Geographic Properties of Internet Routing: Analysis and Implications

Lakshminarayanan Subramanian
University of California at Berkeley

Venkata N. Padmanabhan
Microsoft Research

Randy H. Katz
University of California at Berkeley

September 2001

Technical Report
MSR-TR-2001-89

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
USA

Geographic Properties of Internet Routing: Analysis and Implications

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

Venkata N. Padmanabhan
Microsoft Research
padmanab@microsoft.com

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

Abstract—

In this paper, we study the geographic properties of Internet routing. Our work is distinguished from 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 how circuitous Internet routes tend to be, how multiple ISPs along an end-to-end path share the burden of routing packets, and how much sharing there is between paths at the geographic level. We evaluate these properties using extensive network measurements. We examine both the spatial and the temporal variations of some of the properties. We discuss the implications of these geographic properties and conclude that geography is an interesting tool for analyzing the Internet properties.

Keywords— **Network measurement, Internet topology, routing, geography**

I. INTRODUCTION

In this paper, we study the geographic properties of Internet routing. There have been a number of studies of Internet routing. The key distinguishing feature of our work, however, is that we consider the *geographic path* traversed by packets, not just the network path. 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 looks 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 routers appearing in the network path.

Why is geography interesting in a networking context? As we show in this paper, geographic information can provide new insights into the structure and functioning of the Internet. It also provides a novel way of quantifying various network properties.

The specific issues we study include:

- **Circuitous routing:** Internet routes are not necessarily optimal. In fact, sometimes they are highly circuitous. For instance, we have 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 IV-C). 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 and the geographic distance between the source and destination hosts. A large ratio would be indicative of a circuitous (and possibly anomalous) route.
- **Hot-potato versus cold-potato routing:** In today's Internet, a packet typically traverses multiple ISP networks on its way from source to destination. An ISP typically employs either "hot-potato" routing, where it hands off packets to the next ISP as soon as possible, or "cold-potato" routing, where it carries the packet on its own network as far as possible before handing off to the next ISP. The former policy minimizes the burden on the ISPs network whereas the latter policy gives the ISP greater control over the end-to-end quality of service experienced by packets. Geographic information can be used to quantify the degree to which an ISP's routing policy resembles hot-potato routing or cold-potato routing.
- **Shared network infrastructure:** There have been a number of efforts aimed at discovering the topological structure of the Internet [1], [2], [15], [17], typically using the *traceroute* tool [4] to discover the IP addresses of intermediate routers along an end-to-end path. However, such topology information may be incomplete in that two seemingly independent routers may actually be sharing networking infrastructure (and hence may vulnerable to correlated failures). For instance: (a) multiple "routers" may correspond to separate interfaces on a single router, (b) a set of routers may be located in the same data center belonging to an ISP, so they may share resources such as long-haul links to other locations, or (c) the routers may be located in the same city (whether in the same ISP network or not), so they may be vulnerable to correlated failures should a natural disaster such as an earthquake strike the city. The threat of simultaneous failure is likely to be greatest in case (a) and least in case (c). In any case, it may be useful, from a network fault tolerance viewpoint, to know that the routers are co-located and may share some net-

work resources. So geographic information would serve as a useful complement to pure network topology information.

In this paper, we study these issues using extensive traceroute data gathered from 20 hosts distributed across the U.S. and Europe (and also historic data gathered by Paxson [16]). We examine the circuitousness of end-to-end paths in the Internet (Section IV), the properties of paths that traverse multiple ISP networks (Section V), and the extent of sharing between paths at the geographic level (Section VI).

We would like to clarify that network performance is *not* the focus of this paper. (We are investigating performance issues in ongoing work, as indicated in Section VII.) For instance, we do not claim that a non-circuitous route is more optimal in terms of performance than a circuitous one. Rather, we present geography as an interesting tool for studying network properties.

II. RELATED WORK

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 IP addresses to the corresponding geographic locations. A few Internet mapping projects have used such tools to incorporate some notion of geographic location in their maps. However, we are not aware of any previous study that uses geographic information to analyze Internet routing as we do in this paper.

The Mercator project [2] focusses 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 heuristics for resolving *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 [1] at Bell Labs also uses a traceroute-based approach to map the Internet from a single source. It 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 Skitter project [17] and the commercial Matrix.Net service [15]. A number of tools have been developed for determining the geographic location corresponding to an IP address. These tools use a variety approaches to map from IP address to location: inferring location from *Whois* records [3] (e.g., NetGeo [5]), extracting location information from traceroute data (e.g., VisualRoute [18], GeoTrack [6]), determining the location coordinates using delay measurements (e.g., GeoPing [6]), etc. In our work, we use the GeoTrack tool, which we describe in some detail in Section III-D.

Besides Internet topology discovery and mapping, network path information (as obtained using traceroute) has been used to study the dynamics of network routing. For instance, Paxson [7] studied the temporal stability of Internet routing using an extensive set of traceroute data. In contrast, in this paper we study a different set of (static) properties of Internet routing using traceroute data coupled with geographic information.

III. EXPERIMENTAL METHODOLOGY

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

A. 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. 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. We use traceroute to determine the network path between 19 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 when possible. GeoTrack reports the location at the granularity of a city. We then use an online latitude-longitude server [9] 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 enabled 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,

---

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

and C & D. [2]

As we discuss in Section III-D.1, we are typically able to determine the location of most but not all routers. We simply skip over routers whose location we are unable to determine. So if in the above example the location of C is unknown, then we compute the linearized distance of the path from A to D is computed as the sum of the geographic distances between A & B and B & D. Clearly, skipping over C results in an estimate of linearized distance that is smaller than the true value. However, as noted in Section III-D.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.

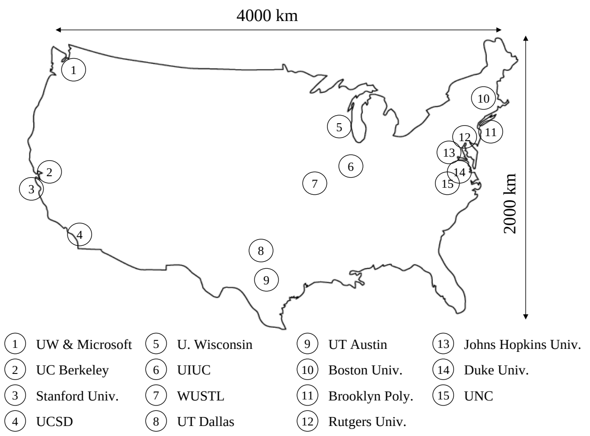

Fig. 1. Locations of our traceroute sources in the U.S. Note that there were two hosts each in Seattle and Berkeley.

B. 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 enables us to study the variations in topological and routing properties as seen from different vantage points. For logistical reasons, it was most 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 [10]. To add some diversity, we had one source in Berkeley, CA connected to a home cable modem network (in addition to one at University of California at Berkeley) and another in Seattle, WA connected to the Microsoft Research network (in addition to the one at the University of Washington at Seattle). These two sources allow us to study (albeit to a limited extent[3]) what impact, if any, the nature of the the source's connectivity has.

We used a set comprising several thousand hosts as the destination set for the traceroutes. 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:* 1205 Web servers of public libraries [12] distributed across 49 states in the U.S.
3. *TVHosts:* 3100 client hosts in the U.S. that connected to an online TV program guide. A majority of these clients were located on non-academic networks such as America Online (AOL).
4. *EuroWeb:* 1092 Web servers [14] 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 84000 end-to-end paths between our traceroute sources and the destination hosts during October-December 2000.

C. Dataset from 1995

To study the temporal variations in Internet properties, we use the traceroute data set collected by Paxson in 1995 [16]. 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. Since we only gathered data from the U.S. and Europe in 2000, we only considered the subset of the 1995 data set that corresponded to the U.S. and Europe.

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.

D. GeoTrack

Once we have gathered traceroute data, we use the GeoTrack tool [6] that we developed previously 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 of-

---

[2] Note that the linearized distance is computed only using nodes, such as routers, that are visible at the IP level. We have no way of discovering the underlying physical links (which may be based on SONET, ATM, or other technologies), so we do not consider them in our computation.

[3] We could have used a diverse set of public traceroute servers [13] to overcome this limitation. However, the large volume of traceroutes that we were looking to run from each source precluded this.

ten assign geographically meaningful names to routers[4], 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 [6] 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* 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.

D.1 Coverage of GeoTrack

Of the 11296 router names in our traceroute data set, 7842 were recognizable (approximately 70%). We compiled a list of 16 major ISPs with nationwide backbones in the U.S. or with international coverage (AT&T, Cable & Wireless, SprintLink, UUNET, etc.) and found that 5966 of the 6859 router names for these ISPs were recognizable (87%). In some individual cases, such as AT&T and UUNET, the recognizability was in excess of 95%. The unrecognizable router names tend to be concentrated in regional or campus networks. (For example, *cmu.psc.net* is a node at the Pittsburgh Supercomputing Center in Pittsburgh, PA. However, since it does not contain a valid city or airport code, GeoTrack is unable to recognize its location.[5]) Such nodes are typically located in the vicinity of the source or the destination of the end-to-end paths that we traced, so the resulting error in linearized distance is minimal.

In the case of the 1995 data set, GeoTrack is able to recognize 1289 out of 1531 router names (approximately 84%).

D.2 Computing Distances

To compute the geographic distance between two locations, we first consulted an online atlas server to obtain the corresponding latitude and longitude values. We then used simple trignometry to compute the distance between the two locations (which is basically the length of the great circle arc connecting the two points).

Having discussed our experimental testbed and methodology, we now turn to the analysis of the data we gathered. We present our findings and discuss their significance.

IV. CIRCUITOUSNESS OF INTERNET PATHS

In this section, we examine how circuitous Internet paths are. 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 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.[6]

We present several different analyses with a view to studying the impact of spatial factors (i.e., the location of the source and destination of the traceroutes) as well as temporal factors (i.e., 1995 versus 2000). Specifically, we present the distribution of the distance ratio for: (a) paths from a single location to different sets of destinations (UnivHosts, TVHosts), (b) paths from hosts in the same location but on entirely different networks, (c) paths from geographically distributed hosts in the U.S., (d) paths that lie entirely within the the U.S. versus paths that lie entirely within Europe, and (e) paths drawn from Paxson's 1995 data set and ones drawn from our 2000 data set.

A. Paths from a Single Source

We consider paths from our traceroute source at UC Berkeley to UnivHosts and to TVHosts. Many of the hosts in UnivHosts (our traceroute source at UC Berkeley among them) connect to the Internet2 high-speed backbone via a local GigaPOP. So much of the wide-area path between the UC Berkeley host 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 the UC Berkeley host to the hosts in TVHosts typically traverse one or more commercial ISP backbones.

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 2 shows, 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

---

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

[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.

[6] Clearly, the ratio can be no smaller than 1 unless network links pass *through* the earth, which they do not!

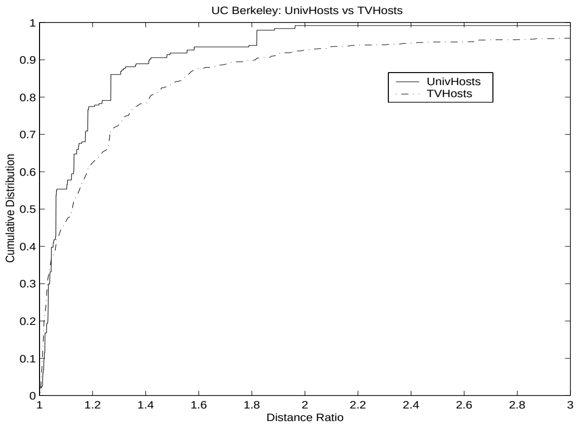

Fig. 2. CDF of distance ratio for paths from UC Berkeley to UnivHosts and TVHosts.

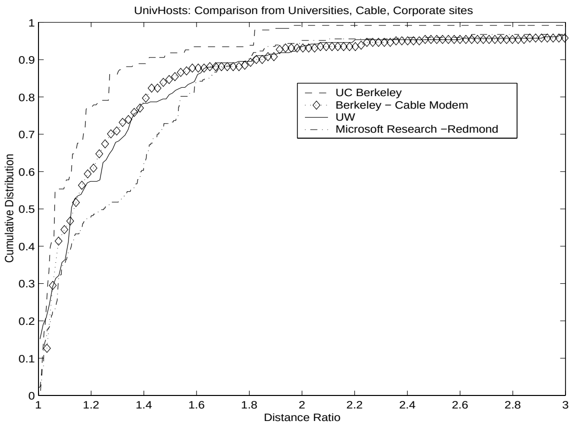

Fig. 3. CDF of distance ratio for paths from pairs of co-located sources to UnivHosts.

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 overhead due to 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 traverses multiple commercial ISPs that have a sub-optimal peering relationship. These trends are qualitatively the same for traceroute source locations other than UC Berkeley.

B. Paths from 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 3 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 such as BBNPlanet (now called Genuity)).

We observe a similar trend for 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 the BBNPlanet network. The path goes all the way down the U.S. west coast to Los Angeles, then across to Carlton, TX, then back up to Indianapolis, and finally to Chicago. 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 the source (and destination) network connectivity has a significant impact on how direct or circuitous the network paths are.

C. Paths from Multiple Sources in Different Locations

Next, 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 IV-B.

As shown in Figure 4, 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

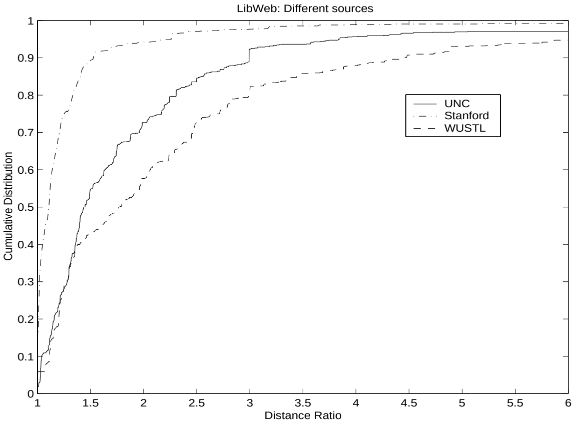

Fig. 4. CDF of distance ratio for paths from multiple sources to LibWeb.

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 V.

Our findings in this section suggest that the distribution of the distance ratio is indicative of the richness of connectivity of a source.

D. 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).

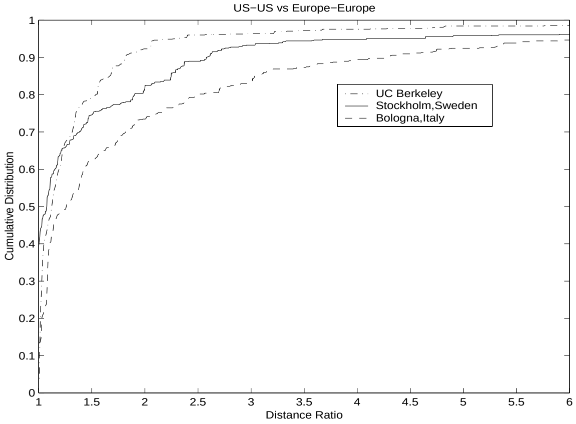

Fig. 5. CDF of distance ratio for paths within the U.S. and those within Europe.

In Figure 5, 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 European sources compared to Berkeley, especially in the tail of the distribution. We attribute this to two causes.

First, paths in Europe are 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. We have heard anecdotal evidence (but do not have any concrete data) of similar detours through the U.S. on paths between countries in east and southeast Asia. Given that a large fraction of Internet communication at these locations tends to be with the U.S. (because, for instance, much of the Web content is located in the U.S.), there is sometimes better connectivity to the U.S. than between these remote locations themselves.

E. 1995 versus 2000

Finally, we compare the distribution of the distance ratio computed from our 2000 data set with that computed from Paxson's 1995 data set [11]. 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. 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.

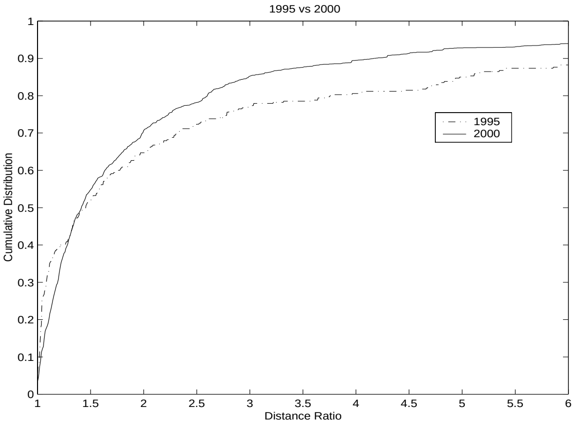

Fig. 6. CDF of distance ratio for paths in Paxson's 1995 data set and our data set from 2000.

Figure 6 plots the CDF of the distance ratio for the 1995 and 2000 data sets. We note 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.

F. Summary

In summary, our analysis in this section has shown that distance ratio is a useful metric for quantifying the quality of connectivity for a source or a path. A large value of the distance ratio enables us to automatically flag paths that are highly circuitous, possibly because of routing anomalies.

V. IMPACT OF MULTIPLE ISPS

Our analysis in Section IV 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 tool 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?

A. Circuitousness of End-to-End Paths versus Intra-ISP Paths

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.

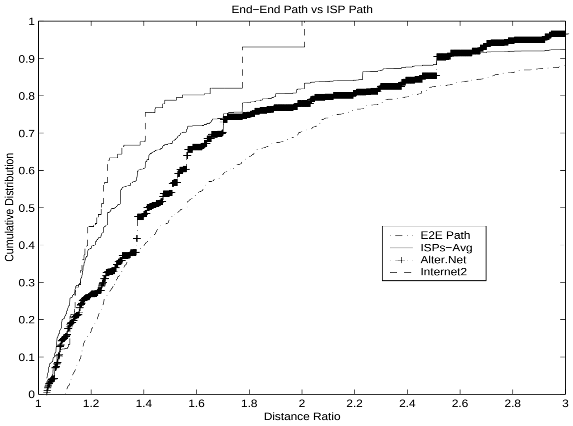

Fig. 7. CDF of distance ratio of end-to-end paths versus that of sections of the path that lie within individual ISP networks.

As shown in Figure 7, 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 because of sub-optimal peering between large ISPs. Inter-domain routing in the Internet largely uses the BGP [8] protocol. BGP is a path vector protocol that operates at the level of ASes. It offers little visibility into the internal structure of an AS's network (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. Another issue is that ISPs employ complex policies that controls which traffic enters their network and at which ingress points. In the example mentioned in Section IV-C, packets need to travel all the way from St. Louis to New York to enter AT&T's network. For these reasons, 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, is used for intra-domain routing. Since the internal topology of the ISP network is typically 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 IV-C, both the St. Louis $\rightarrow$ Chicago $\rightarrow$ New York path within Verio's network and the New York $\rightarrow$ Chicago $\rightarrow$ St. Louis path within AT&T's network are much less circuitous than the end-to-end path.

However, this is not mean that intra-ISP paths are never circuitous. As noted in Section IV-B, 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.

B. Impact of Multiple ISPs on Circuitousness

In Section V-A we hypothesized that the presence of multiple ISPs (with sub-optimal peering between them) 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 $max1$ and $max2$. 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 $max1 = 0.4$ and $max2 = 0.3$. Note that it is possible for $max1$ to be 1.0 (and so $max2$ to be 0.0) if the entire end-to-end path traverses just one ISP network[8].

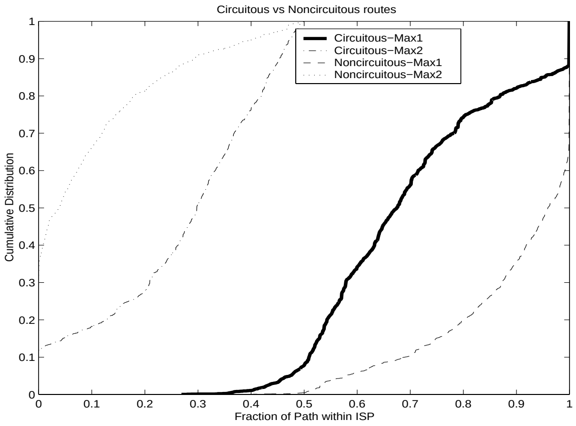

Fig. 8. CDF of the fraction of the end-to-end path that lies within the top 2 ISPs in the case of circuitous paths and non-circuitous paths.

Figure 8 shows the CDF of $max1$ and $max2$ for the circuitous and non-circuitous paths. The difference in the characteristics of these two categories of paths is striking. The $max1$ and $max2$ 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 both 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 $max1$ and $max2$ 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.

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.

C. 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

---

[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.

[8] Just to clarify, "one ISP network" means a single (usually wide-area) network that traverses a long distance. Local networks confined to a city (e.g., a university network) contribute nil to the linearized distance and therefore are ignored.

individual ISPs. We wish to understand how the effort of carrying traffic end-to-end over a wide-area path is apportioned across different ISPs. For this reason, we only consider a set of 13 large ISPs with nationwide coverage in the U.S. The 13 ISPs we considered are Alter.net, Sprintlink, AT&T, Cable and Wireless, Internet2, Verio, BBNPlanet, Qwest, Level3, Exodus, UUNet, VBNS and Global Crossing. For ease of exposition, we term these as the *major* ISPs. For each major ISP, 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.

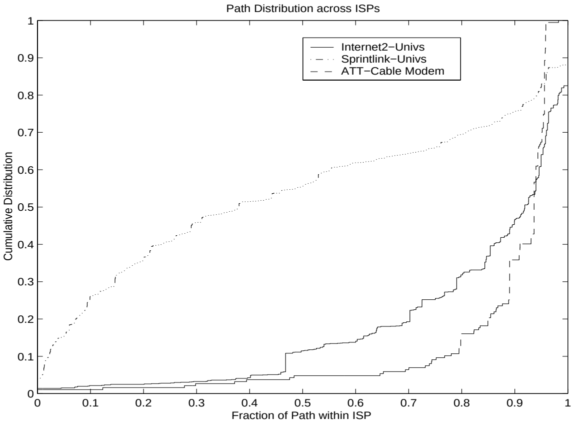

Fig. 9. CDF of the fraction of the end-to-end path that lies within individual ISP networks.

Figure 9 plots the CDF of this fraction for different ISPs. For the purposes of exposition, in the case of Internet2 and Sprintlink, we consider the paths from four geographically distributed university source hosts in the U.S to the LibWeb data set. Our university sources are UCSD, JHU, UT Dallas and Duke. In the case of AT&T, we only consider paths originating at our @Home cable modem host in Berkeley. 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 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 the end-to-end path.

Similarly, the large fraction in the case of AT&T can be attributed to the fact that we only considered paths originating at the cable modem host on @Home's network. Since @Home is owned, in part, by AT&T, it is not surprising that the AT&T backbone accounts for a substantial fraction of the end-to-end path.

The much smaller fraction in the case of Sprintlink is harder to explain definitively. It is likely the result of the specific peering relationships that Sprintlink has with other major ISPs. Next, we discuss hot-potato routing, which might offer a possible explanation. We stress, however, that we are not in a position to make a definitive determination.

D. "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.

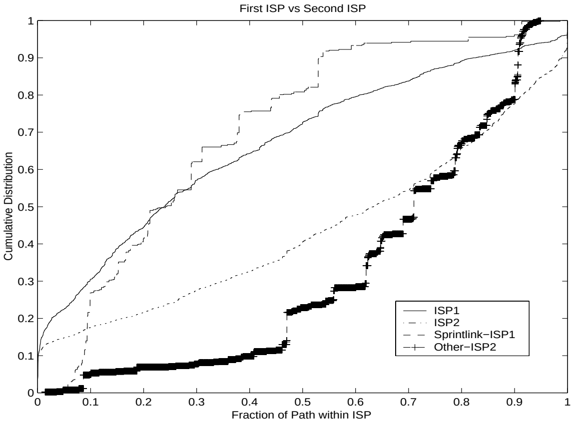

Fig. 10. CDF of the fraction of the end-to-end path that lies within the first and second ISP networks in sequence.

We consider the set of paths from U.S. sources to TVHosts. For each path that traverses two or more major ISPs, we compute the fraction of the end-to-end path that lies within the first ISP (ISP1) and the second 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 10. 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 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 10 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. Note, however, that we can only make indirect inferences based on the data we have. We cannot establish with certainty whether Sprintlink (or any other ISP) does, in fact, employ a hot-potato routing.

E. 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 sub-optimal peering between ISP networks. Furthermore, paths that traverse substantial distances within two or more ISPs tend to be more circuitous than paths that largely traverse a single ISP. 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.

VI. PATH SHARING

In this section, we use geographic information to study the characteristics of sharing between network paths. As outlined in Section I, this investigation is motivated by network fault tolerance considerations. A failure on the shared section of two paths has the potential of adversely impacting both paths.

We consider paths from a pair of hosts (such as traceroute sources) and a common destination. We define three notions of sharing. The basic one, which we term as *Shared-IP*, refers to sharing at the IP level. The Shared-IP path between two paths is the portion that lies between the first common router in both paths (as determined by matching IP addresses) and the destination. This corresponds to the notion of sharing used in most networking studies.

However, as we noted in Section I, Shared-IP does not necessarily capture the true notion of sharing between two paths. Two seemingly separate routers may actually be sharing network resources. The two “routers” may correspond to different interfaces on the same router. Or they may be located in the same data center or in the same city. So the two routers (and the network paths that traverse them) may be vulnerable to correlated failures.

It is hard to determine what, if any, resources are shared between two nodes without detailed knowledge of the actual network deployment. We define two additional notions of sharing (based on geographic information) as approximations of the ideal. The first, termed *Shared-ISPGeo*, refers to sharing at the ISP and geographic levels. Two routers are deemed to be shared if they are on the same ISP network and are located in the same city. The intuition is that such routers are likely to be located in the same data center and/or share resources such as long-haul links. The second, termed *Shared-Geo*, refers to sharing purely at the geographic level. Two routers are deemed to be shared if they are located in the same city, whether on the same ISP network or not. Should the city be struck by a disaster such as an earthquake or a cyclone, these Shared-Geo routers will be vulnerable to coordinated failures.

In our analysis of shared paths, we consider two different pairs of sources: UC Berkeley-Stanford and UCSD-UIUC. UC Berkeley and Stanford share a common connection to the Internet2 backbone via the Calren2 regional network. We would expect a significant amount of sharing among paths originating at UC Berkeley and Stanford given the similarity in their network connectivities and their physical proximity. On the other hand, UIUC and UCSD are located far apart. Moreover, they have very different network connectivities. UIUC has direct connectivity to BBNplanet, Verio, and Internet2 while UCSD has connectivity to CERFnet and Calren2. In many of our traceroutes, we found that UCSD connects to Internet2 indirectly via Calren2 at Los Angeles, CA.

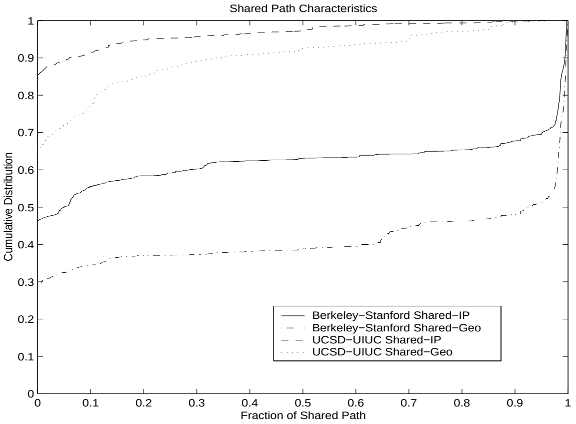

Fig. 11. Shared Path Characteristics: Berkeley-Stanford vs UCSD-UIUC

For both these pairs of sources, we consider paths to common destinations drawn from the LibWeb data set. For each destination, we compute the linearized distance of the Shared-IP path, the Shared-ISPGeo path, and the Shared-Geo path. We then express these as a fraction of the end-to-end linearized distance of each path. Essentially for every destination, we compute two fractions - one from each source. Figure 11 shows the cumulative distribution of the length of the shared path (expressed as a fraction) for the Shared-IP and Shared-Geo cases corresponding to the Berkeley-Stanford and UIUC-UCSD pairs. We note that for approximately 50% of the destinations, the paths from Berkeley and Stanford have a Shared-Geo path that extends over more than 95% of the end-to-end path from either source. However, only 28% of these destinations paths have a Shared-IP path that extends over more than 95% of the end-to-end path from either source. This indicates that there is much greater sharing between paths at the geographic level than at the IP level. The characteristics of Shared-ISPGeo path are very similar to those of the Shared-Geo path for both pairs of sources.

We also note in Figure 11 that the sharing between paths originating from UCSD and UIUC is very small. For more than 80% of the destinations, the Shared-IP is negligible in terms of linearized distance. The Shared-Geo path characteristics indicate that sharing is poor even in geographic terms for UCSD and UIUC.

VII. CONCLUSIONS

In this paper, we have presented geography as a tool for analyzing various aspects of Internet routing. Using our GeoTrack tool, we determine the geographic path between a pair of nodes. We use the distance ratio metric to quantify the circuitousness of the geographic path (and hence that of the underlying network path). We use extensive traceroute data gathered from 20 source nodes distributed across the U.S. and Europe for our analysis.

We find that the degree of circuitousness in paths varies depending on the connectivity of the source and destination nodes. Paths originating from well-connected hosts (e.g., those on the UC Berkeley or Stanford campuses) tend to have far less circuitous routes than hosts that are less well connected (e.g., those on the WUSTL campus). As a region, Europe tends to have more circuitous routes than the U.S. Also, in the U.S., routes measured in the year 2000 were less circuitous than those measured in 1995. We believe this positive change from 1995 to 2000 is the result of the Internet becoming more richly connected.

Paths that traverse substantial distances in multiple ISP networks tend to be more circuitous, presumably because of suboptimal peering between ISPs. 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 travers 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 hot-potato routing policy.

We also use geographic information to analyze sharing between paths at various granularities. We find that often sharing at the geographic level is far greater than sharing at the IP level.

In addition to the specific analyses that we have performed, we believe that a significant contribution of our work is to introduce the idea of using geography as a tool for analyzing network properties. As the Internet becomes more richly connected, we expect geography-based network analysis techniques to become more useful. For instance, the distance ratio can be used to automatically flag routes that are likely to be anomalous.

In ongoing work, we are studying the correlation between network performance and geographic location. In particular, we are investigating the effectiveness of geography-based schemes for wide-area replica selection.

ACKNOWLEDGEMENTS

Vern Paxson made the 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. We would like to thank them all.

REFERENCES

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

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

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

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

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

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

[7] V. Paxson. End-to-End Routing Behavior in the Internet. *IEEE/ACM Transactions on Networking*, Vol.5, No.5, pp. 601-615, October 1997.

[8] John W. Stewart BGP4: Inter-domain Routing in the Internet. Addison-Wesley Publishers, 1999.

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

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

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

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

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

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

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

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

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

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