Quantifying National Information Leakage

Daniel Wendlandt
Department of Computer Science
Stanford University
Stanford, Ca
danwent@cs.stanford.edu

Martin Casado
Information Operations and Analysis Center
Lawrence Livermore National Laboratory
Livermore, Ca
casado@cs.stanford.edu

***Abstract*— The Internet has been become a global communication medium that transcends national boundaries. However, few empirical studies have explored how this network without borders impacts a nation's ability to limit access and control over the information it entrusts to the Internet. In this paper we present our work addressing one facet of this issue: national information leakage. We explore to what degree information is routed outside of a country during communications between two internal networks. Our technique uses trace measurements from multiple vantage points within a nation to a diverse set of destinations within that same country. The traces are then mapped to physical locations using various IP-to-geographic location techniques to determine what percentage of point-to-point communication routes within a target country temporarily leave during transit to the final destination. We present measurements of sources within the United States, Mexico, Germany, Canada and Denmark.**

Our results suggest significant information leakage for many trace sources, including some located in regions with high Internet inter-connectivity. We show that while the behavior of sources can vary greatly within a country, national-level trends are apparent. Furthermore, the existence of leaking Internet paths is reliant not only on the geographic location of the source and destination end-points, but also depends highly on the peering behavior of the service providers offering access to these points. We believe our methodology is a promising approach to determining the degree of foreign access to a nation's internal Internet traffic and provides a basis for exploring the implications of entrusting security sensitive communications to the Internet.

I. Introduction

An increasing amount of security sensitive information is being entrusted to the Internet. For example, it is not uncommon for internal governmental, military, and industrial communications of operations critical data to be disseminated over email. However, the routing architecture of the Internet is largely agnostic to political boundaries, raising the possibility that point-to-point communications within an entity, such as a country, may be "leaked" to a third party during normal Internet operation. In this paper we study the susceptibility of communications within a country to viewing by foreign states. Specifically we explore the question - How much Internet communication between hosts in the same country is routed through another country? We define information leakage as an entity's loss of control over its own data as it is transfered on the Internet. While other interesting scenarios exist, such as third party access to communications between two countries, or the purposes of this paper we limit our focus to Internet paths that have both the source and destination within a single country.

A. Risks of Information Leakage

There are significant security concerns regarding foreign access to national traffic. Passive listening enables sensitive communication identification and interception with possible targets including internal governmental and corporate communications. Everyday Internet traffic contains information that can be used to gain unauthorized access to personal user accounts, provide leverage in a directed scam or facilitate identity theft. Furthermore, passive analysis of network headers allows fairly complete reconstruction of network infrastructure topologies which can aid in targeting. End-to-end encryption, while helpful for securing sensitive data is still vulnerable to event analysis which can reveal critical Internet resources, communicating end points, and communication patterns.

The risks are compounded when access is directly along the path of data-flow (or "on-route"), allowing for redirection or modification of traffic. On-route access to traffic can be exploited to compromise information integrity enabling the malicious destruction or modification of data, including executable code. Also, this level of access can make source spoofing practical, allowing a malicious node the ability to assume the identity of any destination host whose route goes through it. Source spoofing can be used to take advantage of trust relations by assuming the identity of a member inside the trusted group. Even end-to-end security solutions such as public/private key encryption may be subject to man in the middle attacks, although this is largely mitigated through an established certificate authority with a widely known and distributed public key.

Even if a third party with access to the traffic on-route does not have malicious intentions, information leakage can be a crucial concern because of jurisdictional control. Nations are increasingly using local governance and policy initiatives to dictate admissible content and behavior on the Internet. Similarly, different nations have developed separate laws governing the monitoring of data transfered over the Internet. However, as is commonly understood by the research community and demonstrated in our work, effective Internet governance cannot easily partition the Internet along national borders. This issue is particularly relevant with regard to export control restrictions, content filtering and censorship.

B. Overview

The goal of our work is threefold:

- to develop and analyze a methodology for estimating how much of a country's internal communications are leaked to

a third party.
- to quantify the degree to which information is leaked for a limited number of target countries.
- to use this data to explore the root causes of these geographically circuitous routes.

Within this paper we present a methodology for determining route information leakage. Our approach makes use of publicly available traceroute servers as sources for trace probes to destinations selected from advertised BGP prefixes. We use a well-regarded commercially available IP-to-geographic location software package in addition to the parsing of DNS names to determine the geographic locations of hosts returned from our traces. All traces and geographic information are stored in a database and post-processed in aggregate to determine leakage properties.

We present results estimating the percentage of routes that are leaked from sources with the United States, Mexico, Canada, Denmark and Germany. Measurements were taken over a two month period and involved over 100,000 traces. Our data results are compelling not only in that they provide the first quantifiable results on this area, but also because they display significant leakage, even for sources within highly connected regions. For example, we show that nearly all routes from two source within Mexico passed through the United States before returning to Mexico. Furthermore, a significant number of sources, even in areas of rich connectivity like Toronto or Frankfurt, demonstrate leakage of over half of the measured routes. While recent work indicated that geographic location is a strong factor in the circuitousness of routes[13], our results suggest that the peering behavior exhibited by the upstream provider for the source and destination of a path is a more significant consideration than geography in determining whether the route is likely to leak information.

The main contributions of this paper are:

- to describe a methodology and representative implementation for estimating the percentage of routes that are leaked from each IP address in a set of sources to destinations within the same country.
- to provide experimental results that quantify the route leakage between source and destination IP addresses within five target countries: the United States, Mexico, Canada, Germany and Denmark.
- to provide new insights into the root causes of circuitous Internet paths that result in leakage.

The remainder of this paper is organized as follows. In section 2 we provide relevant background. The methodology we used to determine route leakage information is presented in section 3. Section 4 presents the results of our measurements followed by a discussion of the results in section 5. We then evaluate several portions of our methodology in section 6. Section 7 covers related work, followed by a discussion of future research directions in section 8. Finally, we present our conclusions in section 9.

II. Background

In this section we provide a brief overview of the composition of the Internet and describe some of the high-level factors influencing route paths.

A. Autonomous System Connectivity

The Internet is comprised of approximately 18,000 individual networks, each known as an autonomous system (AS) [18]. Typically, each AS represents a unique economic entity, such as an Internet Service Provider (ISP), a large company or a university. These ASs interconnect in order to form the global Internet.

ASs are often divided into different classes depending on the nature of their connection to other autonomous systems. The most simple of these classification schemes partitions ASs into three network sub-types: stub, multi-homed, and transit. Most networks are either stub or multi-homed, which are simply variations on the same theme. Stub networks are end-point networks that are connected to a single other AS, known as the upstream Internet Service Provider (ISP). This provider is responsible for offering the single link of connectivity between the stub network and the rest of the Internet. As a result, stub networks do not carry any traffic except that which is either created by or addressed to a host on that network. Larger or more critical end-point networks may choose to have more than one provider, making them multi-homed.

ISP transit networks make up fewer than 10% of all autonomous systems [15], yet such networks are the critical networks that make the Internet a globally connected entity. Stub and multi-homed networks with limited reachability are the customers of these larger networks and pay transit providers to carry traffic to otherwise unreachable networks.

Transit networks themselves vary in size, from small regional providers operating solely within a single metropolitan area to large backbone providers with intercontinental networks. Transit networks are unique in that in addition to carrying traffic for hosts within that AS, the network carries packets destined to and from smaller networks who are customers of the transit provider. A smaller service provider may in turn be the customer of a much larger service provider, which transitively extends the reach of the customers of the smaller ISP. This economic customer-provider relationship is fundamental to understanding the nature of routes in today's Internet.

B. Inter-domain Routing

The path of autonomous system networks that a packet must follow to reach a given destination is determined using an exterior gateway protocol (EGP), and is tightly coupled to the economic relationships among autonomous systems. The de-facto EGP on the Internet today is the fourth version of the Border Gateway Protocol (BGP). BGP facilitates the propagation of reachability information to provide transit networks with the ability to route packets toward their destination network. Central to the protocol is the notion of an advertised route, in which an autonomous system informs its neighbors that it will accept traffic destined to a particular block of IP addresses, known as a network block or network prefix. Since networks do not wish to overload their transit links, providers advertise only the prefixes of their customers who are paying for the transit service. As a result, highly connected backbone networks will advertise paths to the many smaller stub networks for which it provides transit.

Yet customer-provider connections alone are not enough to create global connectivity, since a single provider cannot reach all possible destinations. In addition to customer-provider relationships, networks participate in peering relationships. The term peering can take on a variety of related connotations within network connectivity discussions, but in this paper we use it to refer to an agreement between two ASs to exchange traffic without a fee. When two ASs connect at a peering point, they exchange only traffic destined to prefixes advertised by the other AS, which is referred as a peer. Since no customer-provider relationship exists, BGP prefix advertisements learned from peering connecections are not forwarded to other autonomous systems. Peering provides the crucial link between the upstream providers of the source and the upstream provider of the destination. The portion of the path before peering is known as the upstream portion, the peering location as the peak, and the later half of the path as the downstream portion. Traditionally, peering occurrs at well known Internet Exchanges (IX) points created specifically for ISPs with routers in a given city to connect to one another and trade traffic. However, in recent years it has become increasingly common for larger ISPs to form private peering relationships in addition to or in replacement of those formed at IXs.

C. Classifying Autonomous Systems

The largest backbone ISPs make up a highly connected group of transit networks known as the dense core. The exact make-up of this group is not strictly defined, but for our purposes we can assume it contains 20-30 of the largest ISPs whose peering relationships form a near-clique[15]. This collection of transit ISPs is a highly connected mesh that ensures connectivity among the many regional transit networks. As a result, if the two smaller networks do not peer directly, the traffic often flows to networks in the dense core where it will almost certainly be able to peer. Companies within the dense core include Level3, UUnet/MCI, Verio, Sprint, Global Crossings, Cable & Wireless, France Telecom, Teleglobe, Telia and others[15]. These companies are referred to as tier-1 or top-level service providers because they often only trade traffic in peering arrangements and thus do not have an upstream provider. Such companies serve mainly as upstream providers of smaller transit providers, but also provide direct Internet access for some small individual customer networks.

III. Methodology

In this section we describe our approach for measuring the frequency with which routes are leaked by a sources during communication with a set of destinations in the same country. We use aggregate measurements from multiple sources within a country to estimate the percentage of routes that transit through a foreign country.

Our process can be broken down into five steps: choosing route sources, choosing route destinations, discovering the routes using network probes, utilizing IP-to-geo-location software for each routing hop, and validating the resulting path data.

A. Route Source Selection

Our work relied on two primary types of sources, Scriptroute servers and public, web-based traceroute servers. Scriptroute is a collection of dispersed computers around the globe, many belonging to the PlanetLab project, set aside specifically for network measurement[14]. These machines are hosted largely at academic sites. For our measurements, Scriptroute accounted for all American sources, 5 of 11 Canadian sources and 1 of 9 Danish sources. No Scriptroute servers were used in either Germany or Mexico.

In the US and Canada educational networks are commonly limited to a single university that is multi-homed with a commercial service provider with an educational provider, like Internet2, used only for routes to other academic or research institutions. In Europe educational institutions are often singly-homed on a national research network that ties into a larger European research network. When necessary we diversified our pool of sources in a country by making use of publicly available, web-based traceroute servers. Most of the public servers we used were posted on traceroute.org[2] or found through simple web searches.

In some countries the number of available sources was limited and we utilized every source we could identify that both worked with automated queries and represented a unique vantage point within the set of sources for that nation. Within countries that offered many potential sources, we limited the number chosen to allow for the collection and validation of all trace-paths in a timely manner. When many diverse Scriptroute vantage points were available, such sources where preferred because of their uniform trace interface and output format. When copious sources where available, we selected them in a largely random fashion, our only guideline being to create a set with diversity of source network types, ranging from tier-1 networks to small local service providers. We did so in the hopes of recognizing any leakage patterns that varied among source type. We did not preview the behavior of sources to determine what type of leakage we might expect from them.

While trace studies to map Internet connectivity topology have been shown to require few sources[19], efforts such as ours that seek to understand the many paths of information flow on the Internet require a large number of vantage points to be effective. This is because each source AS, and potentially even different locations within a source AS, could produce a different geographic pattern for information flow. As a result, any aggregate data for an entire nation is inherently tied to the sources that were selected. For this reason we do not feel it would be valid to provide a single quantitative value describing a nation's information leakage, favoring instead a graphical approach that allows the recognition of general national-level trends.

B. Route Destination Selection

Classifying properties of routes from a set of sources to destinations within the same country requires the identification of suitable IP addresses to serve as traceroute targets. To provide an adequate understanding of the level of information leakage, the networks containing these addresses need to be diverse enough to be representative of the many different service

providers within the country - since routing paths are based on logical AS-level connectivity - yet small enough that the resulting routes can be probed and analyzed in a reasonable timeframe.

Our selection method combined the use of BGP information with IP-to-geo-location provided by third-party Quova GeoPoint software[4]. We first obtained BGP data from an Oregon Routeviews [23] server for the arbitrarily selected date of 7/1/2004. The Routeviews server at the Oregon IX maintains multi-hop eBGP sessions with 44 peering BGP-speakers located in large transit networks dispersed across the globe. This data contained over 98,686 unique network blocks. Because of aggregation, we cannot be sure how well the nearly 100,000 unique prefixes obtained from BGP cover the actual routable address space, but we can ascertain that a significant amount of route specificity is lost either by missing or aggregated prefixes as core routers normally have close to 150,000 prefixes [18].

Since network blocks represent segments of IP space that are individually routed, they provide a means of identifying plausibly different geographic endpoints within a country. As a result, we chose to have the first valid IP address in each prefix represent a possible trace endpoint with equal probability. It should be noted that since there are more routes to ASs that announce more prefixes, the resulting data displaying a percentage of routes leaked will be weighed more heavily by destinations in such autonomous systems. However, such a design choice was logical given that our initial goals seek only to understand information leakage and not to provide actual data on the amount of traffic leaked (see 3.F).

We queried GeoPoint for each of the advertised BGP prefixes, assigning the network blocks to a candidate pool of potential trace destinations for the country returned by geo-location. Our implementation removes networks with less than 256 IP addresses to reduce the impact of ASs that advertise many small prefixes. If the number of destination networks was too large to be handled by our tracing infrastructure in a timely manner, we selected a subset of the total prefixes. To do this, we first estimated the maximum number of traces our infrastructure and validation process could support in a reasonable time-frame, considering our limitations of public traceroute servers and the human validation of leaked routes. We then ordered the prefixes within the candidate pool numerically by IP address, and selected every Nth entry, such that the value of N yielded our desired number of destination addresses. This method guaranteed us diversity among destinations within the IP address space.

There also existed the potential for false geo-location of end-host networks, especially those located near a border and within an IP block from a network provider in a different country. Since IP-to-geo-location partially relies on Internet registry information to determine the country of a network, it is possible that the software may have placed such a network in the wrong country and thus eliminated a potential destination that may have been highly likely to leak data. While we have no means of quantifying such error we have found the geo-location of end-host networks to be highly accurate with GeoPoint and do not believe that such cases introduced significant error.

C. Discovering Route Paths

Traceroutes, run either with Scriptroute or public servers, provided the hop-level routing paths used in our analysis. Using traceroutes for path discovery is an imperfect mechanism for several well-documented reasons [7]. The use of public traceroute servers greatly limited the packet types available for tracing, normally only running a standard Unix traceroute on input limited to the IP address. While this does not allow us to evade the common edge routers practice of blocking default incoming UDP traceroute packets, we did not find this to be a major concern because it was highly unlikely that leakage would occur on the extreme network edge. As a result, for the purposes of consistency we used UDP traceroutes even from Scriptroute servers where we had greater flexibility. The handling of traces which appear to fail is described in section 3.D.

In the case that multiple IP addresses were returned for a single hop, we arbitrarily chose the first one to represent the hop in our trace-path. In practice we found that differing addresses for the same hop only rarely represented a significantly different geographic route. Another recognized caveat of traceroute path discovery is that the tool only provides visibility at the IP level. Therefore our approach can incorrectly label a route as not-leaked if it transits through a country occurs completely at layer-two. Finally, due to rate-limits considered to be courteous for public servers, the collection of paths from a single source spanned 2-5 days. We, however, do not consider route drift to be a significant concern since our study simply attempted to gain an overall understanding of leakage from a source and therefore did require results from within a precise window of time. All traces presented in this paper where run during the weeks of August 9th through August 30th 2004.

D. Geo-Location of Path Hops

Each IP address hop within a completed trace-path, along with the source and destination addresses, was assigned a physical city and country value by the IP-to-geo-location software. While the GeoPoint software is highly effective for identification of end-host network locations, limitations inherent to geo-location make estimating the location a single backbone router difficult. For this reason, our primary method for determining router location was a specialized IP-to-geo-location software package from Quova that is designed specifically for determining the location of routing infrastructure.

Our secondary method to assign physical locations to routers was to parse the DNS names of the routers. While automated means exist to parse this data in Scriptroute, this tool does not cover all of the networks we mapped during our measurements and thus could not provide locations for routers in some AS networks and those without DNS names . The hops which could not be resolved by either of the two methods above were flagged as such and assigned locations based on the value returned by GeoPoint.

E. Validation of Traces and Geo-location

Our goal in validating paths was to accurately identify leaked routes in an efficient manner in order to handle a large number of traces. Keeping with our approach of tolerating some

false negatives for the sake of efficiency while going to great lengths to eliminate false positives, we required human confirmation of all trace-paths containing at least one apparently leaking hop. Validation consisted mainly of confirming whether IP addresses that were not located by either of the two highly reliable methods above, where in fact correctly assigned locations by GeoPoint. The existence of meaningful DNS names aided this effort, as did the use of hop latency information when considered in light of the other already located hops in the trace-path. While human validation of traces greatly limited the total number of probes we could use for the study, we found it necessary in order for us to achieve a high degree of confidence in presenting result that clearly demonstrate the prevalence of information leakage.

Only paths with at least one out of country hop were validated by a human, leaving the possibility for paths containing a false negative when a hop left the country but was not identified as foreign by our two router specific IP-to-geo-location methods or by GeoPoint. While this resulted in some routes not being identified as leaking, our experience demonstrated that false negative paths from geo-location are uncommon. This is because mis-located core routers are most often falsely identified as being located in the home country of the infrastructure owner, which rarely was the same country we were taking measurements in. The notable exception of this fact is the United States. Additionally, in practice, most leakage is not a single hop, but rather several foreign hops consisting of more than one provider. Thus, the likelihood of all hops being falsely identified as belonging to the originating country was further reduced.

In this process we also had to deal with traces that did not represent routable networks in the desired country. If the prefixes were actually located in another nation, such traces would have been first identified as leaking before being purged from the data-set by the validation process. However, failed traces that do not leave the country are errantly counted as legitimate non-leaking paths since they are not inspected by our validation process. This contributed to a lower overall percentage of paths leaked for our sources, yet the impact of this effect should be minimal because our destinations are largely routable because of their selection from real BGP data.

F. Data Aggregation

Once all traces had been labeled as either *leaking* or *non-leaking* we simply calculated and reported the percentage of traces marked as leaking out of the final set of traces. In doing so, we calculated the percent of routes leaked from a country, not the amount of data leaked. Understanding the amount of data leaked would require a notion of how much data is traversing each route we identify, which is significantly more difficult. Furthermore, it is unclear how valuable such information would be, since the danger posed by leaked information is likely to be more dependant on the content of the data rather than on the actual number of bytes transferred.

IV. Experimental Results

Using the methodology outlined above, we explore information leakage within five different countries that offer diversity in both the economic and geographic factors that define their Internet routing topologies. We chose Denmark, Mexico, Germany, Canada, and the United States. We do not claim that these nations are representative of all Internet architectures around the globe, rather we selected them in the hope that their differing geo-political qualities and the resulting network infrastructure would offer interesting comparisons concerning information leakage. For example, Europe is comprised of many small and densely populated nations with strong economic ties resulting from the EU and with relatively equivalent Internet usage [5]. In comparison, the three North American nations, the United States, Canada, and Mexico, are significantly larger and less densely populated. While Internet usage in Canada is on par with that of the United States, Mexico has achieved significantly less Internet penetration than its northern neighbor.

Fig. 1. Sources used within Denmark

A. Data Sets and Results

In this section we present the results of our measurements.

1) Denmark Data-set: Denmark has Internet usage at levels equivalent to the most developed countries in the world [5], yet its small size results in the majority of the connectivity among providers clustering at the centralized Danish Internet Exchange (DIX) near Copenhagen. Also of interest is the close proximity to extremely high volume routing hubs in Frankfurt, Amsterdam and to a lesser degree Stockholm.

The Denmark data-set includes 1,674 total routes after cleaning and verification, consisting of traces to all 186 identified and validated destination prefixes of sufficient length. The nine unique sources presented here represent all of the sources we were able to identify and query in Denmark.

2) Germany Data-set: In comparison to Denmark, Germany is both larger, implying greater isolation, and less centralized, with clusters of high connectivity located throughout the country. Germany also contains one of the largest traffic hubs in Europe located in Frankfurt.

Traces from 14 servers in Germany to 426 out of 2290 potential destination networks generated a total trace count of 5,964. The 14 sources in the data-set represent a diverse subset of the many potential sources we identified in Germany.

TABLE I
DENMARK SOURCES

Name	Location	Type	Leaked
Novo Nordisk IT	Copenhagen	Regional	6%
Tele DK	Copenhagen	Regional	12%
Univ. of Copenhagen	Copenhagen	Educational	12%
Cybercity	Copenhagen	Regional	14%
Tiscali	Copenhagen	Multi-National	26%
Telia	Copenhagen	Tier-1	45%

Fig. 2. Sources used within Germany

TABLE II
GERMANY SOURCES

Name	Location	Type	Leaked
SpaceNet	Munich	Regional	1%
Man DA	Darmstadt	Regional	3%
Velia	Frankfurt	Regional	4%
OSN	Nuremburg	Regional	5%
Trusted Networks	Unterschliessheim	Regional	5%
Probe Networks	Frankfurt	Regional	5%
Helios	Hannover	Regional	5%
Carrier66	Hamburg	Regional	5%
IPB	Berlin	Regional	6%
Level3	Duesseldorf	Tier-1	6%
IP Exchange	Nuremburg	Regional	6%
Telia	Frankfurt	Tier-1	11%
Tiscali	Hamburg	Multi-National	37%
France Telecom	Frankfurt	Tier-1	69%

3) Canada Data-Set: Canada has similar levels of Internet usage as the United States [5] and represents a significantly dispersed population neighboring the fiber-rich United States.

The Canada data-set consists of traces from 11 sources to 335 destination networks, out of a total of 3537 prefixes of sufficient size. The total trace count was 3,685 and the 11 sources represent the surprisingly low number of Canadian sources we were able to discover and query in an automated fashion.

4) Mexico Data-Set: Mexico is interesting due largely to the disparity of Internet penetration it has experienced compared to the neighboring United States [5] . Finding sources in Mexico proved difficult, and as a result our sample is somewhat incomplete and heavily weighted toward academic institutions. Yet the data collected still provides an valuable glimpse into how Mexico's traffic flows. We traced to 626 destinations, representing all identified and valid destination prefixes. Using 5 different sources, we collected a total of 3,130 traces.

5) United States Data-Set: The United States is a unique data-set due to its relatively high internal routing connectivity

TABLE III
CANADA SOURCES

Name	Location	Type	Leaked
Magma	Toronto	Regional	1 %
U. Victoria	Victoria	Educational	2 %
TRLabs	Winnipeg	Educational	3 %
Citenet	Montreal	Regional	18 %
TeraByte	Edmonton	Regional	25 %
Bellnexxia	Toronto	Regional	28 %
U. New Brunswick	Fredericton	Educational	37 %
U. Calgary	Calgary	Educational	40 %
Istop.com	Ottawa	Regional	41 %
U. Toronto	Toronto	Educational	59 %
Teleglobe	Burnaby	Tier-1	77 %

Fig. 3. Sources used within Canada

Fig. 4. Sources used within Mexico

TABLE IV
MEXICO SOURCES

Name	Location	Type	Leaked
U. Nacional Autonoma	Mexico City	Educational	16%
UAM	Mexico City	Educational	17 %
U. of Guadalajara	Guadalajara	Educational	21 %
Telefonica	Monterrey	Regional	95 %

Fig. 5. Sources used within the United States

TABLE V
UNITED STATES SOURCES

Name	Location	Type	Leaked
Intel (AT&T)	Seattle	Tier-1	0 %
Purdue	West Lafayette	Educational	0 %
U. of Minnesota	Minneapolis	Educational	0 %
U. of Rochester	Rochester	Educational	0 %
Carnegie Mellon	Pittsburg	Educational	0 %
U. of Washington	Seattle	Educational	0 %
U. of Oregon	Eugene	Educational	0 %
MIT	Cambridge	Educational	0 %
Columbia	New York	Educational	0 %
U. of Texas	Austin	Educational	0 %
Cornell	Ithaca	Educational	0 %
UC-Berkeley	Berkeley	Educational	0 %
U. of Cincinnati	Cincinnati	Educational	0 %
Dartmouth	Hanover	Educational	0 %
UC-San Diego	San Diego	Educational	0 %
U. of Michigan	Ann Arbor	Educational	0 %
U. of Rochester	Rochester	Educational	0 %

when compared to other national backbones. We utilized 22 unique vantage points, located at universities or other research institutions hosting Scriptroute servers, allowing us to send network probes at a rate signficiantly higher than permitted by public traceroute servers. As mentioned in 3.A, the abundance of American educational sources does not represent a lack of source diversity because institutions use a commercial upstream provider to reach most destinations. Despite having 22 sources tracing to over 4,082 different destinations our methodology identified only a handful of leaking paths. While this amounted to 89,804 traces, the total number of prefixes of sufficient length assigned to the US in our database was 47,776, meaning we explored less than 1 in 10 possible prefixes. The lack of discovered leakage does not, of course, mean US information leakage is certainly at such a low level, since our methodology errors on the side of finding a lower bound. However, our data does suggest that the US is not subject to anywhere near the same levels of information leakage as the other countries we have studied.

V. DISCUSSION OF RESULTS

Within this section we seek to use the data presented in Section 4 to begin a discussion on the topic of national-level information leakage and to explore how the data demonstrates underlying architectural factors contributing to the leakage of individual routes.

A. Quantifying National Information Leakage

While the existence of routes leaving a nation in transit and then returning is not itself surprising given the global nature of the Internet infrastructure, there is little data showing aggregate percantages of routes that display this characteristic for particular countries.

Tables 1-5 demonstrate that the percentage of routes leaked by sources can vary drastically, falling nearly anywhere between 0% and 100%. While significant variation exists even within a single country, the aggregation of data from all five countries in figure 6 demonstrates that significant national-level trends do exist. The United States exhibits essentially no information leakage capable of being measured by our methodology, with only a handful of routes leaking out of almost 90,000 traces. However, information leakage in the other four countries was more than we had anticipated. While Germany demonstrates better control over traffic than the remainder of the countries, even there only 65% of destinations experienced a leakage rate of less than 1 in 10. Yet the tail of the Germany data converges to nearly 100% more quickly than that of Mexico, Canada or Denmark. This demonstrates that overall Germany has relatively strong local connectivity. The other three countries have largely similar leakage levels when variation due to source selection is considered. These countries exhibit leakage levels such that approximately 20% of destinations experienced leakage of at least half of their routes. The overall connectivity of Mexico can be contrasted with the other nations in that is has a higher number of destination that leak all or almost all paths.

Utilizing the results we now look to explore the factors leading to a nation's overall leakage potential as displayed in Figure 6.

Such an approach requires an examination of geo-political considerations: not only the physical size and location of the territory but also the economic and technological factors that determine the overall structure and level of connectivity for networks within a nation. A first observation based on the data is that the landmass of a nation or a consideration of the size of its borders does not provide strong evidence for predicting information leakage. The United States has massive borders, but leaks significantly less than all other countries. Additionally, the two smaller European countries display highly contrasting

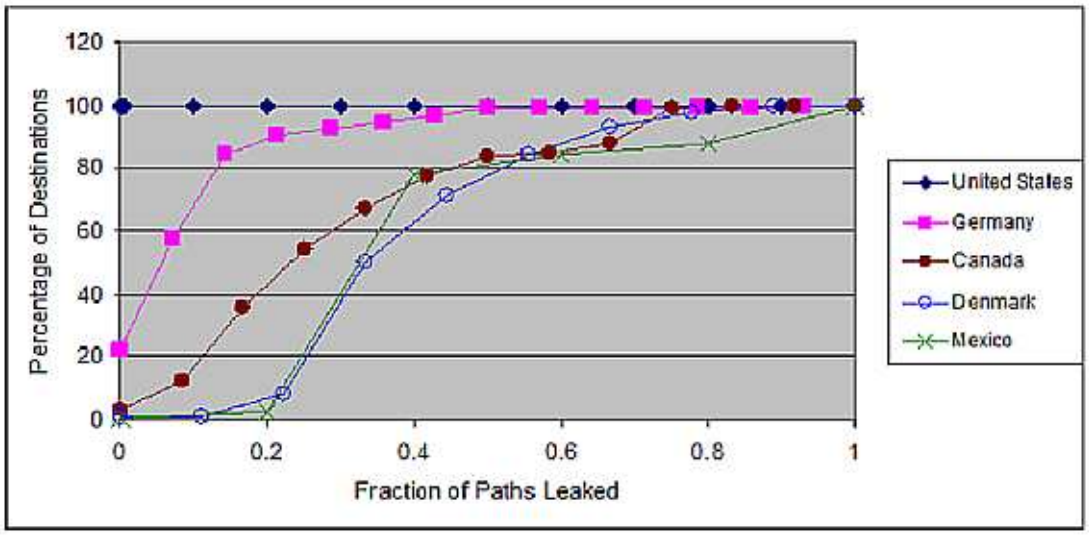

Fig. 6. Per-Nation Cumulative Distribution for Fractions of Paths Leaked by Destinations

results, demonstrating that size is not a primary determining factor.

Our analysis of the data clearly shows that routes cannot be predicted based simply on geography and any predictions must be made with a strong understanding of the network connectivity existing in the region. This is a result of the fact that Internet routing decisions are based on logical connectivity and are made without the explicit consideration of any type of geographic data. For a router, a path showing that a destination is one AS-hop away through an intermediate network on another continent is a preferred over a path that is two AS-hops but never leaves the building.

Our results clearly demonstrate that knowledge of the network connectivity trumped geographic considerations for national-level leakage, with the primary consideration being the existence of hubs of routing connectivity. A routing hub is a city or region that for historic or economic reasons has become a point where tier-1 and tier-2 service providers, in addition to regional service providers, meet to share traffic. These hubs often act as magnets for traffic because the strong concentration of routing infrastructure from major providers and presence of high capacity links means that traffic is often drawn to these locations for either as part of a transit hop or more often for the purposes of peering to reach another providers network.

These major routing hubs impact information leakage in two complementary ways. A nearby domestic hub often makes a path less circuitous, while the presence of such hubs in a neighboring country may draw traffic outside national borders in order to peer. Our findings in this regard are inline with work indicating that most path inflation occurs because of the need of traffic to travel to and from a peering point[22]. Not all routing hubs, however, are created equal and the largest hubs can exert influence over a wide area that includes many smaller local hubs.

Such results are clear when analyzing the results from all five countries. For example, Germanys overall limited amount of information leakage can be attributed to the fact that most paths are able to peer at Frankfurt, which provides extremely rich connectivity. Conversely, Denmark is negatively affected by the lack of strong inter-connectivity of major providers in Copenhagen, leading to traffic frequently leaking to Frankfurt, Amsterdam, Stockholm and even London to peer. Canada likewise shows frequent leakage to northern US hubs such as Seattle, Chicago and New York City, while Mexico leaks significant traffic to hubs in southern California, Texas and Florida. Because traffic flowing within a domain to a peering point is often routed more precisely because of the use of an interior gateway protocol, the geographic proximity of routing hubs is a major consideration and can be useful in predicting where information will leak to.

B. Leakage Potential of an Individual Network

While we are capable of understanding national trends in information leakage based on a geographic understanding of routing connectivity similar to that used in past work, we find such an approach wholly insufficient for predicting the leakage potential for an individual source or destination IP address. We argue that the ability to make such a prediction is important for two reasons. First, the need to understand information leakage for individual critical networks will be a necessary first step for limiting foreign access to the most important data transferred by government or corporate entities over the Internet. Second, understanding the details of information leakage at the level of individual networks leads to a better understanding of what factors, in aggregate, contribute to the national trends for leakage and how their effects can be mitigated.

For individual sources, our data demonstrates results that vary greatly from what would be expected based on a simple geographic network connectivity analysis. For example, from sources within the highly connected Frankfurt region we see vastly different leakage potentials of 4%, 5%, 11% and 69%. Similarly, the Toronto region includes sources as low as 1% and as high as 29% and 59%. These results demonstrate that relying on the nature of network connectivity within a region can prove dangerous when predicting leakage, since a source

within a highly connected area can still demonstrate significant leakage due to circuitous paths.

Our analysis revealed that the peering behavior of the upstream service providers of both the source and destination has the most significant impact in determining the likelihood of a path to transit another country in route to a domestic destination. Particularly, large tier-1 or multi-national service providers may fail to provide strong peering connectivity in the same nation as the source, leading to increased information leakage.

For example, France Telecom leaks 69% of routes from highly connected Frankfurt because it peers locally with only a limited number of providers, connecting to many of the other tier-1 providers via peerings largely in Amsterdam and Paris. Similar behavior is seen in Canada as the use of an upstream provider with significant peering and transport infrastructure in the US often draws traffic south to foriegn routing hubs. This exact pattern was frequently demonstrated in our data-sets as sources within Canada hosted by tier-1 provider Cogent Communications. Cogent rarely peered with sources in Canada, preferring instead to rely on its dense inter-connectivity in Chicago. The same logic can be applied to destinations, noting that networks advertised by providers that peer primarily in another country will often need to leak in order to reach this peering point. Destinations hosted by MCI/UUnet where strong examples of high leakage destinations in all countries analyzed except the United States.

While we demonstrate significance of upstream provider peering with only a few examples due to space considerations, the importance of this fact was uniform across all data-sets and allowed us to provide detailed case study explanations of the caues of leakage potentials for each source used.

One might wonder why large highly connected service providers tend to actually have poor local connectivity at most Points-of-Presence (POPs). An analysis of the business interests of tier-1 and other large transit providers, however, indicates that such companies have little incentive to peer with any provider except others in the dense cor of the Internet. This is because peering with smaller service providers provides no gain in overall reachability and in fact undermines a transit provider's business model by turning potential costumers into a profit-less peers. Additionally, since peering can be expensive both in infrastructure and complexity costs, large service providers do not peer with one another at every city in which they both have a POP. Instead they choose to connect only in major routing hubs, providing a likely imperceptible difference in delay to customers but potentially transporting data across great physical distances[12].

C. Stemming Information Leakage

The data demonstrates the existence of significant leakage due both to geographic network connectivity and the peering properties of the upstream providers for both the source and destination networks. We now move on to discuss what can be done to mitigate information leakage, both from the high-level perspective of a nation-state and additionally on the individual level of a single network.

1) Leakage Mitigation at a National Level: Our data contains both examples of a nation leaking almost no data and nations leaking significant amounts of data. It is reasonable to consider what the latter countries could do in order to emulate the former in hopes of reducing overall traffic exposure to foreign entities.

A commonality among nations with lower leakage, especially those of significant size, was the presence of upstream providers with significant Internet backbones that peered heavily in major hubs throughout the country. Such a network is the reason of the uniformly low results within the United States but can also be attributed to the low leakage from many sources in other countries. For example, Canadian sources such as U. of Victoria and TRLabs, relying primarily on Shaw Communications for upstream connectivity, experience significantly lower leakage because traffic often travels across Canada on the Shaw backbone directly to the destination region for peering. Mexico is an even more striking example of the importance of a strong backbone. The three university sources are connected to many other networks via a centralized national backbone, yielding leakage rates of 20%. This is starkly contrasted with the two other sources experiencing close to 100% leakage because they instead choose to latch on to a service provider in the United States.

It would be in the best interest of a nation looking to reduce information leakage to assure that providers with strong highly peered national backbone networks exist and reach all areas of significance population in the country.

2) Leakage Mitigation for Individual Networks: On a smaller scale, what can a single network location do in order to reduce the number of routes it leaks? As previously mentioned, having an upstream service provider with a strongly peered domestic backbone is a major benefit. However, if this is impossible or prohibitively expensive, a main goal is then to avoid the upstream provider as much as possible by peering traffic locally. The success of this approach is most obvious on a simple network topology like Denmark, with national connectivity focused at the centralized DIX location. Novo Nordisk IT (NNIT), the source network with significantly less leakage than other sources within the physical DIX location benefits greatly from 25 peering relationships with many of the other operators at the exchange [3]. Multi-homing, also used by NNIT, is often a beneficial method for trading traffic locally with several larger service providers who may otherwise not peer with a smaller source ISP for business reasons.

The consideration of a single source in Toronto provides a highly intriguing example of the benefits of peering and multi-homing. The Magma Communications network source is unique in that this service provider has specifically designed its network to avoid routes that pass through the United States [6]. While the cited goal of this design is performance and not security, this case gives an example of the degree of traffic control that is possible given a conscious effort on the part of the source, even in a hostile routing environment such as Toronto. It is of course important to point out, however, that in this work we only measured leakage in one direction and cannot speak to the nature of paths for return traffic.

Magma's strategy combines multi-homing and peering, uti-

lizing four major transit connections to MCI/UUNet, Bell-Nexxia, Primus, and Allstream, while also peering heavily at the Toronto and Ottawa IXs. Having multiple domestic upstream providers increases the chance of finding a very short logical path across a Canadian backbone directly to the destination.

VI. Methodology Evaluation

In this section we look at our original data-sets and newly collected data to evaluate two major considerations within our methodology: the selection of destination networks and the use of a single trace-source per autonomous system.

A. Evaluating Destination Network Selection

Due to our use of public trace infrastructure and the need for manual confirmation of leaked routes, the total number of traces was a limiting factor in this work. Likewise, our collection of destinations via aggregated BGP tables and our methodology of only sampling destination networks hides possible missing destination networks and we would like to estimate whether tracing to only a subset of all networks is sufficient for understanding information leakage.

For both of these reasons an important part of evaluating this methodology is understanding how many traces are required to achieve an accurate picture on what percentage of routes leak. Such an evaluation is necessary to demonstrate both that the number of traces we collected was sufficient and to explore the possibility of achieving results of the same quality with significantly reduced measurements.

First, we are interested in whether our current data-sets where large enough such that collecting additional traces from the same sources would not have significantly altered our results. To estimate this, we use our raw results to simulate the collection of data-sets of 3/4, 1/2, 1/4, and 1/8 times the size of the original and see whether the variation within these data-sets implies that the leakage value for a source was converging to a certain value. If this is the case, we can argue that additional traces would have been unlikely to provide benefit and in fact fewer traces may have been sufficient.

These results are displayed in tables 6-9. In these tables, the Average Difference column displays the average difference across all sources in that country between the original leakage values, and the values arrived at using the modified destination selection strategy. These results suggest that a reduction of our destination set size by 25% and even 50% does not have a substantial impact on the rates of information leakage we see. This suggests that a good estimation of the percentage of routes leaked can be arrived at using significantly fewer traces than our data-sets. We also feel that this fact largely mitigates concerns over the potential loss of more specific destinations resulting from limited BGP data or the selection of only a fraction of destinations networks for actual measurement.

We also look at a method for reducing the number of traces by leveraging the intuition that since routing paths are chosen using attributes of the AS-Path, many traces to the same AS will traverse nearly identical paths and are likely to exhibit the same leakage behavior.

TABLE VI
Denmark Different Destination Selection Strategies

Reduction Method	Traces Eliminated	Average Difference
3/4 Set	46	1.0 %
1/2 Set	93	1.6 %
1/4 Set	139	3.0 %
1/8 Set	162	3.1 %
Scaled Trace	132	1.2 %
Single Trace	132	4.6 %

TABLE VII
Germany Different Destination Selection Strategies

Reduction Method	Traces Eliminated	Average Difference
3/4 Set	105	0.3%
1/2 Set	210	0.6%
1/4 Set	315	1.1%
1/8 Set	367	4.1%
Scaled Single Trace	203	0.3 %
Single Trace	203	1.8%

In analyzing the potential value of a methodology using only one probe per AS, we utilize the full data-set to simulate the alternate method by simply choosing a single random trace per destination AS and calculating the results. Destination AS values are included as part of the data returned by GeoPoint. However, such results actually represent a qualitatively different measurement from our methodology described in Section 4. These ”single trace” results will vary from our original data for two reasons. First, because some autonomous systems (particularly those that are multi-homed) contained destinations with both leaking and non-leaking routes from a single source, randomly selecting one represents a loss in data accuracy. Secondly, because an AS with many prefixes in the original data-set is now represented by only a single endpoint, all destinations ASs in the single probe results are weighted equally unlike in our original results.

To isolate these two causes of information loss, we present a second set of scaled single probe results that restores the weights present in the original data-set. As a result, the difference between these scaled values and the original data results will be due solely to the accuracy loss describe above. Again, the results are again present in tables 6-9.

The extremely low variation seen in the scaled results implies that there is significant promise in a trace-reduction technique that sends one probe per AS and then scales its results to estimate multiple probes to different destinations within that autonomous system.

B. Multiple Vantage Points per AS

We also ran further trace experiments to learn whether an additional source within the same autonomous system brings any benefit as a vantage point. This is useful for understanding whether a single source in our methodology is able to speak for

TABLE VIII
CANADA DIFFERENT DESTINATION SELECTION STRATEGIES

Reduction Method	Traces Eliminated	Average Difference
3/4 Set	82	0.7%
1/2 Set	165	1.7%
1/4 Set	247	2.0%
1/8 Set	289	5.0%
Scaled Single Trace	221	2.2%
Single Trace	221	5.1%

TABLE IX
MEXICO DIFFERENT DESTINATION SELECTION STRATEGIES

Reduction Method	Traces Eliminated	Average Difference
3/4 Set	166	0.8 %
1/2 Set	332	0.8 %
1/4 Set	464	2.1 %
1/8 Set	542	2.6 %
Scaled Single Trace	554	1.4 %
Single Trace	554	10.2 %

all sources within the autonomous system. For small ASs that peer and reach their upstream provider only in a single location, this point is essentially moot. However, the answer in the case of larger networks with multiple exit points is less clear. Even if the source AS choses a uniform next-hop AS when selecting a particular route, it is possible that one location for peering with this next-hop AS causes leakage and another does not. Our experiment looks at two major providers within Germany, Tiscali and Telia, and the Teleglobe network in Canada to provide an intuition on this topic. The results are shown in table 10.

In the case of the Telia source in Dusseldorf, almost all traffic that does not peer locally is sent directly to Frankfurt, leading to essentially identical measurements of information leakage between the two sources. Tiscali demonstrates a significantly different case. While additional network probes indicate that the Tiscali router is connected to Franfurt via a single intermediate hop at Dusseldorf, traces to the same destinations that peer locally for the Frankfurt source are often sent to Amsterdam in the case of the Hamburg source. Such behavior is potentially due to the use of traffic engineering either because of link capacity or an attempt to decrease the overall distance to the peering point. The Canadian Teleglobe sources show a significant but not expansive gap in leakage. Unlike the German sources, the two Teleglobe locations are separated by a significant physical distance and by our investigation are in fact not even connected by a Teleglobe backbone link through Canada. Because fewer destination networks are directly available for peering in Burnaby than in Montreal, the Burnaby source sees higher leakage.

TABLE X
LEAKAGE FROM MULTIPLE SOURCES WITHIN THE SAME AS AND COUNTRY

Provider	Location	Paths Leaked
Telia	Frankfurt, DE	11%
	Dusseldorf, DE	11%
Tiscali	Hamburg, DE	37%
	Frankfurt, DE	4%
Teleglobe	Montreal, CA	64%
	Burnaby, CA	77%

VII. RELATED WORK

Numerous research efforts have sought to understand the logical connectivity of the Internet using hop-limited probes or traceroutes [16], [20], [21], [1], [11], [8]. Rocketfuel is distinguished in this area for leveraging BGP information from RouterViews to optimize destination selection thus limiting the necessary number of traces while still creating highly detailed maps. Rocketfuel is also the first mapping project to make wide use of the PlanetLab infrastructure. Our work builds on the work done by Spring et al. by using Scriptroute, developed by the Rocketfuel team, for traces and selecting destination addresses from advertised BGP prefixes. Unlike efforts in Internet mapping, our work does not focus on logical connectivity but rather explores properties of routes mapped onto the geo-political world. While logical connectivity is important for understanding inter-domain routing between two points, it does not provide sufficient information to determine who has access to traffic on the network.

[22] and [9] study the causes of route inflation on the Internet from a logical perspective. Both studies found that peering policy has a significant influence on logical Internet path lengths. Similarly, our measurements suggest that the probability a route will be leaked is strongly affected by the peering agreements of the top level provider for the source or destination.

The work most closely resembling our own is that of Subramanian et al. [13] in which they study the geographic properties of routing. This work asserts that geographic properties are an interesting means for analyzing routes and their characteristics. In particular, they find that the circuitousness of Internet routes depends on the source network's geographic location and may be impacted by its connectivity provider. Conclusions were drawn from measurements taken from 17 hosts, all within the United States. The work explores circuitousness but it does so purely geographically and not in the context of access to traffic by differing political entities. We believe our work is a natural extension to the work of Subramanian et al. We generalize the approach to work for any target country given availability of public traceroute servers and introduce a methodology for determining appropriate destinations to trace. Our analysis looks specifically at the issue of information leakage, while Subramanian et al. focus on general path circuitousness, comparing the actual distance traveled by a path between two endhost to the point-to-point distance between them. Furthermore, our analysis focuses heavily on the underlying causes of circuitous routes, going into significant detail while considering both geographic network location but also the peering behavior of upstream providers.

VIII. Future Work

The introductory nature of this work opens the door for significant future exploration in the areas of methodology innovation, additional data-collection, and improved analysis.

Beyond the methodology improvements pointed to as a result of analysis in 6.A and 6.B, many possibilities exist for creating better measurements. Our current methodology used an ad hoc method for choosing sources, merely attempting to have a variety of network types and geographic locations represented. A systematic approach to acquiring sources could provide a more accurate picture of the percentage of routes leaked by a country.

Using BGP connectivity information to identify routes likely to leak may also holds potential as an area for further study. We used such a strategy informally while looking for paths that leaked from the United States and found it promising. For example, after identifying Peer1 Networks as a provider of one leaked path through Canada, we analyzed BGP tables to find other US prefixes for which Peer1 appeared in the AS-Path. The presence of this AS in the path for a prefix implied that Peer1 might also offer that network transit through Canada.

Additionally, our methodology takes a high-level approach, attempting to assess information leakage for an entire nation. From an assurance point of view, one may be more interested in understanding the leakage associated with a particular network. This would include not only the flow of information generated by the network, but also traffic flowing to the network and how these patterns change over time.

The scope of data collection is another area upon which future work can improve. We look at two well-developed portions of the world where we could easily acquire public trace sources. The vastly different geography and intriguing politics of Asia suggests considerable potential for further exploration. Locating sources in more remote corners of the world would give insight into whether the understanding of geo-routing dynamics presented in this paper is consistent world-wide or limited to the regions studied.

Analysis capabilities also show significant potential for growth, particularly as they relate to the ability to overcome the fundamental problem of limited vantage points. An unanswered question is whether there can be an effective and automated mechanism for gauging the likely leaking behavior of a source without actually having the ability to send packets from that source. For example, to what degree is the rate of leaked probes to a destination linked to that destination's ability to function as a non-leaking source of information? Multi-homing and other factors contributing to AS-path asymmetry complicate this issue significantly. Another approach could attempt to infer the peering behavior of an AS and its providers based on routing policy, allowing static analysis to infer the likelihood of information leakage. Unlike BGP, routing policy specified in router arbiter databases contains records of the peering behavior of stub ASs which is critical to predicting the flow of information.

Lastly, an important yet unanswered question is - What are the likely future trends for information leakage? Will the precentage of routes leaked increase as networks grow or will increasing connectivity converge on direct paths from all sources to destinations? We believe that insights in this area will help determine if there is a growing threat that will require attention in the future, or if information leakage is an artifact with diminishing relevance.

IX. Conclusion

The amount of Internet traffic leaked outside of a locale's border during its internal communications provides an indication of third party access to that data. Understanding this can be a crucial component of determining the threat environment of a nation's cyber-infrastructure. In this paper we present a methodology and preliminary results from our efforts to explore this problem. Our approach involves using Scriptroute and other publicly accessible traceroute servers in a target country to make traces to a set of destinations within the same country. Routes are then post-processed in aggregate to estimate the percent that are leaked.

We present measurements of over 100,000 traces taken over a two month period from sources in Mexico, the United States, Canada, Germany and Denmark. Our results indicate that most countries experience significant leakage, and that the probability of leaking is strongly linked to the peering behavior of the source and destination's upstream provider, rather than simply thier particular geographic locations. We conclude that exploring "where in the world" traffic flows rather than just its logical path is a compelling approach for determining who has access to Internet traffic between a source and destination and warrants further investigation.

X. Acknowledgements

References

[1] Skitter, http://www.caida.org/tools/measurement/skitter
[2] Traceroute.org, http://www.traceroute.org
[3] Private correspondance with Kristian Bjrnskov of Novo Nordisk IT. September 2004.
[4] Quova, http://www.quova.com
[5] Internet World Stats, http://www.internetworldstats.com/stats.htm
[6] Magma Communcations Limited, http://www1.magma.ca/aboutmagma/backbone.cfm
[7] Lisa Amini, Anees Shaikh and Henning Schulzrinne "Issues with Inferring Internet Topological Attributes", In *SPIE ITcom* , August 2002.
[8] Zhuoqing Morley Mao, Jeniffer Rexford, Jia Wang, and Randy Katz, "Toward an Accurate AS-Level Traceroute Tool" In *SIGCOMM*, August 2003.
[9] Hongsuda Tangmunarunkit, Ramesh Govindan, Scott Shenker, and Deborah Estrin, "The Impact of Routing Policy on Internet Paths" In *IEEE INFOCOM*, April 2001.
[10] Anukool Lakhina, John Byers, Mark Crovella, and Ibrahim Matta "On the Geographic Location of Internet Resources" *Technical Report BUCS-TR-2002-015, Boston University* , 2002.
[11] Vern Paxson "End-to-End Routing Behavior in the Internet" In *IEEE/ACM Transactions on Networking, Vol. 5 No. 5*, October 1997.
[12] William B. Norton "The Evolution of the U.S. Internet Peering Ecosystem" November, 2003.
[13] Lakshminarayanan Subramanian, Venkata N. Padmanabhan, and Randy Katz, "Geographic Properties of Internet Routing" In *USENIX Annual Technical Conference*, June 2002.
[14] Neil Spring, David Wetherall, and Thomas Anderson, "Scriptroute: A facility for distributed Internet measurement" In emphUSENIX Symposium on Internet Technologies and Systems, March 2003.
[15] Lakshminarayanan Subramanian, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz, "Characterizing the Internet Hierarchy from Multiple Vantage Points" In *IEEE INFOCOM*, June 2002.
[16] Neil Spring, Ratul Mahajan, and David Wetherall "Measuring ISP Topologies with Rocketfuel" In *SIGCOMM*, August 2002.
[17] Lixin Gao "On Inferring Autonomous System Relationships in the Internet" In *IEEE Global Internet*, November 2000.
[18] The CIDR-Report, http://www.cidr-report.org

[19] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the Marginal Utility of Network Topology Measurements. In *ACM SIGCOMM Internet Measurement Workshop*, November 2001
[20] H. Burch and B. Cheswick. Mapping the Internet. *IEEE Computer*, 32(4):97-98,102 1999
[21] Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for Internet Map Discovery, *Proc IEEE Infocom* 2000, Tel Aviv, Israel.
[22] Neil Spring, Ratul Mahajan, and Thomas Anderson ”Quantifying the Causes of Path Inflation” In *SIGCOMM*, August 2003
[23] D. Meyer, RouteViews Project, http://www.routeviews.org