraction of the total number of links in $\mathcal{S}_c$ that exist between the corresponding classes. Links are significantly non-hierarchical; noticeable, nearly half of all the links are between domains in $\mathcal{C}_4$, and domains in $\mathcal{C}_1$ and $\mathcal{C}_2$.

in the Internet; in $\mathcal{S}_c$, on average, there were nearly 22 paths to a domain from a fixed backbone provider[2]. Further, this number has more than doubled from $\mathcal{S}_a$ to $\mathcal{S}_c$.

Recall that the route—and implicitly, the path—chosen to a destination is a function of domain policy. For policy reasons, many of these redundant paths may not be available at backbone provider. What is the level of redundancy in the number of paths to destinations? The `AS_PATH` associated with a route update indicates the domain-level path to the address prefixes in the update. We use the number of distinct `AS_PATH`s seen for an address prefix as a measure of the redundancy to destinations within that address prefix.

Figure 4 plots the fraction of address prefixes in each of the snapshots for which n different `AS_PATH`s are seen, for different values of n. To more than half the destinations, we see exactly one domain-level path. To about a fourth, we see exactly two domain-level paths. This is not surprising. Route distribution, and hence, a route's `AS_PATH`, is governed by routing policy. Routing policy is usually set by bilateral transit agreements. In most cases, a domain negotiates a primary and, possibly, one backup transit to a collection of destinations.

Surprisingly, this distribution has *not changed significantly* from $\mathcal{S}_a$ to $\mathcal{S}_c$. This indicates that the routing policies expressed in the Internet have remained relatively simple, despite the growth of the topology. It is conceivable that the lack of tools to configure and analyze inter-domain routing has prevented providers from realizing more complicated policies [1].

Summary and Discussion

In this section, we described several characteristics of the Internet inter-domain routing system and its growth:

---

[2] Advanced Networks and Services

	Useful Paths
$\mathcal{S}_a$	9.1
$\mathcal{S}_b$	13.88
$\mathcal{S}_c$	22.26

Table 5: *Useful Paths:* This figure shows the average number of useful paths (those whose length differs from the shortest path by at most 1) between a fixed point in the Internet's backbone mesh, and other domains.

- The inter-domain topology is relatively sparse. Nearly 75% of the domains have a degree less than 3.
- Despite the rapid growth of the Internet, there appears to be a common natural classification of domains by size. These four classes roughly correspond to national and international backbone providers, large regional providers, smaller regional or large metropolitan area providers, and smaller dial-in providers, or multi-campus corporate networks.
- The connectivity between domains in the upper two layers of the hierarchy has noticeably increased, possibly due to the appearance of several exchange points.
- The inter-domain connectivity does not follow the domain hierarchy. In particular, nearly a fourth of the links in $\mathcal{S}_c$ connect a $\mathcal{C}_4$ domain directly to a backbone provider.
- The diameter of the Internet has remained relatively constant.
- The increased connectivity between providers has not resulted in a greater redundancy of connectivity to destinations.

Our snapshot technique is likely to underestimate the topology. This is particularly true of portions of the topology distant from our observation points on the North American provider mesh. However, it is likely that each snapshot equally underestimates the "instantaneous" topology. We believe, therefore, that these results—obtained by *comparing* snapshots—are reliable estimators for the variation of the inter-domain topology with growth.

5 Route Stability

In this section, we describe inter-domain route stability behavior. Route stability affects reachability to prefixes. From the perspective of routes to a single prefix, two stability measures are of interest: 1) prefix

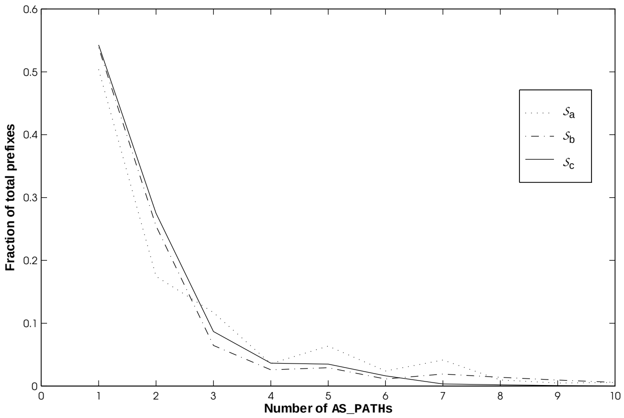


Figure 4: *Path redundancy:* The fraction of all address prefixes in a snapshot for which a given number of paths are seen has not changed significantly over the three snapshots.

availability, the fraction of time that a prefix is reachable, and 2) prefix *steadiness*, the average time that a prefix is continuously reachable. We study the distribution of these measures.

Route stability can vary with the size of the topology. It is important, therefore, to factor out the rapid growth of the Internet when analyzing our traces. For this reason, we use the same trace *snapshot* methodology described in the previous section. We assume that the inter-domain topology has remained static within a snapshot. Recall that a snapshot is long enough to capture at least two transitions for more than 99% of the prefixes seen. On average, however, each snapshot contains more than a hundred transitions per prefix.

To study the effect of growth on route stability, we compare the distributions of these stability measures across snapshots. For this purpose, we consider only snapshots $\mathcal{S}_a$ and $\mathcal{S}_b$, and not $\mathcal{S}_c$. $\mathcal{S}_c$ is extracted from $\mathcal{D}_M$, and this trace was obtained at an experimental facility, the MAE-East Route Server. In these traces, we observe frequent downtimes for peering sessions with clients of the Route Server. These trace flaws prevent a reliable estimation of route stability behavior from $\mathcal{D}_M$ in general. On average, however, route server clients remained highly available, and we have no reason to suspect the collection of `AS_PATH`s seen in $\mathcal{S}_c$, or our conclusions about the growth of topology (Section 4).

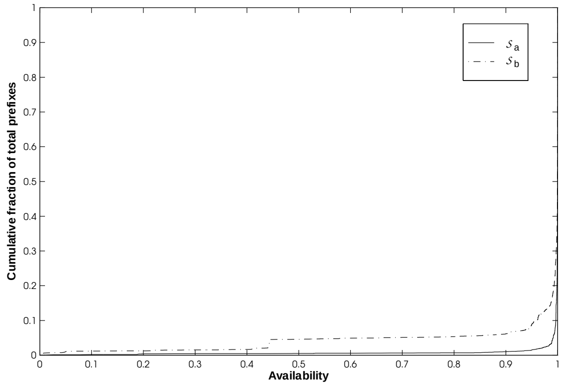


Figure 5: *Prefix Availability:* Prefix availability is defined as the fraction of time that the prefix was reachable, in the duration between when the prefix was first seen and the end of the snapshot. More than 90% of the prefixes are available for more than 95% of the time.

Prefix Availability

We define the *availability* of a prefix as the fraction of time for which it was reachable in the interval between when the prefix was first seen and the end of the snapshot. Figure 5 plots the cumulative fraction of prefixes in a snapshot against prefix availability. In general, prefix availability is high; about 10% of the prefixes were available for less than 95% of the time.

However, between $\mathcal{S}_a$ and $\mathcal{S}_b$, there is some evidence of degradation in availability. In $\mathcal{S}_a$, 90% of the prefixes are available for more than 99% of the time. In $\mathcal{S}_b$, that availability is 97%. There are two possible explanations for this degradation. First, the number of routers and links in paths to prefixes could have increased between the two snapshots. This increases the likelihood of router or link failure in these paths. Second, the errors due to routing misconfiguration could have increased, owing to an increase in the complexity of the inter-domain topology.

To nearly half the prefixes in each snapshot, we saw more than one `AS_PATH` (Section 4). Given that routing on the Internet is based on relatively static metrics, we might hypothesize that one of these paths is used for a significant fraction of the time that a prefix is reachable. Call this most frequently used path to a prefix its *primary* path. We can then define the *primary availability* of a prefix as that fraction of its reachable time for which the prefix was reachable by its primary path. Figure 6 shows the cumulative distribution of the fraction of prefixes against primary availability. For more than 80% of the prefixes, the primary availability is greater than 95%. This indicates that the outages in the primary, most preferred path, are relatively short-lived for most prefixes. The figure also shows improved primary availability in $\mathcal{S}_b$,

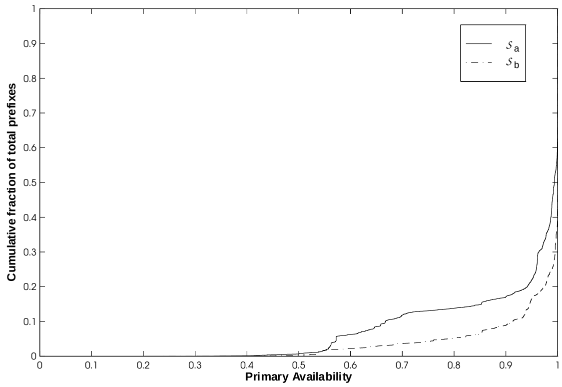


Figure 6: *Prefix Primary Availability:* Prefix primary availability is defined as the fraction of a prefix's reachable time that it is reachable by its primary path. More than 80% of the prefixes are reachable through a primary path for more than 95% of the time.

despite decreased overall availability.

Prefix Steadiness

We define the *steadiness* of a prefix as the mean duration of all intervals in a snapshot over which the prefix was continuously reachable. Figure 7 plots the cumulative distribution of the fraction of prefixes seen in each snapshot against steadiness. There is a wide variation in prefix steadiness. Remarkable, however, is the fact that almost 80% of the prefixes have steadiness values of more than 1 day! There is some evidence that prefix steadiness decreases with growth in the Internet; $\mathcal{S}_b$'s distribution of prefix steadiness is almost completely to the "left" of that of $\mathcal{S}_a$. Furthermore, the steadiness distribution shows several "steps", indicating that there exist clusters of prefixes with similar steadiness behavior. A closer examination of our traces reveals that prefixes in a cluster appear to share a common primary path.

We have already shown that, to most prefixes, there exists a highly available primary path. We might hypothesize that a prefix's steadiness is determined by the number of routers and links in its primary path. To a first approximation, we study a prefix's steadiness as a function of the number of domains in its primary path. Figure 8 plots the cumulative distribution of the fraction of prefixes seen in each snapshot against steadiness, for different path lengths. In a distribution sense, the steadiness of prefixes appears to generally decrease with path length. There is one exception. In $\mathcal{S}_b$, prefixes whose primary path length is three appear steadier than prefixes whose primary path length is two. A possible explanation for this is the existence of other factors, such as routing misconfiguration, that affect steadiness.

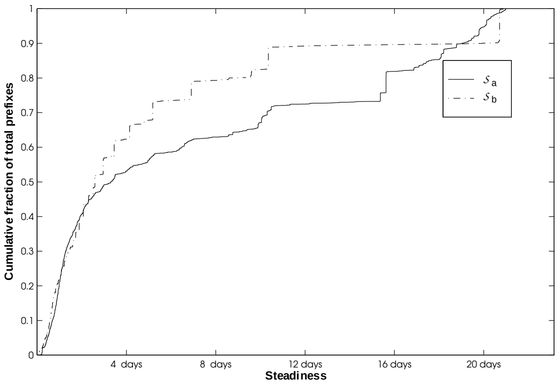


Figure 7: *Prefix Steadiness:* This figure plots the cumulative distribution of prefix steadiness. There is a wide variation in prefix steadiness, but nearly 80% of the prefixes have a steadiness greater than 1 day.

Discussion

The above sections describe the distributions of *prefix* availability and steadiness. Prefixes differ in the number of destinations they "cover". Little is known about the address space utilization of IP prefixes. A first order approximation is that this utilization is proportional to the address space covered by a prefix. Using this approximation, we can *estimate* the distribution of *host* availability and steadiness. These distributions are shown in Figure 9. The distribution of host availability (Figure 9 shows that, in $\mathcal{S}_a$, nearly 99% of hosts are available for more than 99% of the duration of our snapshot interval. In $\mathcal{S}_b$ only 95% of the hosts are available for more than 99% of the time. The steadiness curves are similar to those in Figure 7. However, the disparity in steadiness between $\mathcal{S}_a$ and $\mathcal{S}_b$ is greater in Figure 9.

Our prefix stability behavior is as observed from a single point on the backbone mesh. The behavior of individual prefixes is likely to differ when observed at different points on this mesh. However, we believe that the *distribution* of prefix availability and steadiness is likely to be similar at other backbone providers.

6 Related Work

To our knowledge, no prior work has attempted to characterize the Internet inter-domain topology and its growth. Several researchers have proposed inter-domain topology models [3, 12, 17]. These models posit a three- or four-level provider hierarchy. Some of these models also assume that domains in a class primarily connect to domains in classes above or below their own. Our analysis confirms the existence of a four-level hierarchy. However, our classification exposes a significant level of connectivity within classes, as well as

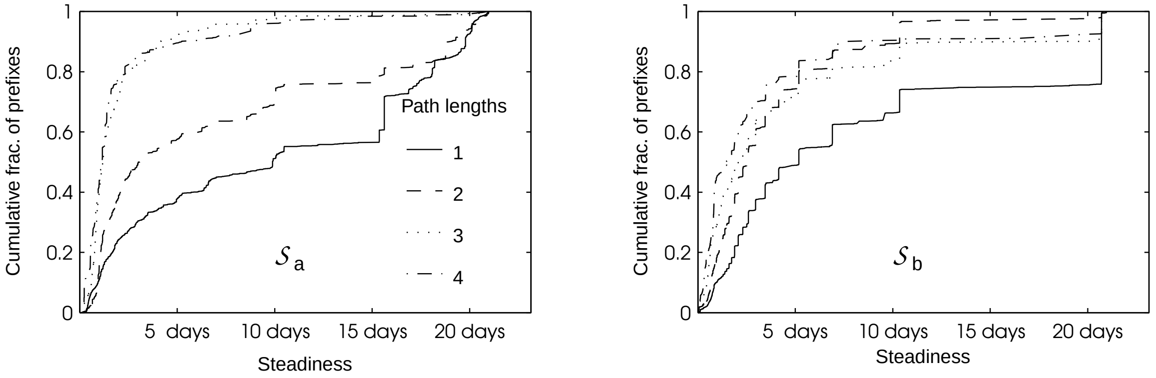


Figure 8: *Prefix steadiness for different path lengths:* This figure plots the cumulative distribution of prefix steadiness, for different primary path lengths, for the two snapshots. The distributions are consistently worse for longer path lengths, with one exception (in $\mathcal{S}_b$ for path lengths 2 and 3).

connectivity between non-adjacent classes.

Two other studies have attempted to characterize Internet route stability. One analyzes a 12-hour trace of routing traffic on the now defunct NSFNET backbone network [7]. The Internet topology has changed significantly since then. This study analyzes routing protocol behavior: the frequency and size of updates, routing protocol convergence times and so on. It does not analyze prefix availability and steadiness characteristics.

A second, more recent, study [15] uses the `traceroute` tool to sample availability and steadiness characteristics of paths between a small set of host pairs. Their end-to-end path availability estimates of 97% or higher appear to match our estimated host availability distribution; they too observe a degradation in host availability with Internet growth. Their distribution of the path with the highest availability, observed at "AS granularity" between host pairs, closely matches our primary prefix availability distribution.

7 Conclusions

There are two major conclusions about Internet inter-domain routing that emerge from our findings. First, the Internet has grown "laterally". Second, route stability behavior has degraded with growth. At a high level, these conclusions are not very surprising. The ubiquity of provider connectivity afforded by Internet exchange points encourages "lateral" growth (many providers at different levels, and constant number of levels) rather than "vertical" growth. The widespread use of manual routing configuration on the Internet today, and the lack of sophisticated routing analysis and debugging tools probably accounts for the decreased

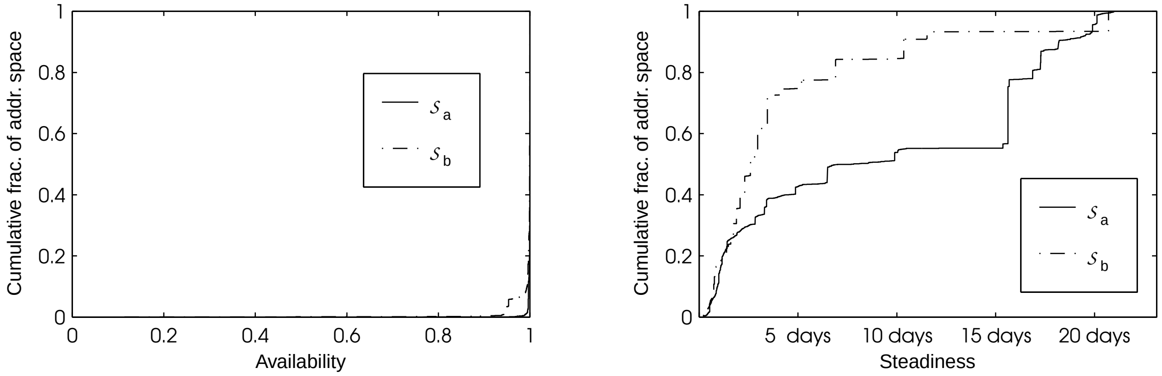


Figure 9: *Availability and steadiness of paths to hosts:* These graphs plot the estimated distribution of host availability and steadiness.

route stability.

We see no reason to believe that the pattern of growth in the inter-domain topology will change significantly in the near future. However, an increasingly commercial and competitive infrastructure is likely to spur the development of automated routing analysis and engineering tools, resulting in better route stability.

References

[1] C. Alaettinoğlu. Scalable Router Configuration for the Internet. *Proceedings of the 1996 International Conference on Networking Protocols*, October 1996.

[2] C. Alaettinoğlu, T. Bates, E. Gerich, D. Karrenberg, M. Terpstra, and C. Villamizar. Routing Policy Specification Language (RPSL). Internet Draft draft-ietf-rps-rpsl-00.txt, Network Information Center, March 1996. Revised, June 1996.

[3] C. Alaettinoğlu and A. U. Shankar. The Viewserver Hierarchy for Interdomain Routing: Protocols and Evaluation. *Journal on Selected Areas in Communications*, 13(8):1396–1410, October 1995.

[4] T. Bates, E. Gerich, L. Joncheray, J-M. Jouanigot, D. Karrenberg, M. Terpstra, and J. Yu. Representation of ip routing policies in a routing registry. Technical Report ripe-181, RIPE, RIPE NCC, Amsterdam, Netherlands, October 1994.

[5] J. C. Bolot. End-to-end Packet Delay and Loss Behavior in the Internet. In *Proceedings of the ACM SIGCOMM Symposium on Communication Architectures and Protocols*, pages 289–298, San Francisco, CA, Sep 1993.

[6] E. J. Bos. Routing Table Growth. `http://nic.surfnet.nl/net-management/ip/cidr`, 1995.

[7] B. Chinoy. Dynamics of Internet Routing Information. In *Proceedings of the ACM SIGCOMM Symposium on Communication Architectures and Protocols*, pages 45–52, San Francisco, CA, Sep 1993.

[8] MFS Datanet. Mae-East Traffic Statistics. `http://www.mfsdatanet.com/MAE/east.stats.html/`.

[9] D. Estrin, Y. Rekhter, and S. Hotz. A Scalable Inter-Domain Routing Architecture. In *Proceedings of the ACM SIGCOMM Symposium on Communication Architectures and Protocols*, pages 40–52, Baltimore, MD, August 1992.

[10] National Science Foundation. Network Access Point Manager, Routing Arbiter, Regional Network Providers and Very High-Speed Services Provider for NSFNET and NREN(SM) Programs. Program Solicitation 93-52, May 1993.

[11] V. Fuller, T. Li, J. Yu, and K. Varadhan. Classless Inter-Domain Routing (CIDR): An Address Assignment and Aggregation Strategy. Request for Comments 1519, Internic Directory Services, September 1993.

[12] S. M. Hotz. *Routing Information Organization To Support Scalable Interdomain Routing With Heterogeneous Path Requirements*. PhD thesis, University of Southern California, 1996.

[13] M. Lottor. Internet Domain Survey. Technical report, Network Information Systems Center, SRI International, October 1993.

[14] V. Paxson. Growth Trends in Wide-Area TCP Connections. *IEEE Network*, pages 8–17, July/August 1994.

[15] V. Paxson. End-to-end Routing Behavior in the Internet. In *Proceedings of the ACM SIGCOMM Symposium on Communication Architectures and Protocols*, San Francisco, CA, September 1996.

[16] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). Request for Comments 1771, Internic Directory Services, March 1995.

[17] E. W. Zegura, K. L. Calvert, and Samrat Bhattacharjee. How to Model an Internetwork. In *Proceedings of IEEE INFOCOMM*, pages 40–52, San Francisco, CA, September 1996.