USENIX Association

Proceedings of the 2002 USENIX Annual Technical Conference

Monterey, California, USA
June 10-15, 2002

Geographic Properties of Internet Routing

Lakshminarayanan Subramanian
University of California, Berkeley
lakme@cs.berkeley.edu

Venkata N. Padmanabhan
Microsoft Research
padmanab@microsoft.com

Randy H. Katz
University of California, Berkeley
randy@cs.berkeley.edu

Abstract

In this paper, we study the geographic properties of Internet routing. Our work is distinguished from most previous studies of Internet routing in that we consider the geographic path traversed by packets, not just the network path. We examine several geographic properties including the circuitousness of Internet routes, how multiple ISPs along an end-to-end path share the burden of routing packets, and the geographic fault tolerance of ISP networks. We evaluate these properties using extensive network measurements gathered from a geographically diverse set of probe points. Our analysis shows that circuitousness of Internet paths depends on the geographic and network locations of the end-hosts, and tends to be greater when paths traverse multiple ISP. Using geographic information, we quantify the degree to which an ISP's routing policy resembles hot-potato or cold-potato routing. We find evidence of certain tier-1 ISPs exhibiting hot-potato routing. Finally, based on network topology information gathered at CAIDA, we find that many tier-1 ISP networks may have poor tolerance to the failure of a single, critical geographic node, assuming the published topology information is reasonably complete.

1 Introduction

The Internet consists of several autonomous systems (ASes) that are under the control of different administrative domains. Routing across these administrative domains is accomplished using the Border gateway protocol (BGP), a protocol for propagating routes between ASes. ASes connect to each other either at public exchanges or at private peering points. The network path between two end-hosts typically traverses multiple ASes. BGP is flexible in allowing each AS to apply its own local preferences, and export and import policies for route selection and propagation. The characteristics of an end-to-end path are very much dependent on the policies employed by the intervening ASes.

Previous work on Internet routing has focused on studying properties such as end-to-end performance, routing stability, and routing convergence that are affected by routing policies. There has also been work on strategies for determining alternate (and hopefully better) routes by using overlay networks to circumvent the default Internet routing. We discuss previous work in more detail in Section 2.

In this paper, we present a novel way of analyzing certain properties of Internet routing. We show how *geographic* information can provide insights into the structure and functioning of the Internet, including the interactions between different autonomous systems. In particular, geographic information can be used to quantify well-known network properties such as hot-potato routing. It can also be used to quantify and substantiate prevalent intuitions about Internet routing, such as the relative optimality of intra-ISP routing compared to inter-ISP routing.

To analyze geographic properties of routing, it is necessary to first determine the *geographic* path of an IP route. The geographic path is obtained by stringing together the geographic locations of the nodes (i.e., routers) along the network path between two hosts. For instance, the geographic path from a host in Berkeley to one in Harvard may look as follows: Berkeley → San Francisco → New York → Boston → Cambridge. The level of detail in the geographic path would depend on how precisely we are able to determine the locations of the intermediate routers in the path. In Section 3, we describe GeoTrack [13], a tool we have developed for determining the geographic path of routes. Our study is based on extensive traceroute data gathered from 20 hosts distributed across the U.S. and Europe and also traceroute data gathered by Paxson [26] in 1995.

Internet routes can be highly circuitous. For instance, we observed a route from a host in St. Louis to one in Indiana (328 km away) that traverses a total distance of over 3500 km (Section 4.2.1). By tracing the geographic path, we are able to automatically flag such anomalous routes, which would be difficult to do using purely network-centric information such as delay. We compute the *linearized distance* between two hosts as the sum of the geographic lengths of the individual links of the path. We then compute the ratio of the linearized distance of the path to the geographic distance between the source and destination hosts, which we term the *distance ratio*. A large ratio would be indicative of a circuitous and

possibly anomalous route. In Section 4, we study circuitousness of paths as a function of the geographic and network locations of the end-hosts.

Our results indicate that the presence of multiple ISPs in a path is an important contributor to circuitous routing. We also find intra-ISP routing to be far less circuitous than inter-ISP routing. Our study of circuitousness of paths provides some insights into the peering and routing policies of ISPs. Although circuitousness may not always relate to performance, it can often be indicative of a routing problem that deserves more careful examination.

There are two extremes to the routing policy that an ISP may employ: *hot-potato* routing and *cold-potato* routing. In hot-potato routing, the ISP hands off packets to the next ISP as quickly as possible. In cold-potato routing, the ISP carries packets on its own network as far as possible before handing them off to the next ISP. The former policy minimizes the burden on the ISP's network whereas the latter gives the ISP greater control over the end-to-end quality of service experienced by the packets. As we discuss in Section 5.4, geographic information provides a means to quantify these notions by using the geographic distance traversed within an ISP as a proxy for the amount of work performed by the ISP. In addition, we can also evaluate the degree to which an individual ISP contributes in the routing of packets end-to-end. Our analysis of properties of paths that traverse multiple ISPs is presented in Section 5.

Another aspect of routing that bears careful examination is its fault tolerance. Fault tolerance has generally been studied in the context of node or link failures based on network-level topology information. However, such topology information may be incomplete in that two seemingly independent nodes may actually be susceptible to correlated failures. For instance, a catastrophic event such as an earthquake or a major power outage might knock out all of an ISP's routers in a geographic region. Geographic information can help in identifying routers that are co-located. In order to analyze the impact of correlated failures, we consider ISP topologies at the geographic level, where each node represents a geographic region such as a city. Using the geographic topology information of several commercial ISPs gathered from CAIDA [24], we analyze the fault tolerance properties of individual topologies and the topology resulting from the combination of the individual ISP networks (Section 6). We find that many tier-1 ISPs are highly susceptible to single geographic node failures. The combined topology however exhibits better tolerance to such failures.

In summary, we believe geography is an interesting means for analyzing and quantifying network properties. In some cases, our analysis provides additional evidence for existing intuition about certain properties of Internet routing (e.g., hot-potato routing, circuitous paths). An important contribution of our work is a methodology for quantifying such intuitions using geographic information. Such quantification enables us, for instance, to automatically flag circuitous paths, something that would be hard to using purely network-centric metrics (and no geographic information).

2 Related work

We classify related work into two categories: (a) Internet routing; (b) Topology discovery and mapping.

2.1 Internet routing

There are several properties of Internet routing that are of interest: end-to-end performance, routing stability, routing convergence, etc. Previous work on Internet routing has focused either on measuring these properties or on modifying certain aspects of routing with a view to improving performance. Our work shows how geographic information can be used to measure and quantify certain routing properties such as circuitous routing, hot-potato routing and geographic fault tolerance.

Network path information, obtained using the *traceroute* tool [8], has been used widely to study the dynamics of Internet routing. For instance, Paxson [14] studied various aspects of Internet routing using an extensive set of traceroute data. They include: routing pathologies, stability of routing, and routing asymmetry. In relation to our work, he studies circuitous routing by determining the geographic locations of the routers in his dataset and uses geographic distance as a metric to quantify it. In addition, he uses the number of different geographic locations along a path to analyze the effect of hot-potato routing as a potential cause for routing asymmetry. We extend this work by studying circuitousness as a function of the geographic and network location of end-hosts. We also analyze the effects of multiple ISPs in a path on its circuitousness. The distance ratio metric that we define can be used to automatically flag anomalies such as the large-scale route fluttering identified in [9, 14].

Overlay routing has been proposed as a means to circumvent the default IP routing. Savage et al. [17] study the effects of the routing protocol and its policies on the end-to-end performance as seen by the end-hosts. They show that for a large number of paths in the Internet, there exist paths that exhibit significantly better performance in terms of latency and packet loss rate. Recently, Andersen et al. [1] have proposed specific mechanisms for finding alternate paths with better performance char-

acteristics using an overlay network. By actively monitoring the quality of different paths, their alternate path selection mechanism can quickly recover from network failures and optimize application specific performance metrics.

Consistent with these findings, our measurements indicate the existence of highly circuitous paths in the Internet. We also find that the circuitousness of a path is correlated with the minimum end-to-end latency along the path.

2.2 Topology discovery and mapping

Discovering and analyzing Internet structure has been the subject of many studies. Much of the work has focused on studying topology purely at the network level, without any regard to geography. Recently several tools have been developed to map network nodes to their corresponding geographic locations. A few Internet mapping projects have used such tools to incorporate some notion of geographic location in their maps.

The Mercator project [6] focuses on heuristics for Internet Map Discovery. The basic approach is to use traceroute-like TTL limited probe packets coupled with source routing to discover routers[1]. A key component of Mercator is the set of heuristics used to resolve *aliases*, i.e., multiple IP addresses corresponding to (possibly different interfaces on) a single router. The basic idea is to send a UDP packet to a non-existent port on a router and wait for the ICMP *port unreachable* response that it elicits. In general, the destination IP address of the UDP packet and the source IP address of the ICMP response may not match, indicating that the two addresses correspond to different interfaces on the same router. In our work we use geographic information to identify points of sharing in the network. We view this as complementary to network-level heuristics such as the ones employed in Mercator.

The Internet Mapping Project [2] at Bell Labs also uses a traceroute-based approach to map the Internet from a single source. The map is colored according to the octets of the IP address, so portions corresponding to the same ISP tend to be colored similarly. The map, however, is not laid out according to geography. Other efforts have produced topological maps that reflect the geography of the Internet. Examples include the MapNet [24] and Skitter [28] projects at CAIDA and the commercial Matrix.Net service [25].

A number of tools have been developed for determining the geographic location corresponding to an IP address. These tools use a variety of approaches to map an IP address to location: inferring location from *Whois* records [7] (e.g., NetGeo [11]), extracting location information from traceroute data (e.g., GeoTrack [13], VisualRoute [30]), determining the location coordinates using delay measurements (e.g., GeoPing [13]), etc. Our previous work on IP2Geo [13] focused on developing several tools, including GeoTrack, to do IP-to-location mapping. In this work, we use the GeoTrack tool to analyze geographic properties of Internet routing.

[1] Actually, router *interfaces* are discovered, not routers.

3 Experimental methodology

In this section, we discuss our experimental methodology. We present the details of our measurement test bed and the data sets we gathered. We also discuss GeoTrack, the tool we used to determine geographic paths in the Internet.

3.1 Overview

Since the goal of our work is to study the geographic properties of Internet routing, much of our measurement work has focused on gathering network path data using the traceroute tool [8]. We are not interested in studying the dynamic properties of Internet routing (e.g., how routes change over time), so we only record a single snapshot of the network path between a given pair of hosts. It may possible that some of the routes in our dataset are backup paths due to failures at the time of our measurement. However, we do not expect the aggregate statistics reported in this paper to be affected by such failures since our measurements were spread over a $2-$month time period. We use traceroute to determine the network path between 20 traceroute sources and thousands of geographically distributed destination hosts.

Once we have gathered the traceroute data, we use the GeoTrack tool to determine the location of the nodes along each network path where possible. GeoTrack reports the location at the granularity of a city. We then use an on-line latitude-longitude server [18] to compute the geographic distance between the source and destination of a traceroute as well as between each pair of adjacent routers along the path. The latter enables us to compute the *linearized distance*, which we define as the sum of the geographic distances between successive pairs of routers along the path. So if the path between A and D passes through B and C, then the linearized distance of the path from A to D is the sum of the geographic distances between A & B, B & C, and C & D.

As we discuss in Section 3.4.1, we are typically able to determine the location of most but not all routers. We simply skip the routers whose locations we are unable to determine. So in the above example, if the location of C is unknown, then we compute the linearized distance of

the path from A to D as the sum of the geographic distances between A & B and B & D. Clearly, skipping over C would lead us to underestimate the linearized distance. However, as noted in Section 3.4.1, most of the skipped nodes are in the vicinity of the either the source or the destination, so the error introduced in the linearized distance computation is small.

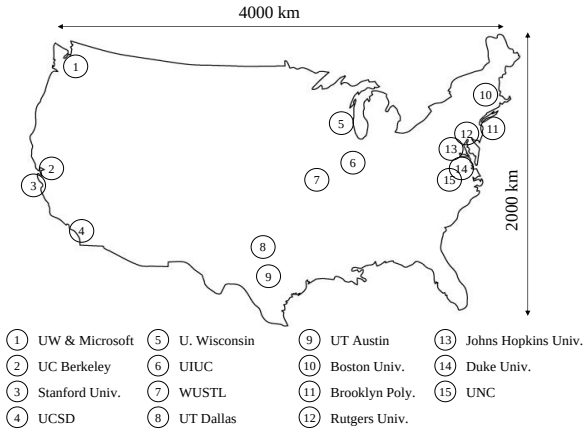

Figure 1: Locations of our traceroute sources in the U.S. Note that there were 17 hosts in 15 locations (two hosts each in Seattle and Berkeley).

3.2 Measurement testbed

We used 20 geographically distributed hosts as the sources for our traceroutes. 17 of these hosts were located in the U.S. (Figure 1) while 3 were located in Europe (at Stockholm (Sweden), Bologna (Italy), and Budapest (Hungary)). The geographical diversity in source locations enables us to study the variations in routing properties as seen from different vantage points. For logistical reasons, it was convenient for us to locate the traceroute sources on university campuses. 18 out of the 20 traceroute sources fell into this category. Furthermore, 9 of the 15 university locations we considered in the U.S. were connected by the Internet2 backbone [19]. To add some diversity, we had one source in Berkeley, CA connected to a home cable modem network (in addition to a host at the University of California at Berkeley) and another in Seattle, WA connected to the Microsoft Research network (in addition to a host at the University of Washington at Seattle). These two pairs of sources allow us to study (albeit to a limited extent[2]) what impact, if any, the nature of the source's connectivity has.

The destination set for the traceroutes comprised several thousand hosts. These destinations hosts fell into 4 categories:

1. *UnivHosts:* 265 Web servers and other hosts located on university campuses in the U.S. The hosts were distributed across 44 of the 50 states in the U.S.

2. *LibWeb:* 1,205 Web servers of public libraries [21] distributed across 49 states in the U.S. We also ensured that the distribution of the geographic locations of these libraries is not skewed.

3. *TVHosts:* 3,100 client hosts in the U.S. that connected to an on-line TV program guide. A majority of these clients were located on non-academic networks such as America Online (AOL).

4. *EuroWeb:* 1,092 Web servers [23] distributed across 25 countries in Europe.

For ease of exposition, we sometimes refer to UnivHosts, LibWeb, and TVHosts as the U.S. hosts and EuroWeb as the European hosts.

This diverse set of destination hosts enables us to investigate the properties of Internet routing in the context of a large set of ISPs. In all, we traced approximately 84,000 end-to-end paths between our traceroute sources and the destination hosts during October-December 2000. Our data is available online at [27].

3.3 Dataset from 1995

To study the temporal variations in Internet properties, we use the traceroute data set collected by Paxson in 1995 [26]. The data set includes traceroutes conducted between pairs of hosts drawn from a set of 33 hosts distributed across (mainly academic sites in) the U.S., Europe, South Korea, and Australia.

Despite the fact that the 1995 data set contains far fewer paths than the 2000 data set, it provides an interesting data point for comparison. The 1995 data set was gathered in late 1995, about 6 months after the demise of the NSFNET backbone (which used to provide connectivity to academic sites in the U.S.) and early in the life of the commercial Internet.

3.4 GeoTrack

Once we have gathered traceroute data, we use the GeoTrack tool, which we developed previously as part of the IP2Geo project [13], to translate the network path between a pair of hosts to the corresponding geographic path. GeoTrack tries to infer the location of a router based on its DNS name. Network operators often assign

[2]We could have used a diverse set of public traceroute servers [22] to overcome this limitation. However, the large volume of traceroutes that we were looking to run from each source precluded this.

geographically meaningful names to routers[3], presumably for administrative convenience. For example, the name *corerouter1.SanFrancisco.cw.net* corresponds to a router located in San Francisco. However, not all router names are *recognizable* (i.e., some router names may not contain an indication of location).

Here is a brief outline of how GeoTrack works; please refer to [13] for a more detailed description. The DNS name of the router is parsed to determine if it contains any location codes. GeoTrack uses a database of approximately 2000 location codes for cities in the U.S. and in Europe. Each ISP tends to use its own naming convention, so there may be multiple codes for each city (e.g., *chcg, chcgil, cgcil, chi, chicago, ord* for Chicago, IL). GeoTrack incorporates ISP-specific parsing rules that specify the subset of valid codes and the position(s) in which they may appear in the router names.

We use the domain name of a router to decide which ISP it belongs to. While this heuristic works reasonably well, it is not perfect because multiple domain names may correspond to the same administrative domain (e.g., *alter.net* and *uu.net*), often due to the merger of what were once independent networks. For the same reason, even AS numbers would not enable us to determine the administrative domain boundaries with complete accuracy.

3.4.1 Coverage of GeoTrack

Of the 11,296 *.net* router names in our traceroute data set, 7842 were recognizable (approximately 70%). We compiled a list of 13 major ISPs with nationwide backbones in the U.S. or with international coverage: Sprintlink, AT&T, Cable and Wireless, Internet2, Verio, BBNPlanet[4], Qwest, Level3, Exodus, PSINet, UUNET/Alter.net, VBNS, and Global Crossing. We found that 5,966 of the 6,859 router names for these major ISPs were recognizable (87%). In some individual cases, such as AT&T and UUNET, the recognizability was in excess of 95%.

By manual inspection, we found that a large chunk of the router names which are unrecognizable by our tool have no meaningful codes to decipher their locations. Many unrecognizable router names tend to be concentrated in regional or campus networks. (For example, *cmu.psc.net* is a node in Pittsburgh, PA. However, since it does not contain a valid city or airport code, GeoTrack is unable to recognize its location.[5]) Figure 2 shows that recognizability is lowest close to the start and the end of the path. (The peak corresponding to the very beginning of the path is due to the source location always being known.) Thus most of the unrecognizable nodes are typically located in the vicinity of the source or the destination, so the resulting error in linearized distance is minimal.

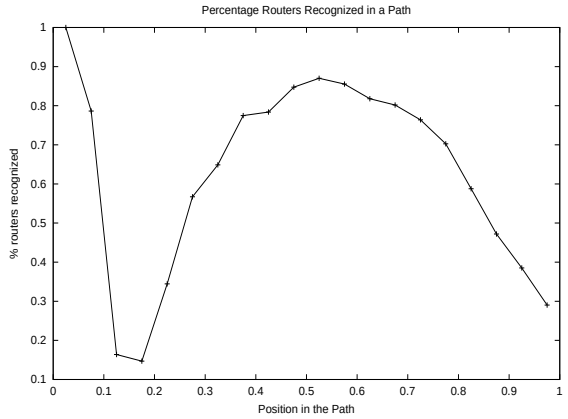

Figure 2: The recognizability of router names as a function of the position of the router in the end-to-end path. The position is quantified by dividing the number of hops leading up to the router by the total number of hops end-to-end.

In the case of the 1995 data set, GeoTrack is able to recognize 1,289 out of 1,531 router names (approximately 84%). Interestingly, we noticed a huge difference in the naming convention used in 1995 and 2000. Hence we needed to create a new set of codes for the 1995 data set.

3.4.2 Possible inaccuracies

First, the city codes used in GeoTrack for computing the location of router given its label are manually determined and encoded. Hence there is always a possibility that the location of a router as determined by GeoTrack is incorrect. However, we have greatly reduced the possibility of such errors by using delay-based verification, ISP specific parsing rules and manual inspection. In delay-based verification, we perform the following simple check: if the difference between the minimum RTTs to two adjacent routers in a path is not high, the distance between them cannot be large. This simple check helped us distinguish between two cities named *Geneva* that had similar city codes — one in Switzerland and the other in

[3]To be precise, DNS names are associated with router *interfaces*, not routers themselves. However, for ease of exposition we simply use the term router.

[4]BBNPlanet is now called Genuity, but the router names are still in the *bbnplanet.net* domain.

[5]Of course, it is possible to include *psc* and *cmu* as codes. However, we refrain from doing so since we only want to include those codes in GeoTrack that inherently indicate location. Doing otherwise would lead us down the path of exhaustive tabulation, which is undesirable.

Texas. We have enumerated specific rules for 52 different ISPs (all major ISPs in our data set) which specify the exact position where a city code is embedded in a label. This, in conjunction with ISP specific city-codes, greatly reduces the chances of a wrong location output. We have also manually inspected the geographic paths corresponding to a large sample of our traceroute data to check for any possible errors.

Second, the linearized distance computed can be distorted if the geographic locations of many routers in a path are unknown. We reduce this distortion by restricting our analysis to paths that have at least 4 recognizable intermediate routers. The linearized distance of a path can also be skewed due to intra-metro distances. Intra-metro distances will affect our analysis only for small values of linearized distances. To reduce this skew, we only consider paths with a linearized distance greater than 100 kms in our study.

3.5 Limitations

We now discuss the limitations of our study arising both due to the inherent limitations of geographic information and due to limitations of our experimental methodology.

1. **Geography does not determine performance:** There is not a perfect relationship between geographic distance and network performance. It is possible that a circuitous path yields better performance than a less circuitous one. For instance, the most optimal path between certain countries may be via the U.S. even if that means a large detour in geographic terms. However, in Section 4.5, we show that there exits a strong correlation between the minimum end-to-end delay between two end-hosts and the linearized distance of their connecting path. In light of this, we view our geographic analysis of network paths as providing (a) hints on paths that are *potentially* anomalous and should be examined more closely to determine if they are indeed anomalous, (b) an indication of how much improvement there could be in end-to-end latency if a non-circuitous path between source and destination were feasible, and (c) a way to quantify network properties such as hot-potato routing, which may provide new insight into these properties.

2. **IP-level topology is incomplete:** Our linearized distance computation only considers the router-level (i.e., IP-level) topology. We have no way of discovering the underlying physical topology (which may be based on ATM, SONET, or other technologies), so in general we would underestimate the linearized distance. While this is a limitation of our methodology, we note that the trend in high-speed networks (OC-48 and faster) is away from separate layer-2 and layer-3 architectures (e.g., IP-over-ATM) and towards an all-IP network [15]. This trend increases the applicability of our methodology.

4 Circuitousness of Internet paths

In this section, we examine the nature of circuitous routes in the Internet. Since there is not a standard measure of circuitousness, we define a metric, *distance ratio*, as the ratio of the linearized distance of a path to the geographic distance between the source and destination of the path. The distance ratio reflects the degree to which the network path between two nodes deviates from the direct geographic path between the nodes. A ratio of 1 would indicate a perfect match (i.e., an absolutely direct route) while a large ratio would indicate a circuitous path.

We present several different analysis with a view to studying the impact of spatial factors as well as temporal factors. Under spatial factors, we study the effect of the geographic and network locations of end-hosts on the circuitousness of paths. To study temporal properties, we compare the circuitousness of paths drawn from Paxson's 1995 data set to the ones drawn from our 2000 data set. Finally, we analyze the relationship between the minimum delay between two end-hosts and the linearized distance along their path.

4.1 Effect of network location

In this section, we will vary the network location of the end-hosts (source and destination) and study its effect on the distance ratio of paths. In our first analysis, we fix a source and compare the distance ratio of paths to destinations in different networks. In our second analysis, we compare the distance ratio of paths from different sources in the same geographic location but with different network connectivities to a set of end-hosts in the same network.

4.1.1 Paths from a single source

We consider paths from our traceroute sources in U.S. universities to two varied set of end-hosts: UnivHosts and TVHosts. Many of the hosts in UnivHosts (including our sources) connect to the Internet2 high-speed backbone via a local GigaPOP. So much of the wide-area path between our sources and a host in UnivHosts traverses the Internet2 backbone. On the other hand, TVHosts is a more diverse set that includes hosts located in various commercial networks (AOL, MSN, @Home, etc.) as well as university campuses. So the wide-area paths

from our sources to the hosts in TVHosts typically traverse one or more commercial ISP backbones.

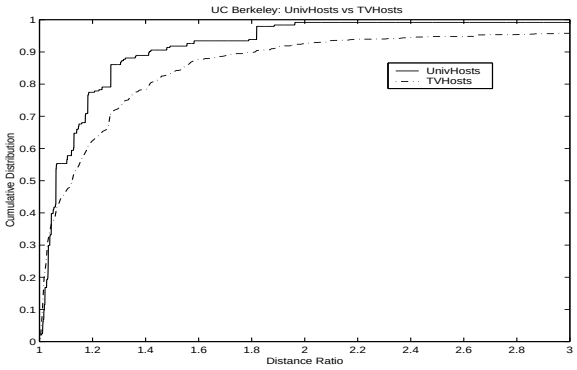

Figure 3: CDF of distance ratio for paths from UC Berkeley to UnivHosts and TVHosts.

This difference between the two groups of destination hosts is reflected in the cumulative distribution function (CDF) of the distance ratio for the two cases. As Figure 3 shows (for source in UC Berkeley), the distance ratio is close to 1 for many of the destinations. The ratio is 1.1 or less (corresponding to a linearized distance that exceeds the end-to-end geographic distance by no more than 10%) for 55% of the destinations in UnivHosts and 45% in TVHosts. This finding is consistent with the rich Internet connectivity of the San Francisco Bay Area (where UC Berkeley is located). The area includes several public Internet exchanges (e.g., MAE-West, PAIX, etc.) as well as private peering points. So a path from the UC Berkeley host to a destination host is often (but not always) able to transition to the latter's ISP within the SF bay area itself. So there is little need to take a detour through another city just to transition to the destination's ISP.

There is a far more pronounced difference between the UnivHosts and TVHosts cases if we look at the tail of the distribution. For instance, at the 90th percentile mark, the distance ratio is 1.41 in the case of UnivHosts but 1.72 in the case of TVHosts; in other words, the detour is 1.75 times as large for TVHosts destinations as it is for UnivHosts (72% versus 41%). The paths to some of the hosts in TVHosts tend to be more circuitous because they traverse multiple commercial ISPs whose peering relationships may cause detours in the end-to-end path. We discuss this issue in more detail in Section 5. We observe qualitatively the same trends for other university sources as well; i.e., the distance ratio tends to be smaller for paths leading to UnivHosts compared to TVHosts.

4.1.2 Multiple sources in the same location

We now consider paths from pairs of hosts in the same location but on entirely different networks to destinations in the UnivHosts set. We consider two such pairs of traceroute sources: (a) a machine on the Berkeley campus and another also in Berkeley but on @Home's cable modem network, and (b) a machine at the University of Washington (UW) campus in Seattle and another on the Microsoft Research network 10 km away.

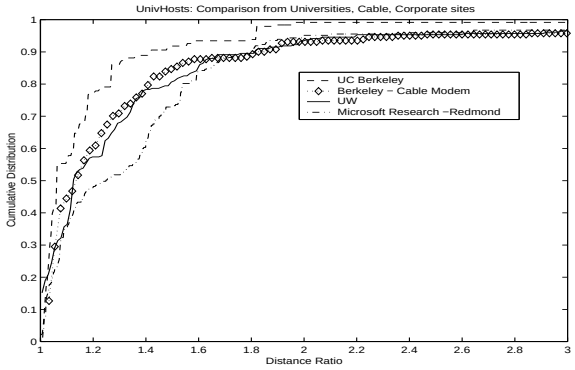

Figure 4: CDF of distance ratio for paths from pairs of co-located sources to UnivHosts.

Figure 4 shows the CDF of the distance ratio for all 4 sources. For the two sources located in Berkeley, we find that the one on the university campus has a significantly smaller distance ratio, especially at the tail of the distribution. For instance, the 90th percentile of the distance ratio for the UC Berkeley source is 1.41 while that for the cable modem source is 1.83. Since the destination set is UnivHosts, the UC Berkeley source tends to have more direct routes (via Internet2) than the cable modem client has (via @Home and other commercial ISPs).

We observe a similar trend for the UW-Microsoft pair. The UW source has more direct routes to other university hosts than does the Microsoft source. For instance, the path from Microsoft to the University of Chicago follows a highly circuitous route through BBNPlanet's (Genuity) network. The geographic path traversed includes Los Angeles, Carlton (TX), Indianapolis and Chicago (in that order). The linearized distance of the path is 4976 km while the geographic distance between Seattle and Chicago is only 2795 km. In contrast, the path from UW (via Internet2) is far more direct: it passes through Denver, Kansas City, Indianapolis, and finally Chicago, for a total linearized distance of 3533 km.

These results indicate that the nature of network connectivity of the source and the destination has a significant impact on how direct or circuitous the network paths are.

4.2 Effect of geographic location

The geographic location of a source indirectly determines its network connectivity. Sources near well-connected geographic locations like the Bay Area can potentially have less circuitous routes since many commercial ISPs will have a POP very close to them. To better understand the effect of geographic location, we compare the distance ratios of sources in different locations to a common set of destination end-hosts. We extend this analysis to study the role of network structures in different continents (U.S and Europe) on the circuitousness of paths.

4.2.1 Multiple sources in different locations

We consider paths from sources in three geographically distributed locations in the U.S.: Stanford, Washington University at St. Louis (WUSTL), and the University of North Carolina (UNC). The destination set is LibWeb, which is a larger and more diverse set than the UnivHosts set considered in Section 4.1.2.

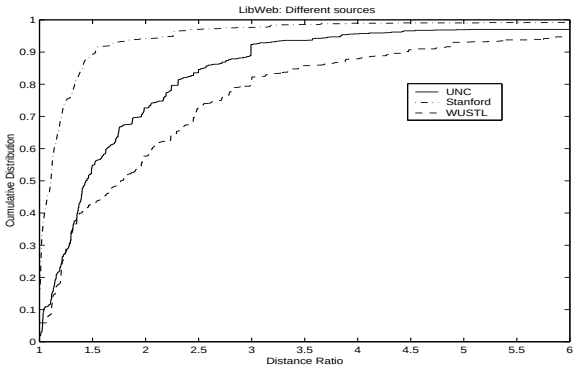

Figure 5: CDF of distance ratio for paths from multiple sources to LibWeb.

As shown in Figure 5, the distance ratio tends to be the smallest for paths originating from Stanford and the largest for those originating from WUSTL. Stanford, like Berkeley, is located in the San Francisco Bay area, which is well served by many of the large ISPs with nationwide backbones. In contrast, WUSTL is much less well connected. Almost all paths from WUSTL enter Verio's network in St. Louis and then take a detour either to Chicago in the north or Dallas in the south. At one of these cities, the path transitions to another major ISP such as AT&T, Cable & Wireless, etc. and proceeds to the destination. Any detour is particularly expensive in terms of the distance ratio because the central location of St. Louis in the U.S. means that the geographic distance to various destinations is relatively small.

In general, paths (such as those from WUSTL) that traverse significant distances in the backbones of two or more large ISPs tend to be more circuitous than paths (such as those from Stanford) that traverse much of the end-to-end distance in the backbone of a single ISP (regardless of who the ISP is). One example of a highly circuitous path we found involved two large ISPs, Verio and AT&T. The path originates in WUSTL in St. Louis and terminates at a host in Indiana University, 328 km away. However, the geographic path goes from St. Louis to New York via Chicago, all on Verio's network. In New York, it transitions to AT&T's network and then retraces its path back through Chicago to St. Louis, before finally heading to Indiana. The linearized distance is 3500 km, more than 10 times as much as the geographic distance. We examine the impact of multiple ISPs in greater detail in Section 5.

While the specific findings pertaining to Stanford and WUSTL may not be important in general, our results suggest that the distribution of the distance ratio is consistent with our intuition about the richness of connectivity of hosts in different geographic locations.

4.2.2 U.S. versus Europe

We now analyze the distance ratios for paths in Europe and compare these to the distance ratios for paths in the U.S. We consider paths from the 17 U.S. sources to destinations in the LibWeb set and also paths from the 3 European sources to destinations in the EuroWeb set. Thus, all of these paths are contained either entirely within the U.S. or entirely within Europe. We do not consider paths from U.S. sources to European destinations (or vice versa) because the distance ratio for such paths tends to be dominated by long transatlantic links (which tends to push the ratio towards 1).

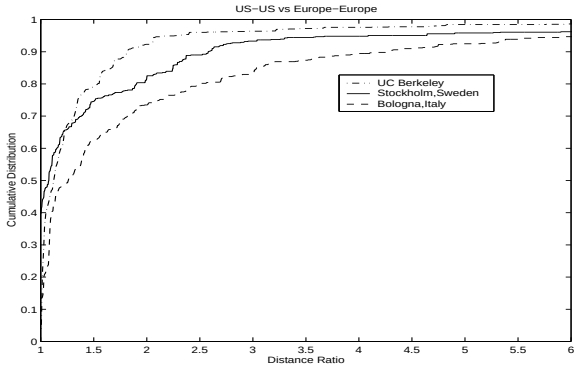

Figure 6: CDF of distance ratio for paths within the U.S. and those within Europe.

In Figure 6, we show the distribution of the distance ratio for three sources: Berkeley in the U.S., and Stockholm (Sweden) and Bologna (Italy) in Europe. We observe that the distance ratio tends to be larger for the Eu-

ropean sources compared to Berkeley, especially in the tail of the distribution. We attribute this to three causes.

First, paths in Europe tend to traverse multiple regional or national ISPs. The complex peering relationships between these ISPs often results in convoluted paths. For instance, a path from Bologna to a host in Salzburg, Austria traverses 3 ISPs – GARR (Italian Academic and Research Network), Eqip/Infonet, and KPNQwest (a leading pan-European ISP based in the Netherlands) – and passes through Milan (Italy), Geneva (Switzerland), Paris (France), Amsterdam (Netherlands), Frankfurt (Germany), and Vienna (Austria). The linearized distance of the path is 2506 km whereas the geographic distance between Bologna and Salzburg is only 383 km.

Second, in some cases the path from a European source to a European destination passes through nodes in the U.S. For instance, a path from Stockholm (Sweden) to Zagreb (Croatia) passes through a node in New York City belonging to Teleglobe, a large international ISP. In the event that the ISPs in Europe have better connectivity to ISPs in U.S., it would be appropriate for them to route their traffic through U.S. though the route may be more circuitous. Third, geographic distances in Europe tend to be smaller than the ones in U.S. As in the case of St Louis in Section 4.2.1, small detours in routing can be particularly expensive in terms of the distance ratio for paths between end-hosts in Europe.

4.3 Temporal properties of routing

To better understand some of the temporal properties of routing, we compare the distribution of the distance ratio computed from our 2000 data set with that computed from Paxson's 1995 data set [20]. The paths in the 1995 data set correspond to traceroutes conducted amongst the 33 nodes (mainly at academic locations) that were part of the testbed. We considered 340 paths between the subset of 20 nodes that were located in the U.S. The 1995 data set includes multiple traceroute measurements between each pair of hosts. In our study, we only use data from one successful traceroute between each pair of hosts. To keep the nature of the measurement points similar, in the 2000 data set we only consider paths between the 15 source hosts located at universities and the 265 hosts in the UnivHosts set.

Figure 7 plots the CDF of the distance ratio for the 1995 and 2000 data sets. By observing the tail of the cumulative distribution, we find that the distance ratios tend to be smaller in the 2000 data set. This improvement is not surprising because the Internet is more richly connected today than it was 5 years ago. There now exist direct point-to-point links between locations that were previously connected only by an indirect path.

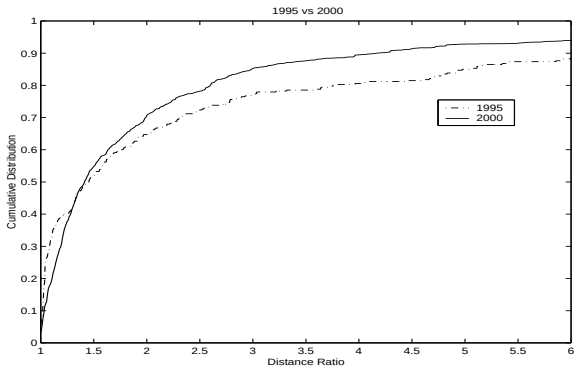

Figure 7: CDF of distance ratio for paths in Paxson's 1995 data set and our data set from 2000.

4.4 Correlation between delay and distance

Finally, we analyze the relationship between geography and the end-to-end delay along a path. Though geography by itself cannot provide any information about many performance characteristics like bandwidth, congestion along a path, the linearized distance of a path does enforce a minimum delay along a path (propagation delay along a path).

To study this correlation, we use the TVHosts data set since it represents a wide variety of end-hosts. In our traceroute data, we obtain 3 RTT samples for every router along the path. Since not all routers in a path are recognizable, we consider the minimum RTT, geographic distance and linearized distance to the last recognizable router along the path. In this analysis, we restrict ourselves to the list of probes in the U.S.

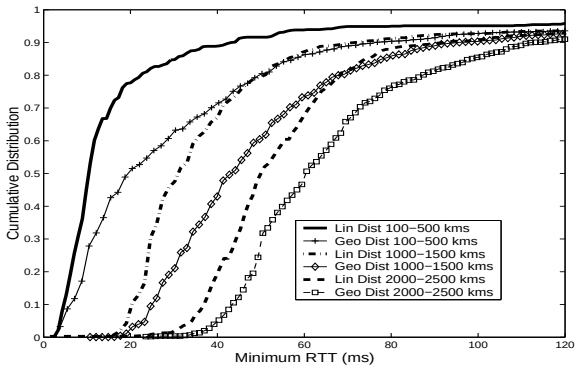

Figure 8: CDF of minimum end-to-end RTT to TVHosts for different ranges of linearized distances and geographic distances of paths

Figure 8 illustrates the correlation of the minimum RTT along a path to the linearized distance of a path and the geographic distance between the end-hosts. We make

three important observations. First, at low values of the linearized distance there exists a strong correlation between the delay and linearized distance for a large fraction of end-hosts especially for small values of linearized distances. We expect this correlation to be much stronger as we compute the minimum over a larger number of samples. Second, linearized distance along a path does enforce a minimum end-to-end RTT which is an important performance metric for latency sensitive applications. Third, the minimum RTT between two end-hosts has lesser correlation to the geographic distance between them as compared to the linearized distance of the path connecting them. We observe that for a given range of linearized distance of a path, the RTT variation is much smaller than its variation for the same range of geographic distance between the end-hosts. Hence linearized distance of a path conveys more about the minimum RTT characteristics of a path than merely the geographic distance between the end-hosts. We also verified that these observations hold across the other data sets we collected. The coarse correlation between minimum delay and geographic distance was used in building GeoPing, an IP-to-location mapping service [13].

4.5 Summary of Results

From Sections 4.1 and 4.2, we observe that the circuitousness of a route depends on both the geographic and network location of the end-hosts. In many cases, the trends we observe in the distance ratio are consistent with our intuition. A large value of the distance ratio enables us to automatically flag paths that are highly circuitous, possibly (though not necessarily) because of routing anomalies. Finally, we show that the minimum delay between end-hosts and the linearized distance of their path are strongly correlated. This relationship indicates that the circuitousness of a route does have an effect on the delay observed along the route (though this does not completely dictate the performance along the route).

5 Impact of multiple ISPs

Our analysis in Section 4 focused on the characteristics of the end-to-end path from a source to a destination. The end-to-end path typically traverses multiple autonomous systems (ASes). Some of the ASes are stub networks such as university or corporate networks (where the source and destination nodes may be located) whereas others are ISP networks. The relationships between these networks is often complex. There are customer-provider relationships (such as those between a university network and its ISP or between a regional ISP and a nationwide ISP) and peering relationships (such as those between two nationwide ISPs). A stub network may be multi-homed (i.e., be connected to multiple providers). Two nationwide ISPs may peer with each other at multiple locations (e.g., San Francisco and New York).

These complex interconnections between the individual networks have an impact on end-to-end routing. In this section, we show that geography can indeed be used as a means to analyze these complex interconnections. Specifically, we investigate the following questions: (a) are Internet paths within individual ISP networks as circuitous as end-to-end paths?, (b) what impact does the presence of multiple ISPs have on the circuitousness of the end-to-end path?, (c) what is the distribution of the path length within individual ISP networks, and (d) can geography shed light on the issue of hot-potato versus cold-potato routing?

5.1 Circuitousness of end-to-end paths versus intra-ISP paths

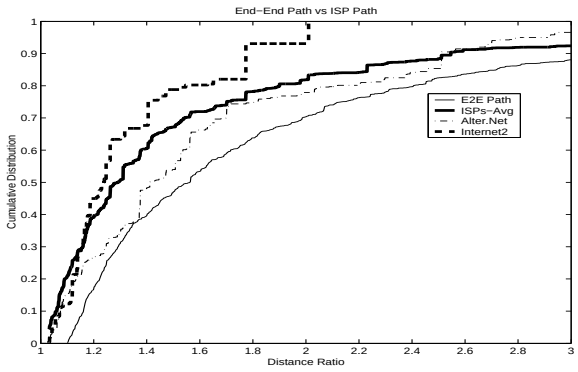

Figure 9: CDF of distance ratio of end-to-end paths versus that of sections of the path that lie within individual ISP networks.

We now take a closer look at the circuitousness of end-to-end Internet paths, as quantified by the distance ratio. We compare the distance ratio of end-to-end paths with that of sections of the path that lie within individual ISP networks. We consider paths from the U.S. sources to the LibWeb data set for this analysis.

As shown in Figure 9, the distance ratio of end-to-end paths tend to be significantly larger than that of intra-ISP paths. In other words, end-to-end paths tend to be more circuitous than intra-ISP paths. Furthermore, the distribution of the ratio tends to vary from one ISP to another, with Internet2 doing much better than the average and Alter.Net (part of UUNET) doing worse.

We believe the reason that end-to-end paths tend to more circuitous is that the peering relationship between ISPs may create detours that would otherwise not be present. Inter-domain routing in the Internet largely uses the

BGP [16] protocol. BGP is a path vector protocol that operates at the level of ASes. It offers limited visibility into the internal structure of an AS (such as an ISP network). So the actual cost of an AS-hop (in terms of latency, distance, etc.) is largely hidden at the BGP level. As a result the end-to-end path may include large detours.

Another issue is that ISPs typically employ BGP policies to control how they exchange traffic with other ISPs (i.e., which traffic enters or leaves their network and at which ingress/egress points). The control knobs made available by BGP include import policies such as assigning a local preference to indicate how favorable a path is and export policies such as assigning a multiple exit discriminator to control how traffic enters the ISP network [5]. These policies are often influenced by business considerations. For instance, packets from a customer of ISP A to a customer of ISP B in the same city might have to go via a peering point in a different city simply because a local service provider in the origin city who peers with both ISP A and ISP B does not provide transit service between the two ISPs.

Such BGP policies may partly explain the example mentioned in Section 4.2.1, where packets from a host in St. Louis to a nearby location had to travel on Verio's network all the way to New York to enter AT&T's network. We have seen several other such examples: a path from Austin, TX to Memphis, TN where the transition from Qwest to Sprintlink happens in San Jose, CA; a path from Madison, WI to St. Louis, MO where the transition from BBNPlanet to Qwest happens in Washington DC. We do not have specific information on the policies that were employed by these ISPs, so we cannot make a definitive claim that BGP is to blame. However, in view of the complex policies that come into play in the context of inter-domain routing, it is not surprising that end-to-end paths tend to be more circuitous.

In contrast, routing within an ISP network is much more controlled. Typically, a link-state routing protocol, such as OSPF [12], is used for intra-domain routing. Since the internal topology of the ISP network is usually known to all of its routers, routing within the ISP network tends to be close to optimal. So the section of an end-to-end path that lies within the ISP's network tends to be less circuitous. Referring again to the example in Section 4.2.1, both the St. Louis → Chicago → New York path within Verio's network and the New York → Chicago → St. Louis path within AT&T's network are much less circuitous than the end-to-end path.

However, this does not mean that intra-ISP paths are never circuitous. As noted in Section 4.1.2, we found a circuitous path through BBNPlanet (Genuity), from Microsoft Research in Seattle to the University of Chicago, that has a linearized distance of 4976 km whereas the geographic distance is only 2795 km. This does not imply that the path is necessary sub-optimal. In fact, the circuitous path may be best from the viewpoint of network load and congestion. The point is that while geography provides useful insights into the (non-)optimality of network paths, it only presents part of the picture.

5.1.1 Impact of path length on circuitousness

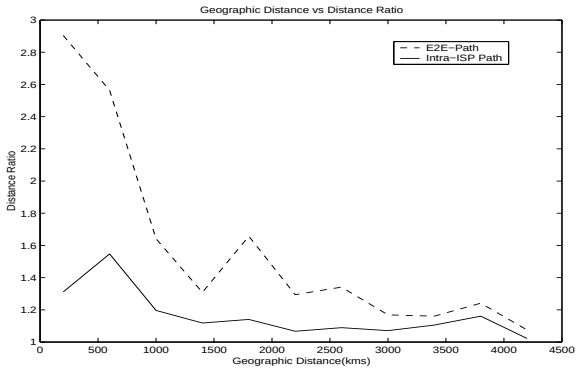

Figure 10: Distance ratio versus the geographic distance between the ends of a path. The median distance ratio is computed over 400 km buckets (0-400 km, 400-800 km, and so on). A minimum distance threshold of 100 km is imposed to prevent the ratio from blowing up, so the first bucket is actually 100-400 km.

One question that arises from the above analysis is whether there is a connection between the circuitousness of a path and its length (i.e., the geographic distance between the two ends of the path). In other words, are longer paths inherently more circuitous, regardless of whether they traverse one ISP or many? If so, the fact that end-to-end paths tend to be longer than intra-ISP paths may explain the greater circuitousness of the former.

However, as shown in Figure 10, the trend is quite the opposite. The distance ratio tends to decrease as the geographic distance increases.[6] The reason is that the impact of a detour is smaller (in relative terms) in the context of a longer path. The distance ratio for the end-to-end path tends to be greater than that for the intra-ISP path,

[6]The jaggedness of the curves arises because of the large variance in distance ratio for small values of geographic distance. The 5th and 95th percentile marks for the 100-400 km bucket are (1.00,20.50) for the end-to-end case and (1.00,4.22) for the intra-ISP case. The corresponding marks for the 4000-4400 km bucket are (1.01,1.57) for the end-to-end case and (1.00,1.18) for the intra-ISP case.

regardless of geographic distance. Thus the greater circuitousness of end-to-end paths is most likely due to the presence of multiple ISP networks in the path.

5.2 Impact of multiple ISPs on circuitousness

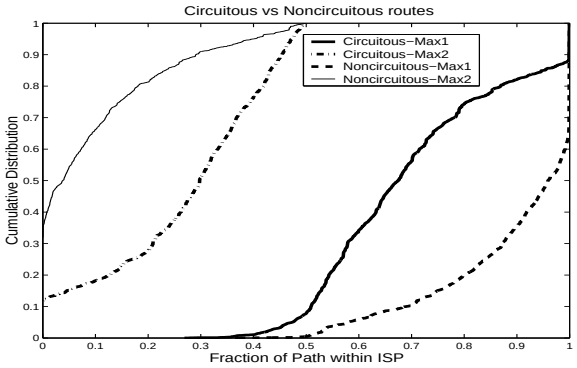

Figure 11: CDF of the fraction of the end-to-end path that lies within the top 2 ISPs in the case of circuitous paths and non-circuitous paths.

In Section 5.1 we hypothesized that the presence of multiple ISPs in an end-to-end path contributes to the circuitousness of the path. We now examine this issue more carefully. We classify end-to-end paths into two categories – non-circuitous (distance ratio < 1.5) and circuitous (distance ratio > 2).[7] For each path in either category, we identify the top two ISPs that account for most of the end-to-end linearized distance. We then compute the fraction of the end-to-end linearized distance that is accounted for by the top two ISPs, and denote these fractions by $\max_1$ and $\max_2$. For example, if an end-to-end path with a linearized distance of 1000 km traverses 400 km in AT&T's network and 300 km in UUNET's network (and smaller distances in other networks), then $\max_1 = 0.4$ and $\max_2 = 0.3$. Note that it is possible for $\max_1$ to be 1.0 (and so $\max_2$ to be 0.0) if the entire end-to-end path traverses just one ISP network. We note that local-area networks confined to a city (e.g., a university network) contribute nil to the linearized distance and therefore are ignored.

Figure 11 shows the CDF of $\max_1$ and $\max_2$ for the circuitous and non-circuitous paths. The difference in the characteristics of these two categories of paths is striking. The $\max_1$ and $\max_2$ curves are much closer together in the case of circuitous paths than in the case of non-circuitous paths. In other words, in the case of circuitous paths, the end-to-end path traverses substantial distances in each of the top two ISPs (and perhaps other ISPs too). In contrast, non-circuitous paths tend to be dominated by a single ISP. For instance, the median values of $\max_1$ and $\max_2$ in the case of circuitous paths is approximately 0.65 and 0.3, respectively. In other words, the top two ISPs account for 65% and 30%, respectively, of the end-to-end path in the median case. However, the fractions for the non-circuitous paths are approximately 95% and 4%, respectively – much more skewed in favor of the top ISP.

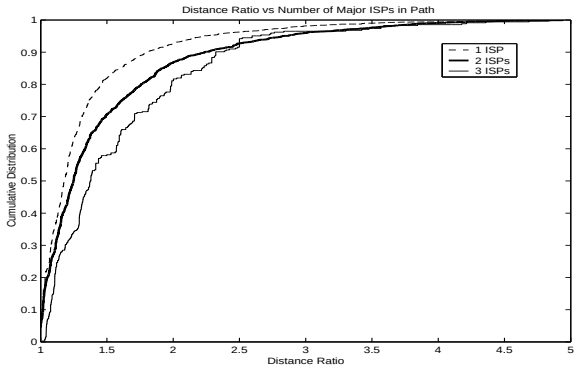

Figure 12: CDF of the distance ratio as a function of the number of major ISPs traversed along an end-to-end path. There were few paths that traversed more than 3 major ISPs.

We also consider the impact of the number of *major* ISPs traversed along an end-to-end path on the distance ratio. Figure 12 shows a clear trend: the distance ratio tends to increase as the path traverses a greater number of ISPs. For instance, the median distance ratios are 1.18, 1.25, and 1.38, respectively with 1, 2, and 3 major ISPs. The 90th percentile of the distance ratio is 1.81, 2.26, and 2.35, respectively. A path that traverses a larger number of major ISPs may span a greater distance. However, as noted in Section 5.1.1, this would not explain the larger distance ratio. In fact, a greater geographic distance would tend to make the distance ratio smaller, not larger

These findings reinforce our hypothesis that there is a correlation between the circuitousness of a path (as quantified by the distance ratio) and the presence or absence of multiple ISPs that account for substantial portions of the path.

5.3 Distribution of ISP path lengths

In this section, we further examine the distribution of the end-to-end linearized distance that is accounted for by individual ISPs. We wish to understand how the effort

[7] While the choice of these thresholds is arbitrary, they capture the intuitive notion of circuitous and non-circuitous routes. Note that there may be paths that do not fall into either category.

of carrying traffic end-to-end over a wide-area path is apportioned between different ISPs. For each of the 13 nationwide ISPs in the U.S. listed in Section 3.4.1, we consider the set of paths that traverse one or more nodes in that ISP's network. For each such path, we compute the fraction of the end-to-end path that lies within the ISP's network.

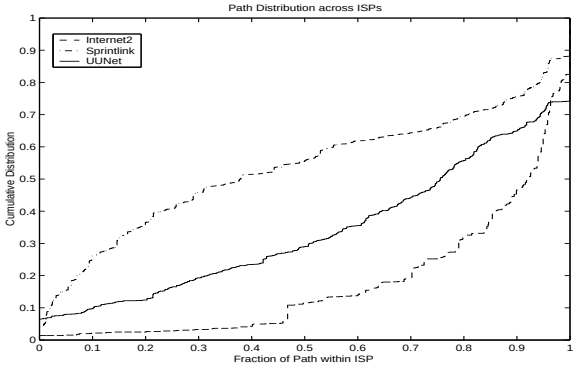

Figure 13: CDF of the fraction of the end-to-end path that lies within individual ISP networks.

Figure 13 plots the CDF of this fraction for a few ISPs. In each case, we consider the paths from the U.S. university sources to the LibWeb data set. We observe that the distributions look very different. For instance, the median fraction of the end-to-end path that lies within Sprintlink is only about 0.35 whereas the corresponding fraction for UUNet is 0.75 and for Internet2 is over 0.9. Internet2 is a high-speed backbone network that connects many university campuses in the U.S. An end-to-end path that traverses Internet2 typically originates and terminates at university campuses. Therefore, the Internet2 backbone accounts for an overwhelming fraction of such end-to-end paths. UUNET accounts for a larger fraction of the paths that traverse its backbone than any other commercial ISP we considered. This may reflect the close relationship between UUNET's parent company, Worldcom (which runs the vBNS backbone [29]), and academic sites.

The much smaller fraction in the case of Sprintlink is harder to explain definitively. From our conversations with people at Sprint [3, 10], we have learned that academic sites are not their major customers, so Sprintlink participates minimally in carrying academic traffic. The location of our traceroute sources at academic sites may explain why Sprintlink only accounts for a small fraction of the end-to-end path.

We stress, however, that the point of our analysis is not to make general claims about certain ISPs being better or worse than others. Rather it is to show that geographic analysis of end-to-end paths yields interesting insights into the role played by multiple ISPs in specific contexts (e.g., academic sites) and that these insights are consistent with our intuition.

5.4 Hot-potato versus Cold-potato routing

Finally, we investigate whether geographic information can be helpful in assessing whether ISP routing policies in the Internet conform to either hot-potato routing or cold-potato routing. In hot-potato routing, an ISP hands off traffic to a downstream ISP as quickly as it can. Cold-potato routing is the opposite of hot-potato routing where an ISP carries traffic as far as possible on its own network before handing it off to a downstream ISP. These two policies reflect different priorities for the ISP. In the hot-potato case, the goal is to get rid of traffic as soon as possible so as to minimize the amount of work that the ISP's network needs to do. In the cold-potato case, the goal is carry traffic on the ISP's network to the extent possible so as to maximize the control that the ISP has on the end-to-end quality of service. In general, an ISP's routing policy would lie somewhere in between the extremes of hot-potato and cold-potato routing.

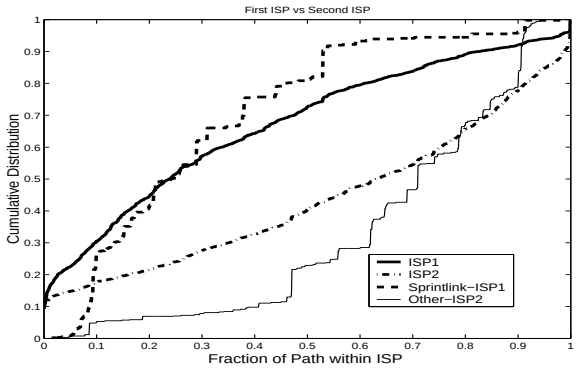

Figure 14: CDF of the fraction of the end-to-end path that lies within the first and second ISP networks in sequence.

We consider the set of paths from U.S. sources to TVHosts. For each path that traverses two or more major ISPs (with nationwide backbones), we compute the fraction of the end-to-end path that lies within the first major ISP (ISP1) and the second major ISP (ISP2) in sequence. We use these fractions as measures of the amount of work that these ISPs do in conveying packets end-to-end. The distributions of these fractions is plotted in Figure 14. We observe that the fraction of the path that lies within the first ISP tends to be significantly smaller than that within the second ISP. For instance, the median is 0.22 for the first ISP and 0.64 for the second ISP. This is consistent with hot-potato routing behavior because the first ISP tends to hand off traffic quickly to the second ISP who carries it for a much greater distance.

Figure 14 also plots the distributions of the path lengths in the case where the first ISP is Sprintlink. We find that the difference between the ISP1 and ISP2 curves is even greater in this case. Again, this is consistent with hot-potato routing behavior on the part of Sprintlink for routes from academic locations.

5.5 Summary

In this section, we have used geographic information to study various aspects of wide-area Internet paths that traverse multiple ISPs. We found that end-to-end Internet paths tend to be more circuitous than intra-ISP paths, presumably because of the peering relationships between ISPs. Furthermore, paths that traverse substantial distances within two or more ISPs tend to be more circuitous than paths that largely traverse only a single ISP. Some of this circuitous routing behavior can be attributed to sub-optimal geographic peering between ISPs. Finally, the findings of our geography-based analysis are consistent with the hypothesis that ISPs generally employ hot-potato routing. The presence of hot-potato routing may also explain for why some major ISPs only account for a relatively small fraction of the end-to-end path.

6 Geographic fault tolerance of ISPs

An important component of studying Internet routing is to understand its fault tolerance aspects. Fault tolerance of a network is normally studied at the granularity of router or link failures. However such a failure model does not capture the fact that two seemingly independent routers can be susceptible to correlated failures.

We ask the question: what is the tolerance of an ISP's network to a *total* network failure in a geographic region, i.e., a failure that affects all paths traversing the region? We refer to such a failure as a *geographic failure*. Potential reasons for such a failure include natural calamities such as earthquakes or power blackouts.

By using the geographic location information of the routers, we can identify routers that are co-located and thereby construct a *geographic topology* of an ISP. In this topology, each geographic region is associated with a node and an edge between two nodes signifies the existence of at least one long-haul backbone link that connects the corresponding geographic regions.

We obtained the geographic topologies for 9 of the 13 major ISPs listed in Section 3.4.1 from the CAIDA Mapnet site [24]. These are: AT&T, Cable and Wireless, Sprintlink, Genuity, Qwest, PSINet, UUNet, Verio and Exodus. Many of these topologies are obtained from information published at the ISPs' Web sites and are between 6-12 months out of date. Although it may be possible to construct an ISP's geographic topology using extensive traceroute measurements, it would be hard to assess the completeness of the constructed topology. Hence we restrict ourselves to the geographic topologies obtained from CAIDA. However, as acknowledged by CAIDA [24], it is possible that these topologies may themselves be incomplete. This may be due to limited tracing or the presence of backup paths in routing. We will perform our analysis under the assumption that these topologies are reasonably complete and only have a few missing links.

6.1 Degree distributions

The degree of a node provides a first-level quantification of the fault tolerance of that node in a given topology. A node with a degree k can tolerate up to k geographic failures before getting completely disconnected from all other nodes in the topology. In particular, a leaf node is not resilient to the geographic failure of its neighbor, but the failure of a leaf node itself has minimal impact on the rest of the network. On the other hand, the failure of a node with a very high degree would impact its many neighbors (corresponding to many different geographic regions).

Given complete freedom in placing $E = k * N$ edges on N nodes, it is possible to construct a topology that has a minimum vertex-cut of $2k$. In other words, the E edges can be placed in such a way that even in the presence of any $2k - 1$ node failures in the graph, the resulting topology will still remain connected. We term such a placement of edges that maximizes the size of the vertex cut as an *optimal placement*. In the optimal placement, all the vertices have the same degree, viz. $2 * k$. For the simple case of $k = 1$, the optimal placement results in a ring topology. Although this optimal placement may be difficult to construct due to practical constraints, it provides us a nice reference point for comparing the fault tolerance of ISP topologies. In order to contrast an ISP's topology from the optimal scenario, we look at the degree distribution of the nodes. We say that a graph has a *skewed* degree distribution if its node degrees are distributed over a wide range with a few large node degrees and a high percentage of the nodes are leaves. The Internet topology exhibits a skewed degree distribution which can be characterized by a power law as described in [4].

Among the 9 commercial ISPs, some of them such as AT&T and Genuity have a very skewed degree distributions while other ISPs such as PSINet and Verio have much less skewed degree distributions (closer to optimal). The degree distribution will not be affected much due to a few missing links. Figure 15 shows the degree distributions of AT&T and PSINet. AT&T's topology has the maximum percentage of leaves among the 9

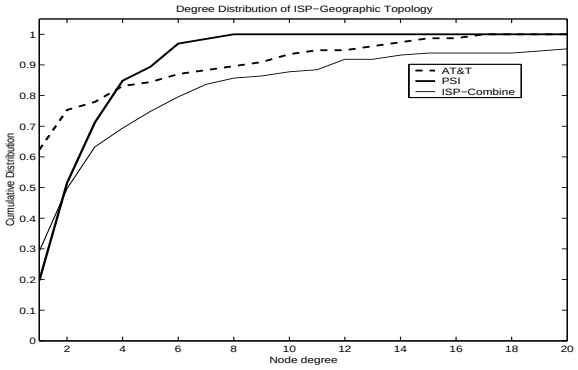

Figure 15: Degree Distribution of Geographic Topologies of ISPs

ISP topologies (62%) and has a few nodes with a degree greater than 12 (Chicago, Dallas). On the other hand, more than 50% of PSINet's nodes have a degree of either 2 or 3. This matches the optimal degree for Verio given that it has an edge to node ratio $k = 1.5$, which corresponds to an optimal degree of $2 * k = 3$. The ISP-Combine curve shows the degree distribution of the geographic topology obtained by combining the topology graphs of all 9 ISPs. The geographic nodes corresponding to the same city in the individual ISP topologies map to a single node in the combined topology. The combined topology still has a significant skew in its degree distribution. 29% of the nodes continue to be leaves. This happens despite the combined topology having an edge to node ratio of $k = 2.5$, which corresponds to an optimal degree of 5. On the other hand, nodes located in the important networking hubs of U.S. (e.g, San Jose, Washington DC, Chicago) have a degree of more than 20 in the combined topology.

6.2 Failure of high connectivity nodes

The skewed degree distributions of many tier-1 ISPs indicate that many geographic regions of an ISP may get disconnected if some high connectivity geographic nodes fail. To evaluate this, we consider the failure scenario where the f nodes of highest degrees in a graph fail.

We define a pair of geographic nodes that are connected by a network path and can communicate with each other as a *communicating pair*. A connected topology of N nodes can support $N(N+1)/2$ communicating pairs. (Since each node represents a geographic *region*, we also consider intra-node communication of a node with itself.) Under the scenario where the f nodes of highest degrees fail, the graph is disconnected into a forest where a node can only communicate with other nodes in its connected component. A connected component with $m < N$ nodes can support $m * (m+1)/2$ communicating pairs. In the simple case where the parent of a leaf node fails, it produces a connected component of size 1 which supports exactly one communicating pair.

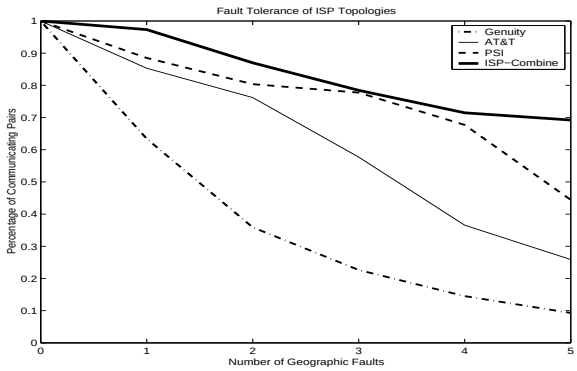

Figure 16: Tolerance to Geographic Failures

Figure 16 shows the percentage of communicating pairs supported in the various ISP networks in face of a varying number of geographic failures. The combined topology of the 9 ISPs supports 68% of the communicating pairs even after the removal of 5 important networking hubs in the US (San Jose, New York, Washington DC, Chicago, Los Angeles). Among the 9 ISPs, while Genuity and PSINet exhibit the least and the best fault tolerance characteristics. In the face of a single node failure, most of the ISPs lose between 15% and 30% of their communicating pairs in the worst case.

It is important to note that these results may represent a near-worst case failure scenario for the ISPs. If, however, many backup links are missing from our topology, the fraction of communicating pairs may be much higher than what we have portrayed. However, our essential message from this analysis is that a balanced degree distribution is a good feature for building a fault tolerant topology for an ISP.

7 Conclusions

In this paper, we have presented geography as a means for analyzing various aspects of Internet routing. First, our analysis based on extensive traceroute data shows the existence of many circuitous routes in the Internet. From the end-to-end perspective, we observe that the circuitousness of routes depends on the geographic and network locations of the end-hosts. We also find that the minimum delay along a path is more strongly correlated with the linearized distance the path than it is with the geographic distance between the end-points. This suggests that the circuitousness of a path does impact its minimum delay characteristics, which is an important end-to-end performance metric. In ongoing work, we are

studying the correlation between geography and network performance.

Second, a more careful examination shows that many circuitous paths tend to traverse multiple major ISPs. Although many of these major ISPs have points of presence in common locations, the peering between them is restricted to specific geographic locations, which causes the paths traversing multiple ISPs to be more circuitous. We also found that intra-ISP paths are far less circuitous than inter-ISP paths. An important requirement to reduce the circuitousness of paths is for ISPs to have peering relationships at many geographic locations.

Third, the fraction of the end-to-end path that lies within an ISP's network varies widely from one ISP to another. Furthermore, when we consider paths that traverse two or more major ISPs, we find that the path generally traverses a significantly shorter distance in the first ISP's network than in the second. This finding is consistent with the hot-potato routing policy. Using geographic information, we are able to quantify the degree to which an ISP's routing policy resembles hot-potato routing.

Finally, our analysis of geographic fault tolerance of ISPs indicates that the (IP-level) network topologies of many tier-1 ISPs exhibit skewed degree distributions which may induce a low tolerance to the failure of a single, critical geographic node. The combined topology of multiple ISPs exhibits better fault tolerance characteristics, assuming that the ISPs peer at all geographic locations that are in common.

Acknowledgments

Vern Paxson made his 1995 data set available to us. Arvind Arasu, B. R. Badrinath, Mary Baker, Paul Barford, John Byers, Imrich Chlamtac, Mike Dahlin, Kevin Jeffay, Craig Labovitz, Paul Leyland, Karthik Mahesh, Vijay Parthasarathy, Jerry Prince, Amin Vahdat, Srinivasan Venkatachary, Geoff Voelker, Marcel Waldvogel, and David Wood helped us obtain access to a geographically distributed set of measurement hosts. Vern Paxson and the anonymous USENIX reviewers provided useful comments on an earlier version of this paper. We would like to thank them all.

References

[1] D. G. Andersen, H. Balakrishnan, R.Morris, and F. Kaashoek. Resilient Overlay Networks, *ACM SOSP*, November 2001.

[2] B. Cheswick, H. Burch, and S. Branigan. Mapping and Visualizing the Internet, *USENIX Technical Conference*, June 2000.

[3] C. Diot. Personal communication, November 2001.

[4] M. Faloutsos, P. Faloutsos and C. Faloutsos. On Power-Law Relationships of the Internet Topology. *ACM SIGCOMM*, August 1999.

[5] L. Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE Global Internet*, November 2000.

[6] R. Govindan and H. Tangmunarunkit, Heuristics for Internet Map Discovery. *IEEE Infocom*, March 2000.

[7] K. Harrenstien, M. Stahl, and E. Feinler, NICKNAME/ WHOIS, *RFC-954, IETF*, October 1985.

[8] V. Jacobson, Traceroute software, 1989, *ftp://ftp.ee.lbl.gov/traceroute.tar.gz*

[9] C. Labovitz, J. Malan, and F. Jahanian. Internet Routing Instability. *ACM SIGCOMM*, August 1997.

[10] B. Lyles. Personal communication, August 2001.

[11] D. Moore et.al. Where in the World is netgeo.caida.org? *INET 2000*, June 2000.

[12] J. Moy. OSPF Version 2. *RFC-2328, IETF*, April 1998.

[13] V. N. Padmanabhan and L. Subramanian. An Investigation of Geographic Mapping Techniques for Internet Hosts. *ACM SIGCOMM*, August 2001.

[14] V. Paxson. Measurements and Analysis of End-to-End Internet Dynamics. Ph.D. dissertation, UC Berkeley, 1997. *ftp://ftp.ee.lbl.gov/papers/vp-thesis/dis.ps.gz*

[15] C. Semeria. Traffic Engineering for the New Public Network. Juniper Networks Whitepaper, September 2000.

[16] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). *RFC-1771, IETF*, March 1995.

[17] S. Savage, A. Collins, E. Hoffman, J. Snell and T. Anderson. The End-to-end Effects of Internet Path Selection, *ACM SIGCOMM*, pp 289-299, September, 1999.

[18] http://geography.about.com/

[19] Internet2. *http://www.internet2.org/*

[20] Internet Traffic Archive. http://ita.ee.lbl.gov/

[21] List of Public Libraries in the U.S. *http://sunsite.berkeley.edu/Libweb*

[22] List of Public Traceroute Servers *http://www.traceroute.org/*

[23] List of Web servers in Europe *http://pauli.uni-muenster.de/w3world/Europe.html*

[24] MapNet: Macroscopic Internet Visualization and Measurement. *http://www.caida.org/tools/visualization/mapnet/*

[25] Matrix.Net, *http://www.matrix.net*

[26] NPD-Routes data set, Internet Traffic Archive. *http://ita.ee.lbl.gov/html/contrib/NPD-Routes.html*

[27] Traceroute data used in this paper. *http://sahara.cs.berkeley.edu/rawtraces*

[28] Skitter project at CAIDA. *http://www.caida.org/tools/measurement/skitter/*

[29] vBNS: very high performance Backbone Network Service. *http://www.vbns.net/*

[30] VisualRoute, Visualware Inc. *http://www.visualroute.com/*