

Leveraging Correlation Between Capacity and Available Bandwidth to Scale Network Monitoring

Praveen Yalagandula, Sung-Ju Lee, Puneet Sharma and Sujata Banerjee
Hewlett-Packard Labs, Palo Alto, CA

Abstract—Recently, there has been a tremendous growth in the number of installed distributed computing platforms such as those for content distribution networks, cloud computing infrastructures, and distributed data centers. Such distributed platforms need a scalable end-to-end (e2e) network monitoring component to provide Quality of Service (QoS) guarantees to the services and improve the overall performance. An important challenge for a network monitoring infrastructure is the periodicity of the measurements as this aspect trades off the monitoring overheads with staleness of the results. In the *Network Genome* project, we explore the relationships between different e2e network metrics with the aim of leveraging such relationships for reducing monitoring costs while maintaining measurement accuracy.

We perform our analysis using long range network measurements from PlanetLab, where we have been collecting e2e network data (route, number of hops, capacity bandwidth and available bandwidth) as part of the S³ system since January 2006. In this paper, we focus on the correlation between the *Capacity* and *Available Bandwidth* metrics between host pairs in the PlanetLab testbed. Our analysis shows that the ranking of hosts with respect to their Capacity to/from a set of nodes is a good indicator of the ranking of hosts with respect to their Available Bandwidth to/from the same set of nodes.

I. INTRODUCTION

Content distribution systems (CDN, e.g., Akamai [1]), cloud computing infrastructures (e.g., Amazon EC2 [2]), and federated large-scale testbeds (e.g., PlanetLab [3], GENI [4]) are increasingly becoming popular. A scalable network monitoring capability is essential in these systems. For example, in a content distribution network, a node needs to not only determine the set of nodes holding a replica of a requested object but also determine the node from which it can download the replica in the shortest time. A scalable network monitoring tool that captures the dynamic state of the end-to-end (e2e) network paths in near-real time enables such replica selection.

An important question for any network monitoring system is the periodicity of the measurements. If all end-to-end metrics of interest are measured as frequently as possible on all paths, the monitoring system might consume significant network and end-host resources and also interfere with the other traffic [5]. On the other hand, if these metrics are measured at very low rates, the monitoring system might be unable to capture significant network change events in time to avert the performance degradation in the overlay services.

The overheads for measuring different metrics on a path vary drastically. Measuring latency and round-trip time on a path does not require significant resources (e.g., a small number of ICMP pings is enough for these metrics). However,

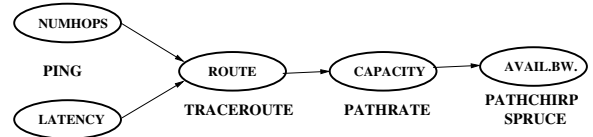


Fig. 1. Correlations between e2e network metrics explored in our project; also shown are the measurement tool(s) used for each metric.

tools for measuring metrics such as end-to-end path capacity and available bandwidth [6], [7], [8] have significant probing overheads - both packet overhead and delay (seconds to multiple minutes) to obtain a statistically significant estimate.

The dynamicity of different metrics on a path vary significantly as well. For example, metrics that depend on physical characteristics of a path such as propagation delay and capacity have a major change only when the underlying path has link or router changes. On the other hand, metrics such as available bandwidth depend on the cross traffic in the network and can vary significantly over short periods of time.

In the “Network Genome” project, our aim is to explore the relationships between various e2e network metrics with the intent of minimizing the monitoring overhead while maintaining high accuracy. We consider several types of correlations: (i) auto-correlation of a single metric on a path, (ii) auto-correlation of a single metric across different paths (over the entire network), (iii) cross-correlation between different metrics on a single path, and (iv) cross correlation between different metrics across the network. Past research has primarily focused on a small fraction of such relations. For example, latency inference techniques [9], [10], [11], [12], [13] assume auto-correlation in the latency metric across paths in a network and exploit that to reduce the monitoring costs. In contrast, our work focuses on quantifying the above correlations for several different metrics and leverage such correlations (if any) for designing an optimized monitoring system.

Since January 2006, we have been monitoring the e2e paths between PlanetLab machines via the S³ (Scalable Sensing Service) monitoring system [14]. We monitor several different metrics: latency, number of hops, route, capacity, and available bandwidth. In a previous paper [15], we studied cross-correlations between number of hops, RTT, route, and capacity of end-to-end paths. In this paper, we study correlations between capacity and available bandwidth metrics. We show these different end-to-end network metrics in Figure 1 as well as some of the measurement tools available.

In this paper, we target the correlations between the capacity and the available bandwidth metrics. Note that the available

bandwidth between two given end hosts can be highly dynamic and depends on the usage of the network. On the other hand, the end-to-end capacity between two hosts changes only when links on the network paths between the hosts is upgraded or network routing changes occur - thus capacity changes less frequently. But if our analysis shows that there is indeed a strong correlation between capacity and available bandwidth, an application can use the slow-changing capacity values to estimate the more dynamic available bandwidth metric.

We analyze the relation between these two metrics using three different methods: (i) utilization factor, (ii) rank correlation, and (iii) top-k correlation. Our results show that the capacity cannot be used to estimate the available bandwidth between a pair of end hosts precisely. However, there is a strong rank correlation and top-k correlation. This is useful in several applications. For example, in a CDN, for a given client, the ordering of the content server nodes according to their available bandwidth is more important than precise information about the available bandwidth.

Peer-to-peer file sharing applications such as Gnutella, BitTorrent, and variants of these systems sort the results of a search based on the capacity between the user's machine and the peer machines. They implicitly assume that capacity is a good indicator of available bandwidth. However, to the best of our knowledge, this paper is the first to quantify such correlations between end-to-end path capacity and available bandwidth to minimize the monitoring cost.

Our results are based on the analysis of the data from the PlanetLab testbed. Though this testbed spans across the Internet, the placement of nodes and distribution does not necessarily uniformly sample the overall Internet. While the results can be reliably trusted and leveraged in designing systems on top of the PlanetLab, service developers on a different overlay network should use our framework and methodologies to repeat and confirm the results on their network.

II. S³ DATA

In this section, we describe the data set we use for this analysis and about the S³ service on PlanetLab from which we collected the data. Our Scalable Sensing Service (S³) [14] is running on PlanetLab since January 2006. The S³ system is a loosely coupled Service Oriented Architecture (SOA) with a web-services interface for tools and collects different all-pair metrics: latency, available bandwidth, capacity bandwidth, and lossrate. For latency, we perform traceroutes from all nodes to approximately 20 "landmark" nodes distributed across the globe, once about every 30 minutes, and use NetVigator [11] to infer the all-pair latency. We use Pathchirp [7] and Spruce [8] for available bandwidth, Pathrate [6] for capacity, and Tulip [16] for lossrate measurements. While many of these tools have been developed a while ago, deploying them in the large scale is still a challenge [14]. Significant engineering effort has been spent in making sure that the tools run reliably and with reasonable accuracy.

For available bandwidth, we use two different tools. The Spruce tool needs capacity values as input and we run it using

the measurement values from Pathrate. Spruce measures the fraction of bandwidth used and uses the provided capacity value to compute the remaining available bandwidth. The other tool, Pathchirp is run for thirty seconds and we average the values that pathchirp outputs - one for every roundtrip time averaged for a window of 11 samples. Both of these tools for available bandwidth measurement are run in a round-robin fashion for all hosts: spruce and pathchirp measurements are run in succession before moving to the next destination host. It takes about 12 hours on average to complete a cycle for a few hundred nodes.

To obtain quick estimates of capacity, we run Pathrate in the Quick Termination mode. With each measurement, Pathrate tool provides a coefficient of variation (COV) value representing the confidence in the measurement. We use results only when the COV value is between 0 and 1. We run these measurements in a loop at each source node measuring each destination in a round-robin fashion. It takes approximately a day on average to complete an entire cycle of measurements for all PlanetLab nodes.

While we have been collecting data for all production PlanetLab nodes, obtaining a complete set is difficult because of the churn in the system and network state at each node. We do not depend on the completeness of the data or expect it to be without errors - after all, this is unlikely in real world networks. We used error bounds reported by the tools to clean up the data set and provide analysis based on the available data.

III. CAPACITY AND AVAILABLE BANDWIDTH

We explore the correlation between end-to-end path capacity and the end-to-end available bandwidth across all paths for each host. Capacity of a path from node A to node B corresponds to the maximum data rate at which data can be transferred from node A to node B assuming no other flows in the network. Given a network route, the capacity of an end-to-end path depends on the physical properties of the links on the path. Available bandwidth corresponds to an instantaneous rate at which node A can push data to node B and is affected by other traffic on the path. End-to-end capacity changes rarely as the routes are static for most of the time in the Internet and links are not updated often. But, end-to-end available bandwidth is a highly dynamic metric.

Many distributed applications such as content distribution networks and distributed data stores need to monitor end-to-end bandwidth between individual nodes. This information is needed to ensure lower response time for clients. For example, consider a content distribution network (CDN) such as Akamai [1] and CoDeen [17]. In CDNs, each file object is replicated at several CDN servers. When a client requests a web object from a CDN server C, if that server does not locally have that object, the server downloads the object from one of the servers that has the object. The CDN server then responds to the client with the downloaded object. Since each object gets replicated at several servers, a CDN server has multiple options to choose from when downloading an object. By downloading

the object from a server S such that the available bandwidth on path S to C (i.e., *DOWNLOAD* bandwidth) is larger than the available bandwidth on paths from all other servers to C , the overall response time for the client can be reduced. In applications such as online photo stores, we need to consider *UPLOAD* bandwidth to decide which particular store to upload photos to, so that overall transfer time is lower.

Unfortunately, measuring available bandwidth between all hosts at high frequency is not feasible. In [5], we show that even a few *simultaneous* measurements for end-to-end available bandwidth can significantly affect the accuracy of the measurements and lead to high CPU and memory load on the measuring machine. In large systems with few hundreds of end nodes, measuring in a sequential fashion such as round-robin can take long for even a single round to complete. Moreover, since available bandwidth is dynamic in nature, the measured values might be stale.

Thus, we explore if capacity, a slow varying end-to-end metric, can be used to estimate the goodness of a path with respect to its available bandwidth. Note that a single capacity measurement (e.g., using the PathRate tool) consumes more bandwidth and time than a single available bandwidth measurement (e.g., using Spruce or PathChirp). But, since capacity does not change often, we can perform capacity measurements at a much slower rate. Also, we can leverage the techniques presented in our previous paper [15] to further reduce this monitoring overhead.

In the following, we present our three methodologies for measuring correlation among these two metrics and present results of our analysis with the S^3 dataset.

IV. UTILIZATION

While available bandwidth (ab) is certainly related to capacity (cap), $ab = cap * (1 - util)$, our investigation is to determine the coefficient of that correlation. Note that the available bandwidth is influenced by both utilization of the path and the capacity. We hypothesize that the capacity of a path plays a larger role in determining available bandwidth than the utilization and hence capacity can be used as a good estimator for the available bandwidth. If this hypothesis holds, we should observe either very low utilization or high uniformity in the utilization across all paths.

We compute the utilization using Spruce and Pathchirp measurements as follows. For every available bandwidth measurement (from Spruce and Pathchirp) performed at time t , we pick the Pathrate capacity measurement closest to t for the particular path. We then compute the utilization as a fraction of the available bandwidth to the capacity value. Hence, with this analysis, we have several samples of utilization values for each path.

A couple of notes before presenting the results. First, we ignore the samples of a path where the computed utilization is < 0 . This can happen in Pathchirp as PlanetLab bandwidth caps affects the Pathrate and Pathchirp tools differently. Second, since S^3 runs Pathrate in quick-termination mode, many measurements fail or end with an output value with COV

> 1 . For Spruce, this has no impact as Spruce measures the utilization and uses the input capacity value to compute the available bandwidth. But this can affect utilization computation for Pathchirp. Hence, for a Pathchirp’s measurement at time t , we look for a successful capacity measurement in $[t - 1day, t + 1day]$ with $0 < COV < 1$. If there is no successful Pathrate measurement, we ignore that Pathchirp measurement in the analysis.

In Figure 2, we present the CDF of utilization of all paths. We present CDF curves for three different statistics: mean, median, and 90th-percentile. For each path, we compute these statistics using the utilization samples. So, a point (x, y) on 90th-percentile curve in this graph denotes that y fraction of paths have 90% samples below utilization x . From the analysis, we observe that utilizations are fairly modest for many of the paths – 80% of the paths have average and median utilization lower than 40%. But we can not conclude that the capacity measurements can be used to infer available bandwidth as there is enough spread in the utilization as can be observed from 90th-percentile curves.

V. RANK CORRELATION

We study the rank correlation between capacity and available bandwidth for each host in the following manner. We consider two cases for this analysis: (i) *DOWNLOAD*: when a host can download content from multiple sources, can it use the capacity values to rank the nodes and determine the best source such that the available bandwidth from that source to the host is the maximal? (ii) *UPLOAD*: when a host can upload its content to multiple servers, can it use capacity ranks and determine the best server? For each of these cases, we study the Spearman’s rank correlation for each host using the data from the S^3 measurements.

To determine the Spearman’s rank correlation coefficient, we rank the paths of a host—from all other nodes to this host in the *DOWNLOAD* and from the host to all other nodes in the *UPLOAD*—according to the ascending order of their capacity values and available bandwidth values. Let the capacity rank and available bandwidth ranks of an i th path with capacity x_i and available bandwidth y_i be r_i^x and r_i^y . The Spearman’s rank correlation coefficient ρ can be computed using Equation 1 where n is the number of paths used in the computation.

$$\rho = 1 - \frac{6 \sum_i (r_i^y - r_i^x)^2}{n(n^2 - 1)} \quad (1)$$

We have more than two years of measurements for Spruce, Pathchirp, and Pathrate measurements. So, for each host, we compute the correlation coefficient for each day’s measurements. Thus, we compute several samples of coefficients for each host. We consider the average, median, and 10th-percentile statistics from the samples for each host. Note that the range of ρ is $[-1, 1]$ and a value of 1 implies a perfect positive correlation and values close to 1 imply a strong positive correlation.

In Figure 3, we plot the Spearman rank correlation value across hosts for *DOWNLOAD* and *UPLOAD* for the Spruce

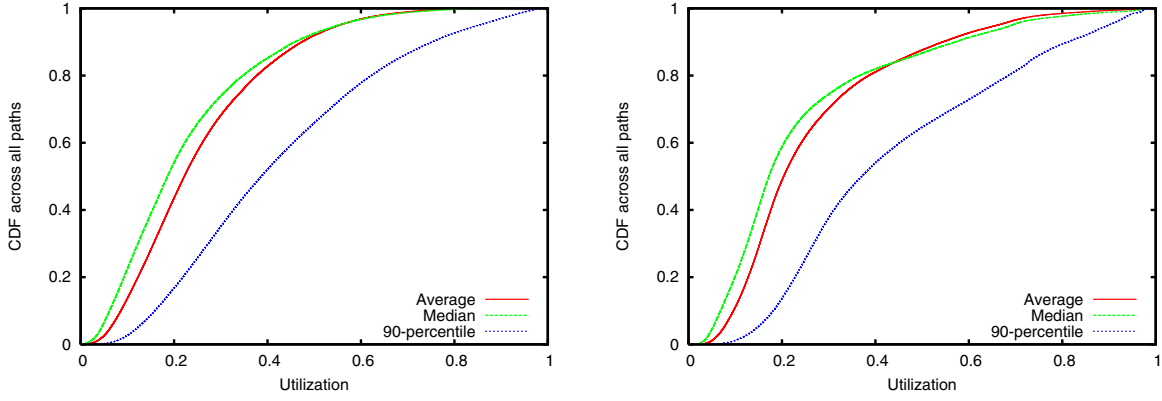


Fig. 2. CDF across paths for different percentiles of utilization: (a) Spruce (b) Pathchirp

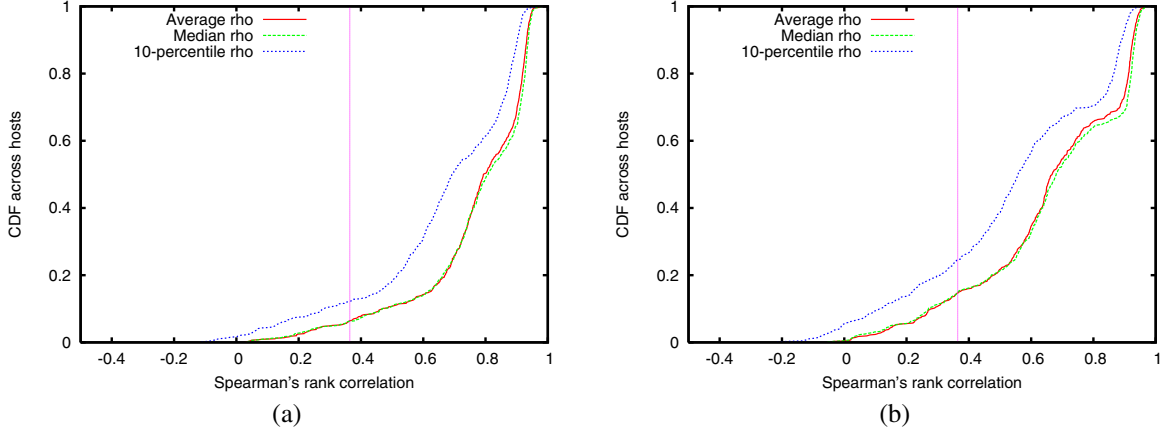


Fig. 3. Spruce: Spearman rank correlation value across hosts for (a) DOWNLOAD, (b) UPLOAD

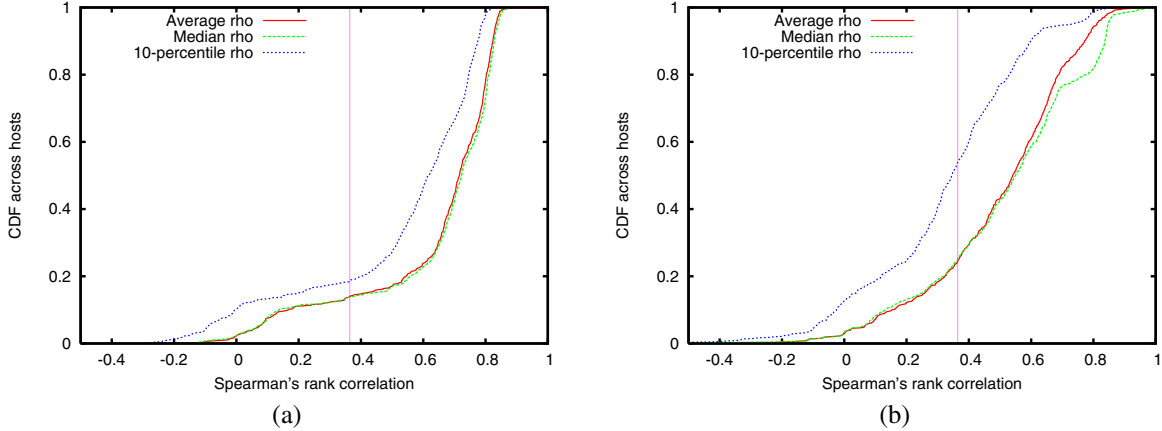


Fig. 4. Pathchirp: Spearman rank correlation value across hosts for (a) DOWNLOAD, (b) UPLOAD

data. In both cases, we have data for about 650 hosts. Both graphs show a strong positive correlation between capacity ranks and the available bandwidth ranks. In the graphs, we also show the line for $\rho = 0.364$ which is the critical value for 0.05 significance (i.e., only 5% chance for such ordering to happen by chance) for a sample with 30 points. More than 90% hosts in the DOWNLOAD and more than 80% hosts in the UPLOAD have average ρ greater than this critical value. Also, more than 80% hosts in the DOWNLOAD and more than 70% hosts in the UPLOAD have 10th-percentile ρ greater than the critical value i.e., 90% of the samples of these hosts have ρ greater than the critical value. This implies a strong

correlation between capacity ranks and available bandwidth ranks. We observe high rank correlations with Pathchirp data too as shown in Figure 4 except for the 10-percentile curve in the UPLOAD. We are further investigating the reason for this case.

Overall, there is a strong rank correlation between the capacity and the available bandwidth metric on the PlanetLab paths. Thus in cases where nodes need to be ranked based on the available bandwidth, capacity measurements can be used to get a good estimate of that order.

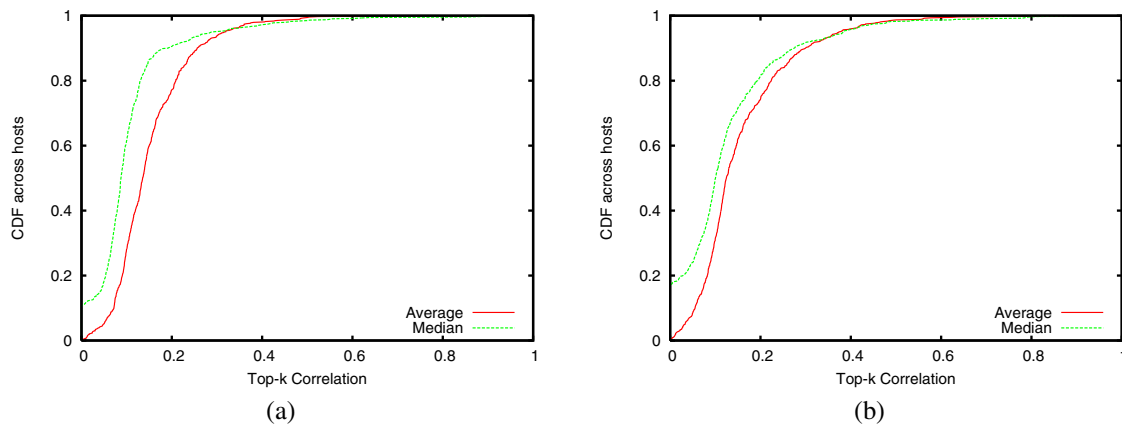


Fig. 5. Pathchirp: Top-k correlation (with $k=5$) across hosts for (a) DOWNLOAD, (b) UPLOAD

VI. TOP-K CORRELATION

The Spearman rank correlation metric compares the rank order based on capacity and available bandwidth for paths from/to *all* other hosts with respect to a host. In this section, we consider a specific correlation that is more significant from a system building perspective. For the CDN example described in Section III, the CDN server C is interested in just another server from which to download the content. Hence, we care about only the highest bandwidth path.

With respect to a host, we define *Top-k Correlation* from end-to-end capacity to end-to-end available bandwidth as follows: Given a node, consider k paths with highest capacity from that node to the other nodes. Suppose the maximum available bandwidth among those k paths be max_k . Suppose max be the maximum available bandwidth across all paths. We define the fractional difference $(max - max_k)/max$ as the Top-k correlation factor. This value always lies between 0 and 1 and smaller is better. In simple words, this correlation measures the goodness of the top-k paths chosen by the capacity metric are in terms of having good available bandwidth.

In Figure 5, we present the CDF for this correlation for $k = 5$ for DOWNLOAD and UPLOAD cases. Observe that at least 80% of the hosts have median correlation smaller than 0.2 and 90% have average and median correlation smaller than 0.4. We observe similar graphs for Spruce dataset. This denotes a strong Top-5 correlation. Hence, a simple system can be designed to exploit this: instead of tracking highly variable available bandwidth across all hosts, only track capacity and when needed pick the top-5 capacity paths and perform available bandwidth measurement on only those paths.

VII. CONCLUDING REMARKS

An important challenge for any large scale e2e network monitoring system is to properly tune the periodicity of the measurements performed on the network. Too many measurements overload the network and interfere with the other traffic; too few measurements lead to inaccurate state of the system. Different e2e metrics have different monitoring overheads and different dynamicity properties. Our goal is to study and leverage the correlations among different metrics to minimize the monitoring overheads while preserving the accuracy. In this paper we focus on the correlation between

the capacity and the available bandwidth metrics. We study the relationship using three different correlation techniques applied on the data that we collected on the PlanetLab testbed since January 2006. Our results show that the capacity can not be used to estimate available bandwidth of a single path precisely. However, there is a strong rank correlation and top-k correlation. This is useful in several applications (such as content distribution networks) where ordering of the nodes according to their available bandwidth is more important than precise information about the available bandwidth.

REFERENCES

- [1] <http://www.akamai.com>.
- [2] <http://aws.amazon.com/ec2/>.
- [3] <http://planet-lab.org>.
- [4] <http://geni.net>.
- [5] H. H. Song and P. Yalagandula, "Real-time End-to-end Network Monitoring in Large Distributed Systems," in *Proc. IEEE COMSWARE*, 2007.
- [6] C. Dovrolis, P. Ramanathan, and D. Moore, "Packet-Dispersion Techniques and a Capacity-Estimation Methodology," *IEEE/ACM Transactions on Networking*, vol. 12, no. 6, Dec 2004.
- [7] V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient Available Bandwidth Estimation for Network Paths," in *Proc. of the PAM 2003*, April 2003.
- [8] J. Strauss, D. Katabi, and F. Kaashoek, "A Measurement Study of Available Bandwidth Estimation Tools," in *Proceedings of the ACM IMC 2003*, Miami, FL, October 2003.
- [9] T. S. E. Ng and H. Zhang, "Predicting Internet Network Distance with Coordinates-Based Approaches," in *Proceedings of the IEEE INFOCOM 2002*, New York, NY, June 2002.
- [10] B. Wong, A. Slivkins, and E. G. Sirer, "Meridian: A Lightweight Network Location Service without Virtual Coordinates," in *Proceedings of the ACM SIGCOMM 2005*.
- [11] P. Sharma, Z. Xu, S. Banerjee, and S.-J. Lee, "Estimating Network Proximity and Latency," *ACM Computer Communications Review*, vol. 36, no. 3, pp. 41–50, July 2006.
- [12] H. Song, L. Qiu, and Y. Zhang, "NetQuest: A Flexible Framework for Large-Scale Network Measurement," in *Proc. of the ACM SIGMETRICS 2006*.
- [13] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, "Vivaldi: A Decentralized Network Coordinate System," in *SIGCOMM'04*.
- [14] P. Yalagandula, P. Sharma, S. Banerjee, S.-J. Lee, and S. Basu, "S³: A Scalable Sensing Service for Monitoring Large Networked Systems," in *SIGCOMM INM Workshop*, 2006.
- [15] P. Yalagandula, S.-J. Lee, P. Sharma, and S. Banerjee, "Correlations in End-to-End Network Metrics: Impact on Large Scale Network Monitoring," in *Proc. of GI Symposium*, 2008.
- [16] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson, "User-level Internet Path Diagnosis," in *ACM SOSP 2003*.
- [17] L. Wang, K. Park, R. Pang, V. S. Pai, and L. Peterson, "Reliability and security in the codeen content distribution network," in *USENIX*, 2004.