

An AS-Level Study of Internet Path Delay Characteristics

Amgad Zeitoun

University of Michigan, Ann Arbor
azeitoun@eecs.umich.edu

Chen-Nee Chuah

University of California Davis
chuah@ece.ucdavis.edu

Supratik Bhattacharyya

Sprint ATL
supratik@sprintlabs.com

Christophe Diot

Intel Research Labs
christophe.diot@intel.com

Abstract— According to conventional wisdom, links connecting different Autonomous Systems (ASes) are the performance bottlenecks in the core of the Internet. This paper presents an empirical evaluation of delays across inter-AS links using hop-limited active probes. The measurements cover a diverse set of Internet paths starting from locations within three large transit Internet Service Providers (ISPs). We find that most inter-AS links on the Internet paths covered by this study do not contribute significantly to end-to-end delays. The few exceptions are long-haul links with large propagation delays. Furthermore, the delay estimates are fairly stable across days, making it possible for ISPs to choose inter-domain paths or perform traffic engineering based on delay measurement feedback. Our observations also suggest that a very large component of the end-to-end delay for the measured paths usually occurs within a single AS.

I. INTRODUCTION

The Internet consists of inter-connected Autonomous Systems (ASes), where each AS is administered by a single authority with its own choice of routing protocols, configuration, and policies. The Border Gateway Protocol (BGP) [1] is used to route data packets across multiple ASes. Hence, a packet traversing the Internet crosses a sequence of ASes connected by links, often referred to as “peering links.” These links are dimensioned based on the relationship between the connecting ASes. If the connecting ASes belong to the same AS tier [2] and have a “peer-to-peer” relationship, then the peering link capacity is negotiated based on how much traffic each expects to exchange with the other. If the ASes share a “customer-provider” relationship (typically a smaller AS paying the larger AS for transit to the rest of the Internet), then the peering link capacity depends on how much the customer is willing to pay for transit services. Alternatively, several ASes may peer with each other at public Internet exchange points. Given the variability in Internet traffic, these arrangements may result in underprovisioning of certain peering links.

In this work, we investigate how inter-AS links impact end-to-end delays in the Internet and whether there is a dominant delay component across a single inter-AS link. A thorough understanding of delay characteristics of Internet paths at the AS level has several benefits. ISP networks may use this information to decide when and how to upgrade links to other networks or to select among alternate AS-level paths to reach destination networks. Applications that build overlay networks [3] can take into account the magnitude and variation in delay across inter-AS links to improve wide-area routing performance. Finally, network topology generators can produce more realistic graphs for simulation studies of the Internet.

Our measurement methodology is based on hop-limited TCP-based probe packets from selected points within three large tier-1 ISPs to a large number of destinations in the Internet. We do not claim that we cover *all* paths on the Internet, but we strive to diversify our set of paths so that they cover different networks and continents. The probe paths cover 172 ASes in four continents and a wide variety of networks such as academic institutions, ISPs, and enterprise networks. Each set of measurements are collected across several days.

The delay estimates of inter-AS links consist of two major components—propagation delay and queueing delay due to congestion. In order to understand the contribution of propagation delay to inter-AS link delays, we categorize inter-AS links by the geographic distances they traverse. This is done by mapping both ends of each link to geographic locations using a variety of techniques such as NetGeo [4] and location hints from router names resolved by reverse DNS lookups. Congestion level on these links is examined by studying variations in the delay estimates.

Results show that most inter-AS links do not contribute significantly to end-to-end delays for the Internet paths covered in our study. The largest inter-AS delay component for 70% of the paths accounted for less than 25% of the end-to-end delay. There are exceptions—for 7% of the paths, there was a single inter-AS link along each path that accounted for 80% or more of the end-to-end delay. A closer inspection shows that most of these links span large geographic distances (e.g., continents or oceans), and therefore have large propagation delays. Furthermore, the delay estimates are fairly stable on the time-scale of days, suggesting that these links do not experience significant congestion levels.

Our results also seem to suggest that the largest component of the end-to-end delay often occurs within a single AS along the path. This finding is consistent with the prevalent practice of hot-potato routing [1], where each transit ISP hands off a packet to another ISP as soon as possible unless the packet is destined for one of its customers.

The rest of this paper is organized as follows. Section II discusses related work and highlights the unresolved challenges that our paper attempts to address. Section III describes our measurement methodology. Section IV presents the results of our data analysis. Finally, we conclude the paper in Section V.

II. RELATED WORK

Extensive literature exists on the subject of characterizing packet delay in the Internet. Most of the existing work, however, have focused on end-to-end delays [5–7], and none

has addressed the problem of delay measurements at the AS level.

The authors of [8] studied the characteristic of RTT, bandwidth, and losses on five paths from the US to hosts in Brazil, South Africa, and Bangladesh. They observed that packet suffers high queueing delays on international links, while bandwidth bottlenecks are usually located outside the US. The work, however, is limited to an extremely small number of Internet paths. More recently, authors of [9] have shown that bandwidth bottlenecks in the Internet are roughly equally split among intra-AS and inter-AS links. Our work differs from this study in two aspects. First, we study Internet paths across several continents. Second, we investigate how inter-AS link delays affect end-to-end delays, regardless of the bandwidth available on these links.

In [10], authors developed and evaluated techniques to infer the lossy links by passively monitoring traffic exchanged between server and clients. This study found that losses on the Internet are more likely to occur on inter-AS links that have large delays. However, it does not study the prevalence of such inter-AS links with high delays, which is the focus of our work.

In [8–10], delays on individual links were estimated using traceroute RTT measurements. However, they do not address the issue of asymmetry in Internet routes, which brings in question the accuracy of the RTT estimates. On the other hand, the methodology we develop for our study carefully addresses the issue of Internet path asymmetry.

III. METHODOLOGY

A basic technique in measuring delays on the AS-level requires identifying routers on the border of each AS along the path, and then estimating delay between border routers. Using active measurement technique, we can send hop-limited probes towards a given destination, identify border routers by mapping them to ASes, and estimate the delay between adjacent border routers. Fig. 1 illustrates how the delay across an inter-AS link is estimated. Consider two consecutive ASes AS_x and AS_y along the path from source S to destination D . Let R_x^e be the router by which the packet exits from AS_x , and R_y^i be the router through which it enters AS_y . First, a probe packet is sent from S towards D to terminate at router R_x^e . This is done by setting the time-to-live (TTL) field in the IP header of the packet to be equal to the number of router-level hops between S and R_x^e . This way the source can estimate the round-trip time to R_x^e . Next, S sends another probe packet towards D to terminate at R_y^i and obtains an estimate of the round-trip time to R_y^i . The difference between these two measurements yields an estimate of the RTT on the inter-AS link (R_x^e, R_y^i). Assuming that the link is symmetric, the one-way delay across the link is half of the difference between the two RTT measurements, i.e.,

$$\Delta_{x,y} = \frac{RTT_y^i - RTT_x^e}{2}, \quad (1)$$

where RTT_y^i is the *minimum* RTT to the *ingress* router in AS_y and RTT_x^e is the *minimum* RTT to the *egress* router in AS_x .

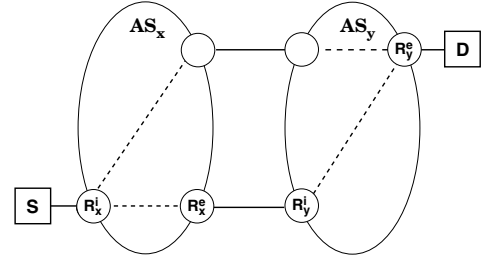


Fig. 1. Measuring AS-level delays.

Despite the simple approach outlined above in estimating AS-level delays, one needs to be careful about a set of challenges that may introduce inaccuracies in measurements. Next, we briefly present a set of challenges and our approach in addressing them.

Probe Type: Many routers discard UDP TTL-limited probes, hence, traditional `traceroute` tool cannot identify some routers along a path. Therefore, we use `TCPtraceroute` [11], a traceroute-like tool that uses TCP packets instead. The latency introduced by routers to generate reply messages to terminated TTL-limited probes is usually negligible [12].

Path Asymmetry: In practice, the Internet path may be asymmetric either at the router level or the AS level. That may happen due to multi-homing and hot-potato routing practiced by most service providers. In Fig. 1, AS_x and AS_y are multi-homed. Packets traveling from source (S) to destination (D) and from D back to S follow two different router-level paths. Each AS hands off a packet to the other AS as quickly as possible since the packet is destined for a network that is not a customer of the sender's AS. Obviously, an asymmetric AS-level path is also router-level asymmetric, but the reverse is not always true.

Our study only requires paths that are symmetric at the AS level, but not necessarily at the router level. This is sufficient to ensure the correctness of our inter-AS link delay estimation technique for all inter-AS links along the probe paths. In Fig. 1, for example, packets going from S to D that terminate at R_x^e and R_y^i traverse the same common portion of the forward path from S to R_x^e , because they are targeted to the same destination D . On the reverse path, R_y^i 's reply goes to R_x^e , due to hot-potato routing, and from there follows the same reverse path followed by the reply from R_x^e to S .

IP-to-AS Mapping: IP-to-AS mapping is critical due to the fact that routers are identified by the IP addresses received in their ICMP replies. Routers usually set the source IP address field in the ICMP reply packet to the IP address of the interface which the probe was received on [13]. There are two ways to map an IP address to an AS, using either BGP tables or Internet Routing Registries (IRR) [14]. The BGP approach suffers from the dynamic nature of BGP information while the IRR approach suffers from the presence of outdated and incomplete information in the IRR databases. We have designed a heuristic to combine both approaches to improve mapping accuracy (detailed description of our heuristics are presented in [15]). To ensure that IP addresses of border

TABLE I
DATASET SUMMARY.

Dataset	$\mathcal{D}1$	$\mathcal{D}2$	$\mathcal{D}3$	$\mathcal{D}4$	$\mathcal{D}5$
Probe source	Sprint	Sprint	Sprint	AT&T	Verio
AS num.	1239	1239	1239	7018	2914
Duration (days)	3	3	4	7	7
Paths	95	92	80	53	31
Num of ASes	116	114	99	70	58
Num of inter-AS links	153	155	132	115	73

routers belong to the address space of their corresponding ASes, we developed a heuristic based on practical router IP address assignment performed by major ISPs. We have also matched our mapping against Mao *et al.* [16].

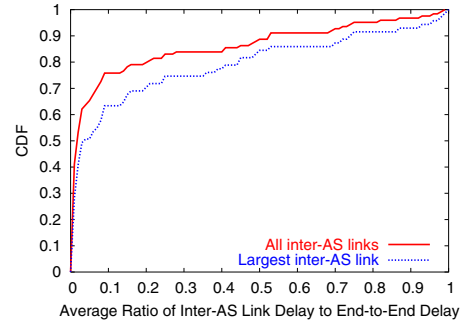
Once we identify Internet paths that are symmetric at the AS-level, we determine the ingress and egress routers at each AS along the path. For every router, 10 probes, each in a different train, are sent with the TTL field set to expire at each router. The set of 10 probe trains is called *probing period*. The time between consecutive probing periods is exponentially distributed with a mean of 15 minutes. This prevents biasness in our measurements and makes the measurement traffic reasonably low to prevent stressing routers and raising false security alarms.

IV. RESULTS

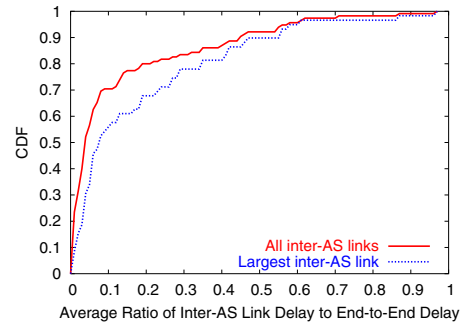
Five sets of data were collected for this study (Table I). The first three sets consist of Internet paths originating from Sprint (AS1239), while the remaining two consist of paths originating from AT&T (AS7018) and Verio (AS2914), respectively. The target destinations consist of 124 traceroute gateways and routers spread across 115 ASes in 22 countries. In all, a total of 103,662 RTT measurements across 405 unique inter-AS links have been collected.

A. Inter-AS Links Delays

We first examine the contribution of inter-AS link delays to the total path delay. The total path delay is defined as the delay from the source machine up to the ingress router of the destination's AS. For this purpose, we divide the inter-AS link delay by the total path delay for each measurement and compute the average of these ratios over all measurements for a given link. Fig. 2 shows the cumulative distribution of these delay ratios for data sets $\mathcal{D}3$ and $\mathcal{D}4$. Other data sets show similar characteristics and are omitted due to space limitations. We observe that the delay across the majority of inter-AS links is small relative to the total path delay. For example, 70% of the inter-AS links in $\mathcal{D}3$ and $\mathcal{D}4$ contribute to less than 25% of the total path delays. Fig. 2 also shows the cumulative distribution of the single largest inter-AS delay component along a path as a ratio of the total path delay. The results show that for 80% of the paths, the single largest inter-AS component contributes less than 48%, for $\mathcal{D}3$, and 40%, for $\mathcal{D}4$, of the total path delay. While some of these links may not have very large capacities [9], our results suggest that they are adequately provisioned in most cases (for the traffic that they are expected to carry) and do not contribute significantly to the measured end-to-end delay.

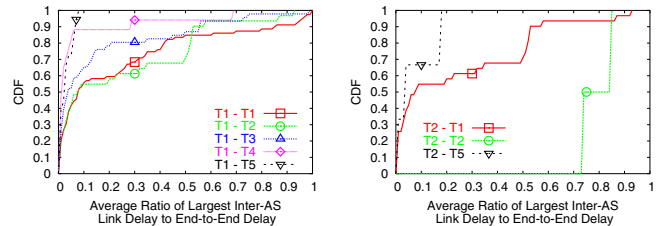


(a) $\mathcal{D}3$



(b) $\mathcal{D}4$

Fig. 2. Cumulative distribution of the ratio of inter-AS link delays to end-to-end delays, from (a) Sprint (AS1239) and (b) AT&T (AS7018).



(a) From Tier-1

(b) From Tier-2

Fig. 3. Cumulative distribution of the ratio of largest inter-AS link delays to end-to-end delays, classified by AS tiers.

Next, we examine the characteristics of inter-AS links relative to their locations in the inter-AS topology. We do so by classifying each inter-AS link by the tier of the two ASes it connects (e.g., tier-1). The AS tiers are determined using the methodology in [2] as tier-1 (T1), tier-2 (T2), etc. In this classification, the direction of the link is immaterial, e.g., T1-T2 is the same as T2-T1. Fig. 3 shows the cumulative distribution of the largest inter-AS delay component in each class as a ratio of the total path delay.

In general, most of the links in each class do not add significantly to the total path delay. However, delay on links from tier-1 ASes to tier-1/tier-2 ASes contribute more than those to tiers-3, 4, or 5. (Fig. 3 (a)). As we will show later, the majority of high delay links are between tier-1 and tier-1 or tier-2 ASes, which are sometimes geographically distant. Also, links between tier-2 ASes contribute significantly to end-to-end delays. Of the largest delay components between tier-2 ASes, almost all contributed to more than 70% of the total path delay. On closer investigation, these links were found to be trans-oceanic links between ASes in the US and Australia.

The results for inter-AS delay components from tier-3 or tier-4 to lower tiers are similar, and are omitted due to space limitations.

In all of the graphs discussed so far, there is a small percentage of inter-AS links that are responsible for a large portion of the path delay. For 12%–19% of the paths in Fig. 2, there is a single inter-AS link that accounts for more than 50% of the total path delay. This could be either due to propagation delay or congestion (leading to queueing delays). Since propagation delay increases with the physical length of a link, we now examine the geographic span traversed by inter-AS links with high delays.

First, we identify the subset of paths for which the largest inter-AS delay component accounts for at least 30% of the total path delay. For each such path we identify the link causing this delay. In order to avoid paths with short end-to-end delays, we eliminate all such links with a delay of less than 10 ms. We refer to the selected links as “high-delay” links.

Next, each router at both ends of a high-delay link is mapped to geographic location (i.e., latitude and longitude). To minimize mapping errors, we use a variety of techniques to map IP address to location, e.g., *NetGeo* [4], performing a reverse DNS lookup and using location hints from the router’s name, looking at the location of next-hop routers, etc. The high-delay links are classified into three categories based on the geographic distance between their end-points—less than 100 miles (in a metropolitan area), between 100 miles and 2,500 miles (within the US), and larger than 2,500 miles (inter-continental).

We observe that the geographic distance has a significant correlation with high delays—of the 45 high-delay links identified, 31 links traverse more than 2,500 miles, and 9 links traverse between 100 and 2,500 miles. The majority of the high-delay links, about 62%, are trans-oceanic. For the few cases where the link delay is high but the distance is small, we found that link characteristics induce large delays, e.g., DSL links. Some of these high-delay low-distance links connect low tier ASes.

In summary, we find that only a small fraction of the inter-AS links that we study contribute significantly to end-to-end delay. Most of these high-delay links span large physical distances, therefore have large propagation delays.

B. Inter-AS Delay Variation

To gain some insight into the fluctuations in congestion levels (if any) across inter-AS links, we examine the variability of inter-AS link delays. For a given link, we order all the probe periods (each consisting of 10 probe trains) by time and denote them by probe period $0, 1, 2, \dots, N$. Let $d(i)$ be the delay estimate for probe period i , computed using the approach in Section III. Then the variation in the delay estimate between successive probe periods i and $i + 1$ is computed as $\delta(i) = |d(i) - d(i + 1)|$.

Fig. 4 shows the cumulative distribution function for the absolute delay variation values, i.e., $\delta(i)$, for two sets of

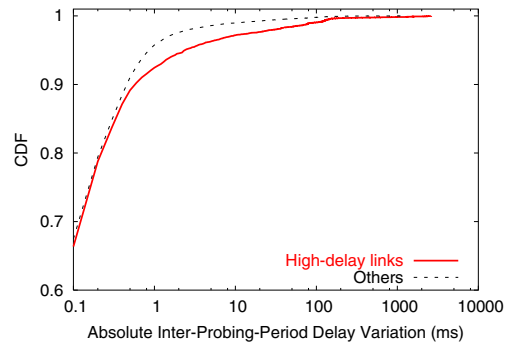


Fig. 4. Absolute probe delay variation across probe intervals.

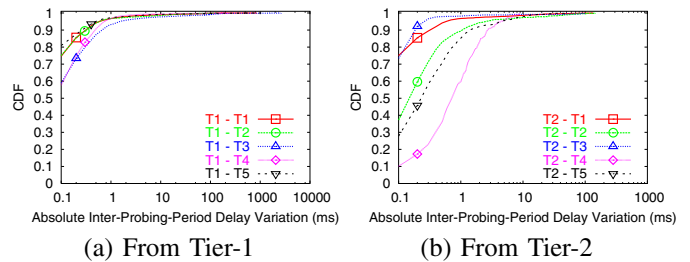


Fig. 5. Cumulative distribution of the absolute delay variation across probe intervals, classified by tiers.

links—the high-delay links (average delay more than 10 ms and accounting for at least 30% of the end-to-end delay) and the rest. From the figure, we can see that the delays are stable for the majority of links. Almost 67% of the delay variations are less than or equal to 0.1 ms, with the 99th percentile being 91 ms and 10 ms for the high-delay links and the rest, respectively. The low variability observed in inter-AS delays, in conjunction with the earlier observation that propagation delay contributes significantly to delays across high-delay links, suggests that the majority of links do not experience a lot of congestion over a long period of time.

We also classify inter-AS links based on the AS-tiers that they connect (Fig. 5). Results show that links that connect tier-1 ASes to other ASes are significantly stable. More than 55% of the T2-T2 inter-AS links are high-delay links (over 140 ms), and for 37% of these links, the variations are less than or equal to 0.1 ms. The variations of delays on T2-T4 and T2-T5 inter-AS links are more prominent, indicating that these links may be more prone to congestion. The likely reason for this is that the core of the Internet is better provisioned than the edges where tier-4 and tier-5 ASes are located. However, the absolute delay values on these links are fairly low—less than 13 ms.

C. Intra-AS Delays

The results presented so far show that most inter-AS links do not contribute significantly to the end-to-end delays on the Internet paths covered by this study. Therefore, the major delay components on most of these Internet paths must occur within ASes. This is not surprising, since a packet may cross several routers within an AS. Moreover, our probes originate from within large transit networks which deliver

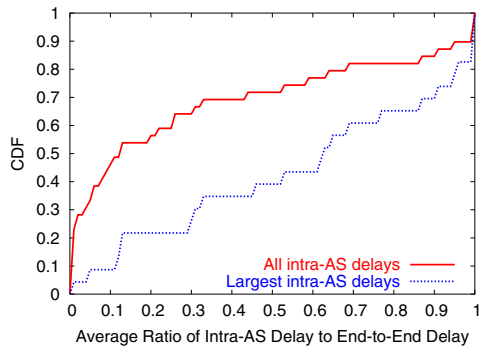


Fig. 6. Cumulative distribution of the ratio of intra-AS delays to end-to-end delays, for router-level symmetric paths.

packets across very long geographical distances. However, it is interesting to examine how the delay is distributed across multiple AS hops. Unfortunately, it is impossible to answer this question unless router-level symmetry in end-to-end paths is guaranteed. A majority of Internet paths are asymmetric at the router-level [17]. Even when the router-level path is symmetric, traceroute may not always be able to determine this path accurately [13].

Our definition of a router-level symmetry is strict. First, the number of hops traversed on the forward and backward paths must be equal. Second, the corresponding IP addresses at each hop should belong to the same /24 subnet. We identified 44 router-level symmetric paths in our study. For these paths, we estimated the delay across each AS along the path in a manner similar to the inter-AS delay estimation approach described in Section III. Probes are sent to the ingress and egress routers of each AS, and the difference between these values yields an estimate of the intra-AS delay. Fig. 6 shows the cumulative distribution of the average intra-AS delays as a ratio of the end-to-end delays along these 44 paths. We notice that packets tend to spend more time traveling through an AS. For example, in 64% of the cases, the time taken to cross an AS accounts for 30% of the end-to-end delay.

The figure also shows the cumulative distribution of the largest intra-AS component as a ratio of the end-to-end delay. Interestingly, in 60% of cases, the largest intra-AS component is more than 50% of the end-to-end delay. By classifying the intra-AS delay components by AS tiers, we find that the largest component along a path occurs in a tier-1 AS 70% of the time. This is a consequence of the choice of our probe initiation point within large transit ISPs—for many of the paths, this ISP carries the traffic most of the way toward the destination. After that the packet is rapidly handed off from one AS to another until it reaches the destination (hot-potato routing).

V. CONCLUSIONS AND FUTURE WORK

We have presented a study of inter-AS link delays across a variety of Internet paths, starting from three large transit backbones. We observed that the inter-AS links covered in our study generally do not contribute significantly to end-to-end delays. The few exceptions are usually links that traverse large physical distances and hence have high propagation delays.

Moreover, our delay estimates have low variability, implying that congestion on these links is not significant.

This paper is the starting point towards understanding where and why packets are delayed as they traverse multiple ASes across the Internet. The next step in this study will be to refine the measurement methodology in order to measure a wider and more representative set of inter-AS links. This may be achieved by relaxing the requirement of AS-level symmetry requirement for the Internet paths measured. In order to estimate the delay across a given inter-AS link, the only requirement is to find a probe initiation point such that the path from the probe initiation point up to the link in question is symmetric. Moreover probes need to be initiated from different types of ASes in order to increase the diversity of the data set.

It is often speculated that public exchange points are greater performance bottlenecks than private peering links. We plan to expand our data sets to include more paths through these exchange points to verify this.

VI. ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (NSF) under the CAREER Award No. 0238348. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] S. Halabi and D. McPherson, *Internet Routing Architectures*, 2nd ed. Cisco Press, 2000.
- [2] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, "Characterizing the Internet Hierarchy from Multiple Vantage Points," *Proc. of IEEE INFOCOM*, 2002.
- [3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A Scalable Content-Addressable Network," *Proc. of ACM SIGCOMM '01*, August 2001.
- [4] "NetGeo," <http://www.caida.org/tools/utilities/netgeo/>.
- [5] V. Paxson, "End-to-End Internet Packet Dynamics," *Proc. of ACM SIGCOMM '97*, Sep. 1997.
- [6] J.-C. Bolot, "Characterizing End-to-End Packet Delay and Loss in the Internet," *Proc. of ACM SIGCOMM '93*, pp. 289–298, Sep. 1993.
- [7] D. Sanghi, O. Gudmundson, K. Agrawala, and B. Jain, "Experimental Assessment of End-to-End Behavior on Internet," *Proc. of IEEE INFOCOM '93*, pp. 867–874, Mar. 1993.
- [8] M. Habib and M. Abrams, "Analysis of Bottlenecks in International Internet Links," *Proc. of International Conference of Computers and Information Technology*, Jan. 2001.
- [9] A. Akella, S. Seshan, and A. Shaikh, "An Empirical Evaluation of Wide-Area Internet Bottlenecks," IBM TJ Watson Research Center, Technical Report RC 22753, Mar. 2003.
- [10] V. Padmanabhan, L. Qiu, and H. Wang, "Server-based Inference of Internet Link Lossiness," *Proc. of IEEE INFOCOM '03*, Mar. 2003.
- [11] "TCPTraceroute," <http://michael.toren.net/code/tcptraceroute/>.
- [12] R. Govindan and V. Paxson, "Estimating Router ICMP Generation Delays," *Proc. of PAM '02*, 2002.
- [13] L. Amini, A. Shaikh, and H. Schulzrinne, "Issues with Inferring Internet Topological Attributes," *SPIE ITCOM 2002*, July 2002.
- [14] "The Merit Network. Internet routing registry database." <ftp://ftp.radb.net/radb/dbase/>.
- [15] A. Zeitoun, C. N. Chuah, S. Bhattacharyya, and C. Diot, "An AS-Level Study of Internet Path Delay Characteristics," Sprint ATL, Research Report RR03-ATL-051699, May 2003.
- [16] Z. Mao, J. Rexford, J. Wang, and R. Katz, "Towards an Accurate AS-Level Traceroute Tool," *Proc. of ACM SIGCOMM '03*, 2003.
- [17] V. Paxson, "End-to-End Routing Behavior in the Internet," *Proc. of ACM SIGCOMM '96*, pp. 25–38, Aug. 1996.