On the Marginal Utility of Network Topology Measurements

Paul Barford, Azer Bestavros, John Byers, Mark Crovella

Abstract—
The cost and complexity of deploying measurement infrastructure in the Internet for the purpose of analyzing its structure and behavior is considerable. Basic questions about the *utility* of increasing the number of measurements and measurement sites have not yet been addressed which has led to a "more is better" approach to wide-area measurement studies. In this paper, we step toward a more quantifiable understanding of the marginal utility of performing wide-area measurements in the context of Internet topology discovery. We characterize the observable topology in terms of nodes, links, node degree distribution, and distribution of end-to-end flows using statistical and information-theoretic techniques. We classify nodes discovered on the routes between a set of 8 sources and 1277 destinations to differentiate nodes which make up the so called "backbone" from those which border the backbone and those on links between the border nodes and destination nodes. This process includes reducing nodes that advertise multiple interfaces to single IP addresses. We show that the utility of adding sources beyond the second source quickly diminishes from the perspective of interface, node, link and node degree discovery. We also show that the utility of adding destinations is constant for interfaces, nodes, links and node degree indicating that it is more important to add destinations than sources.

Keywords— **Network measurement, traceroute, topology discovery, Internet tomography**

I. Introduction

An emerging strategy to gain insight into the conditions and configuration inside the Internet is the use of end-to-end measurements from a set of distributed measurement points. A design goal of the Internet has been to emphasize simplicity in its internal components; for this reason, measurements made at network endpoints are especially attractive. An example of this approach is the use of `traceroute` [17] for the discovery of network connectivity and routing.

While traceroute is remarkably flexible and informative, it is an open question how useful traceroute is for uncovering topological information about the Internet. In this paper we study the use of traceroute as a tool for Internet topology discovery. We consider the common case, in which active measurement sites (traceroute sources) are relatively scarce, while passive measurement sites (traceroute targets) are plentiful. In such experiments, each traceroute source is able to discover a directed graph, induced by IP routing, from itself to all of the destinations.[1] We are interested in the properties of the graph that is formed by the union of these individual views.

In order to find the union of these views it is necessary to identify routers that advertise multiple interfaces, and to associate each advertised interface with a router. Our first contribution is to discuss our experiences in solving this problem and assess the importance of this issue when merging multiple traceroutes into a single graph.

Our main contribution is to show how studying this graph helps clarify how end-to-end paths pass through the Internet. A principal observation is that the marginal utility of adding additional active measurement sites declines rapidly after the second or third site. This motivates a rough model for the routing graph discovered by traceroute as a richly-connected "switching core" fed by ingress and egress paths ("feeders"). Our work indicates that the core consists of a relatively small fraction of nodes and we show that almost all paths in our data pass through this core.

If the source-destination pairs in our data are representative of typical endpoint pairs for IP flows, then the switching core is common to most end-to-end paths taken in the Internet. Thus the properties of the core are especially

This work was partially supported by NSF research grants CCR-9706685 and ANIR-9986397. The data used in this research was collected as part of CAIDA's skitter initiative, http://www.caida.org. Support for skitter is provided by DARPA, NSF, and CAIDA membership.

P. Barford is with the University of Wisconsin, Madison. A. Bestavros, J. Byers and M. Crovella are with the Computer Science Department at Boston University. E-mail: pb@cs.wisc.edu, {best,byers,crovella}@cs.bu.edu .

[1] We make the simplifying assumption that IP routing paths are stable over the timeframe of individual traceroute executions; while this assumption is sometimes incorrect, it frees us to focus on a different set of questions. Note that this assumption does not imply that the resulting directed graph from a source is a tree.

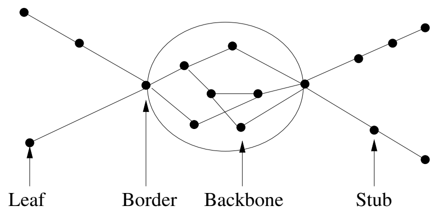

Fig. 1. Classification of Internet nodes

interesting for understanding Internet performance. We note that, compared to the set of all switching core nodes present in our dataset, the majority are visible from only a single measurement source. That is, sets of IP flows originating from different locations tend to pass through similar sets of switching nodes on their way to common destinations. This makes it relatively less productive to discover new switching nodes by adding sources, even when the set of measurement destinations is large.

To assist us in our task, we have leveraged detailed routing traces gathered by CAIDA (Cooperative Association for Internet Data Analysis) for the Skitter project [11]. These traces span thousands of routes between 8 sources and 1277 destinations taken repeatedly over the course of several months. While we can provide no guarantee that the CAIDA measurement sites were chosen in a representative way, the location of the sites are geographically diverse, spanning North America, Europe and Asia. Compiling together all nodes and edges of the graph visited by routes in these traces, we built up a partial picture of the way the Internet backbone appeared in May 2000. Then, using this picture as our baseline, we go back to the traces to observe which paths, or collections of paths, were most productive in generating the overall map.

To understand the topology discovery process in greater detail, we employ a node classification technique which organizes nodes into one of four types: leafs, stubs, border and backbone illustrated in Figure 1. For the graph that we evaluate (after resolving routers that advertise multiple interfaces to a single node) over half of the nodes discovered are classified as backbone nodes while less than 10% are border nodes, giving a picture of the collected IP routes as consisting of a large backbone with somewhat limited ingress and egress. Much of our analysis focuses on marginal utility with respect to the discovery and characterization of *backbone* nodes.

The rest of the paper is organized as follows. In Section II we describe related analytical work in evaluating the effectiveness of deploying wide-area measurement infrastructure with a focus on topology mapping. In Section III, we establish basic definitions for the network discovery problems we consider and outline how to cast these problems in a marginal utility framework. In Section IV, we describe our data set, our graph classification procedure and the limitations of our approach. We present our statistical results for interface disambiguation, node classification and marginal utility in Section V. We define information-theoretic tools and results from their application to the data in Section VI. We summarize, conclude and discuss future work in Section VII.

II. Related Work

A number of research groups have generated maps of the Internet using route tracing tools such as `traceroute` [8], [11] and have built repositories of Internet mapping information. We now survey the most closely related of those works to ours.

Work by Govindan [24], [14] outlines heuristic techniques for generating complete domain maps. One of the challenges in this area goes far beyond the capabilities of traceroute, and lies in mapping out the nooks and crannies of regions in autonomous systems (AS's) which do not transit a substantial amount of data. This work also discusses the problem of *alias resolution* in detail, which is the same as our interface disambiguation problem. They employ the same techniques as we do to resolve multiple interfaces at a single node.

Jamin et al [18] study algorithms for effective placement of Internet instrumentation in the context of their IDMaps project, a project which seeks to provide an Internet-wide distance estimation service, following the architecture designed in [12]. The majority of their work focuses on algorithmic approaches for placing a fixed set of measurement sites on generated topologies, and measurements on the effectiveness of the placement. While their work mentions diminishing returns in the context of infrastructure placement, it does not provide analytical results in this area.

Pansiot and Grad [19] report on the topology resulting from a detailed collection of end-to-end routes they collected in 1995 with the goal of constructing representative multicast trees. Using traceroute, they traced routes to 5000 geographically distributed hosts chosen from their network accounting database. Then they chose a subset of 11 of these hosts to be additional sources of routes, and ran traceroute from these 11 hosts to each of the original 5000 hosts (with the assistance of the Loose Source Routing option). In the topology revealed by this experiment, they found that the routes from *any* subset of six sources contained nearly 90% of the nodes and edges ultimately discovered. They also provided a classification of nodes similar to the one we provide and present the distribution of the degree of nodes of the graph they discover, a distribution which clearly follows a power law. (This power law and evidence of other power laws in this data set, as well as in other data sets, were reported in [10]). However, they provide no qualitative discussion of the characterization of the topology that they obtain, nor do they attempt to quantify the marginal value of information gained as measurements are added.

Broido and Claffy [5] also leverage traceroute datasets from CAIDA to build up and study the aggregation of a set of tree topologies induced by IP routing. While their effort does provide useful characterizations and insights into these topologies, it does not focus on the questions of marginal utility which we study here.

Paxson [20], [21], [22] deployed a "network probe daemon" (NPD) at 37 sites in the wide-area, which used `traceroute` to investigate end-to-end routing behavior and later, performance of transport protocols between all pairs of sites over several weeks. His work emphasized the importance of exploring a large number of paths to observe rare and occasionally anomalous routing behavior. Paxson also studied the issue of interface disambiguation in [21] from the perspective of resolving nodes to geographic locations and not necessarily specific routers. Wide-area measurement and analysis continues to be a focus of many research and industry groups including NIMI [2], WAWM [3] and Surveyor [25]. Another piece of generally related work are the Internet weather reports such as [27], [26]. These are general compilations of the packet loss and round trip time measurements from Internet monitoring boxes deployed in the wide-area.

Finally, other recent studies have used measurement-based approaches to study aspects of the Internet topology, albeit using different tools. Some researchers have used logs collected in the wide-area by BGP-capable routers to study the effects of policy-based routing, with an emphasis on quantifying the inflation in route lengths [16], [31], [28]. At a higher level of abstraction, there has been considerable work on understanding AS-level connectivity [13], [4] including work which leverages traceroute measurements and BGP routes to help infer AS-Level connectivity [7], [6], These pieces of work, like ours, emphasize the importance of incorporating snapshots taken from multiple vantage points to providing the most complete reflection of the overall topology.

III. Definitions and Objectives

The network discovery problems we consider have a natural graph-theoretic formulation, study of which may be of independent interest both to theorists and to researchers who wish to better characterize network topologies. Consider a network topology represented by an undirected graph $G = (V, E)$ in which $|V| = n$. The central question which we study is the extent to which the underlying topology can accurately be characterized as the number of end-to-end observation points grows. In practice, we assume that k sources and m destinations are chosen uniformly at random from the vertex set of this graph. Then we consider the fraction of the vertex set and edge set that is covered by the set of routing paths from the sources to each of the destinations, using the following terminology.

Definition 1: Given a graph $G = (V, E)$ and a subgraph $G' = (V', E')$ of G, the *node coverage* of G by G' is the ratio $\frac{|V'|}{|V|}$. Similarly, the *edge coverage* of G by G' is the ratio $\frac{|E'|}{|E|}$.

Definition 2: Take a set of source vertices $S \subseteq V$ and a set of destination vertices $D \subseteq V$. Also assume that we have a routing algorithm R which selects fixed routes between all pairs $s \in S, t \in D$. We define the union of the set of (s, t) paths in G to be the *subgraph of G induced by* R on all pairs of routes from S to D.

The subgraph induced by a routing algorithm corresponds to overlays of "projections" from multiple sources, i.e. the union of individual directed graphs rooted at these vantage points to the set of destinations. The functions defined below describe how coverage increases as the number of endpoints used to generate the induced subgraph grows.

Definition 3: For a graph G with routes induced by R and for parameters k and m, let $v_G(k, m)$ denote the expected node coverage of G by the subgraph induced by a randomly chosen set of sources S of cardinality k, a randomly chosen set of destinations D of cardinality m. Similarly, let $e_G(k, m)$ denote the expected edge coverage of G by such a subgraph.

The rate at which v_G and e_G change with respect to k and m give insight into the benefit of conducting additional measurements or deploying additional measurement sites in discovering a given graph G. These functions which we consider are a general family of scaling properties of which some specific cases are also being carefully studied. For example, consider these functions for the special case of $k = 1$, which describes the scaling behavior of a multicast tree to m clients. This relationship was first considered by Chuang and Sirbu [9]. Their work, and subsequent work by Phillips, Shenker and Tangmunarunkit [23], demonstrates that the number of hops in a multicast tree, i.e. $e_G(1, m)$, scales as $\mathbf{E}[H_G]m^{0.8}$ for $m << n$, where $\mathbf{E}[H_G]$ captures the average path length in G. In our work, we consider cases in which $k > 1$ and where m can be moderately large (we note that another interesting special case arises when $m = n$).[2] A related direction of future interest lies in the characterization and understanding of those regions of the Internet topology which are relatively difficult to uncover using traceroute. Such a study could conceivably lead to a better understanding on the connection between topology and routing behavior or provide further insight into relationships between topology and peering agreements.

We focus specifically on *marginal utility*, i.e. the incremental benefit obtained by conducting one or more additional measurements. For edge coverage, we define the marginal utility of adding targets as follows (related definitions are similar):

Definition 4: The *marginal utility* of conducting edge coverage measurement $i + 1$ on graph G' from a set of k sources is $e'_G(k, i+1) - e'_G(k, i)$.

This and related quantities will be the primary focus of the rest of the paper. We first study marginal utility from a purely empirical perspective, focusing on the distinction between the core of the network and feeder networks. We then return to the problem from a theoretical perspective, developing and studying an information-theoretic formalism of marginal utility in this context.

IV. Experimental Methodology

We now present the experimental methodology we used to investigate scaling behavior in the Internet. The traceroute datasets we use in this section deviate in several ways from the ideal theoretical framework we prescribed in Section III, and a significant portion of this section is devoted to a discussion of additional assumptions which we made and a description of mechanisms for post-processing of actual datasets.

A. Internet Trace Data

The topology data used in this work was supplied by the Skitter project at CAIDA. The Skitter project has a number of goals including Internet mapping, route characteristic analysis and performance analysis. At the time the primary dataset for this study was collected (May 2000), the Skitter infrastructure consisted of 16 source nodes deployed around the world; we received data from 8 of those nodes. Each source node sends traceroute-like probes to destination nodes located world-wide. All of the destination nodes are Web servers. Our primary data set contains results from traces run to 1277 destinations; The source nodes and the corresponding upstream providers (listed in parentheses) were located in Hamilton, NZ (University of Waikato); Tokyo, Japan (APAN), Singapore, SG (provider unknown); San Jose, CA, USA (Worldcom); San Jose, CA, USA (Qwest); Ottawa, CA (CANET); London, UK (RIPE); and Washington DC, USA (AboveNet). On average, probes are sent to each destination once every 30 minutes. While it is not clear precisely how destinations for destinations are selected in Skitter, the Skitter web site states that destinations are randomly sampled from a "crawl of IP address space" [11]. We also include results from a larger dataset with 12 sources and over 300,000 destinations. This dataset includes the eight sites listed above, with the exception of Singapore, plus Marina Del Rey, CA, USA (ISI); Moffett Field, CA, USA (NASA), Palo Alto, CA, USA; San Diego, CA, USA (CAIDA) and London, UK (AboveNet).

B. Node and Edge Classification

Using the experimental results we gathered, it was immediately apparent that the network graph under observation was not a random network, but consisted of two constituent components: 1) a central routing core, and 2) a set of "feeder" links which feed into the backbone. We then focused on how successfully traceroute could be used when applied to identifying these two constituent components, which had evidently different properties. A central challenge to doing so is to develop an automated procedure which classifies nodes (and edges) into these categories. Using the terminology of Zegura et al [33] to describe their GT-ITM topology generator, we assume that there is a natural and identifiable separation between transit domains, which comprise the Internet backbone, and stub domains, which only transit traffic either originating or terminating in their domain. In this model, the set of transit domains typically forms a highly connected backbone, with a number (at least two and often many more) of node-disjoint paths between any two transit domains, while stub domains typically consist of trees with a single connection to the transit domain backbone.

The objective of our classification algorithm is to take our observations of a topology and determine the boundary between where the backbone ends and stub domains begin based on the available evidence. There are a number of reasons why our classification procedure may fail to classify nodes correctly – in future work, we intend to conduct validation trials to measure the effectiveness of our classification methods from traceroute measurements. Routes to destinations which did not respond to the traceroute requests were discarded, but routes in which intermediate hosts failed to respond to ICMP requests were included. Even using a relatively small number of measure-

[2] While our work is primarily experimental in nature, we believe that the theoretical study of these properties on graphs of interest (such as power-law graphs vs. random graphs) with idealized routing algorithms (such as use of shortest-path routes) may help provide deeper insight.

ment sites, a clear distinction between backbone links and stub links in this subgraph G' emerged (we will demonstrate this and quantify how much error was removed from our classification process as the number of measurements increased).

Given this subgraph, our classification procedure now amounts to a labelling of the nodes and edges of G'. To this end, nodes which correspond to routers and Internet hosts are classified as *core routers*, *border routers*, *stub routers* and *leaf nodes*. Our node classification procedure is performed as follows. First, leaf nodes are identified and labelled as such, and edges adjoining leaf nodes are classified as stub links. Then, in a bottom-up fashion, internal nodes which adjoin a set of edges all but one of which are stub links, are classified as stub routers.

Upon completion of this procedure, the logical trees forming the visible portion of stub domains in G' are established. The remaining unclassified nodes all satisfy the property that at least two of their incident edges are unlabeled – that entire unlabeled portion of the graph G' is the network backbone, and we classify it as such. Within the network backbone, unlabelled nodes which adjoin at least one stub link are classified as border routers, all remaining nodes are classified as core routers, and those links which are not yet classified are backbone links. Figure 1 provides a simple diagram of the results of a classification procedure.

C. Coverage vs. Marginal Coverage

In the examples we have described so far, our classification procedure labels the subset of Internet nodes and links visible in one or more of the end-to-end measurements in our study. Since we are primarily interested in characterizing the Internet backbone, and since we have no expectation of completely mapping stub domains, we would ideally like to measure the *coverage* of the Internet backbone achieved by our experiments, using the definitions presented in Section III. However, this approach is infeasible, as the exact topology of the graph which comprises this backbone is not known a posteriori. While we cannot measure total coverage directly, we can measure the *marginal* improvement in coverage as we conduct additional measurements. To quantify this approach, we take the aggregated information from all of the collected traces as the baseline graph for our study, and measure how well small subsets of the measurements manage to cover that baseline graph. This point highlights an important distinction between marginal coverage and overall coverage — the fact that additional measurements may provide low marginal coverage *does not necessarily* imply that the overall coverage obtained is high — it may be the case that the coverage is poor, but the additional measurements chosen are not productive.[3]

D. Interface Disambiguation

One of the unfortunate issues about building network maps based on traceroute is the existence of routers with multiple interfaces, each with different network addresses. This issue is pervasive – in our study we found that nearly twenty percent of all the nodes we classified as backbone nodes used multiple interfaces with distinct IP addresses to transmit packets. Clearly, studies which disregard this issue, by treating each distinct Internet address as if it were a distinct node, generate inaccurate maps.

The technique we employed to disambiguate multiple interfaces at a single node uses the same basic principle as the one originally used by Pansiot and Grad [19]. The key to this technique is that when transmitting an ICMP message, a router will typically transmit that packet with a source address equal to that of the outbound interface on which the packet is sent. Therefore, if one suspects that a router has two interfaces I_1 and I_2, one can transmit a UDP packet to an unused port at each of those interfaces from a *common* source. If the interfaces are in fact on the same router, the router will respond with two ICMP Port Unreachable messages, both of which will have the same source address I_3, possibly equal to I_1 or I_2. By performing post-hoc probes of this form from a common source (Boston University) to all potentially distinct interfaces, we are able to detect and collapse hosts with duplicate interfaces. Unfortunately, this technique is not infallible. First, approximately 10% of the core routers never responded to UDP messages transmitted to unknown ports; others respond extremely sporadically – we conjecture that the likelihood of response may be correlated with the load on the router. For those routers, disambiguation appears to be impossible with this current technique. Second, our technique relies upon routers responding with a source address equal to the outbound interface. If routers instead respond with a source address equal to the UDP destination address, our technique would be rendered useless. We have no way of estimating the likelihood of this event; however, the fact that we frequently observe routers which respond with addresses which differ from the target address gives us some informal level of confidence that routers do in fact behave according to specification.

[3] An analogous situation arises when choosing black-box test cases to provide coverage of code paths in a software module.

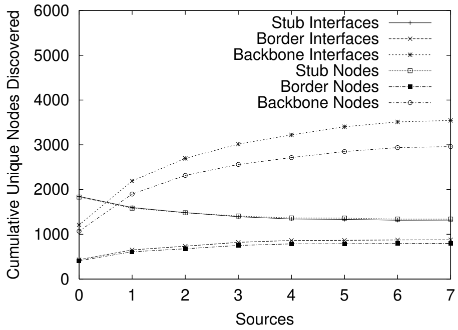

Fig. 2. Class of nodes and interfaces discovered as sources are added (greedily) when classification is not known a priori.

E. Accuracy of Classification

One central aspect of node classification is the accuracy with which we perform classification. With a small number of sources (less than five), many backbone nodes are misclassified as either stub nodes or border nodes by virtue of the fact that the observable Internet is the union of a small number of trees. Figure 2 depicts the relative classification of nodes and links as sources are increased. In some plots in this paper, the order in which sources are added has a significant impact on the overall results. A *greedy* ordering adds the sources in the order which maximizes at each step the total number of distinct nodes observed. A *random* ordering averages over a set of trials in which sources are added purely at random (without replacement) for each trial. In the context of accuracy of classification, behavior of greedy and random orderings were similar; the greedy ordering is depicted.

As we increase the number of sources, our classification procedure increases in accuracy. For example, once we have amassed sufficient evidence to classify a node as a backbone or border router, no set of additional measurements will reverse that classification decision. On the other hand, nodes which we initally classify as part of a stub domain may in fact be backbone nodes, and we may uncover evidence to that effect with additional measurements. In general, we expect to underestimate the fraction of backbone nodes and overestimate the fraction of stub nodes in our classification. The diagram in Figure 2 quantifies that intuition when the number of measurement sites is small, but it is also interesting to note that for this dataset, classification stabilizes after only about five measurement sites (vantage points) are used.

F. Limitations of the Approach

The metrics we propose are difficult to use directly, first because the graph which comprises the Internet is neither fixed nor given in advance. Moreover, even if the graph comprising the Internet were known in advance, our measures of coverage may fluctuate, since the behavior of the routing algorithms in the Internet is non-deterministic, due to the effect of routing policies [28], [32]. Also, while one might hope to quantify topology scaling laws on certain classes of graphs (such as on power-law random graphs) when shortest-path routing is in effect, policy-based routing at the level of AS's skews (or "inflates") routes, making the problem of accurately modeling these scaling laws much more difficult. We note that factors ranging from a wide variety of routing metrics and protocols, variability in network load, and policy-based economic agreements between autonomous systems cause the routes chosen to be quite different than an observer with access only to topology information might expect.

# of Interfaces	1	2	3	4	5	6	7	10
# of Routers	4892	602	169	54	29	13	3	1

Fig. 3. Distribution of observed interface density across routers.

V. Results

The results in this section are divided into five parts: (1) the results of interface disambiguaton run on all nodes in the primary data set, (2) a quantitative evaluation of the number of nodes and links discovered in the backbone as the number of sources and destinations vary, (3) an evaluation of the estimated distribution of node degree in the backbone as the number of sources and destinations vary, (4) fitting the evidence of these evaluations to statistical models and (5) assessing the accuracy of the node classification procedure itself.

A. Results of running the disambiguation procedure

Approximately three weeks after the traceroute data was collected by CAIDA, we ran our interface disambiguation tool to all network interfaces which we had classified as part of the network backbone. An early lesson we learned in our preliminary experiments with the disambiguation software was that a substantial fraction of routers responded to our probes with very low frequency. In an effort to elicit responses from as many responding interfaces as possible, we transmitted five ICMP messages to each interface every twenty minutes for 12 successive hours.

Of the 7451 interfaces on our list, 6510 responded to one or more of our probes and the remaining 941 (12.6%) never responded. We recorded pairs of the form [Target Address, Response Address] and recorded 6709 distinct pairs from the 6510 targeted interfaces which responded. We suspect that this slight (3%) discrepancy is due to route fluctuation affecting the first hop of the return path to B.U.

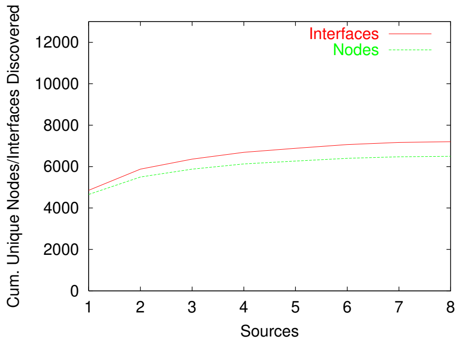

Fig. 4. Number of nodes discovered as sources are added (greedily)

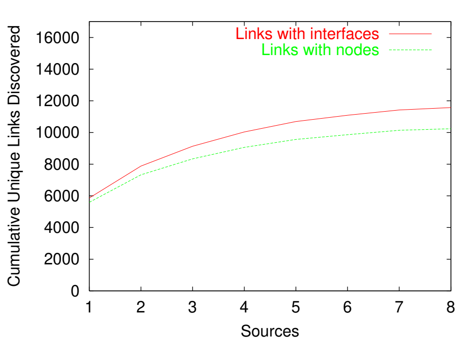

Fig. 5. Number of links discovered as sources are added (greedily)

and does not represent anomalous behavior. The next step we took was to represent the set of addresses present in our list of pairs as nodes in a graph. We drew a correspondence between each connected component of this graph and a single router, where the nodes of the component correspond to distinct addresses for interfaces of the router. Using this strategy, the 6510 targeted interfaces mapped to 5763 distinct routers. The distribution of multiple interfaces we observed is depicted in Figure 3. Using the results in this table, we observed an incidence rate of multiple interfaces of $\frac{871}{5763} = 15.1\%$.

B. Estimating the set of nodes and links in the Internet

In the results below, we have the goal of taking measurements over a set of paths which cover at least n distinct nodes (resp. links) in the Internet. Our first set of experiments demonstrates rapidly diminishing marginal returns as sources are added to trace routes to the full set of 1277 destinations, while our second set demonstrates nearly constant marginal returns as destinations are incrementally added to a destination set targeted by the full set of 8 sources.

In Figures 4 and 5, we demonstrate how the node coverage and link coverage in the Internet improve as sources are added. In both of these plots, there is pronounced evidence of diminishing returns as sources are added, which is highly evident even when running traceroute between a small number of sources (8) and a much larger number of destinations (1277). In each figure we also demonstrate the effect of node and link discovery before and after interface disambiguation.

In Figures 6 and 7, we demonstrate how the node coverage and link coverage in the Internet improve as destinations are added. In both of these plots, there is a relatively constant addition as destinations are added. A simple slope calculation shows that after 200 destinations, approximately 3 new nodes are discovered and 4 new links are discovered when a new destination is added. Each

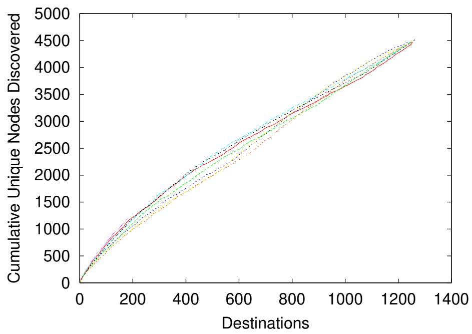

Fig. 6. Number of nodes discovered as destinations are added (randomly). Each line is for a single source

of these figures shows effects after interface disambiguation. Results for interface discovery are approximately the same.

Next, we break down node discovery by node classification. In Figure 8 we show how nodes and interfaces are discovered as sources are added when the node classification is known a priori. This result shows that we primarily discover new backbone nodes and interfaces as additional sources are added, but backbone discovery show diminishing marginal utility.

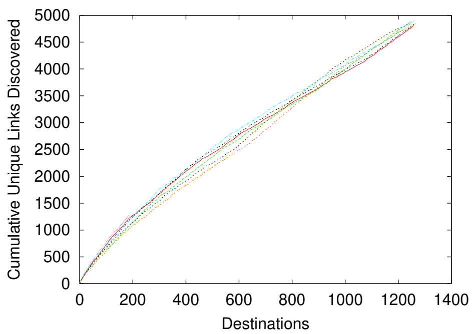

Fig. 7. Number of links discovered as destinations are added (randomly). Each line is for a single source

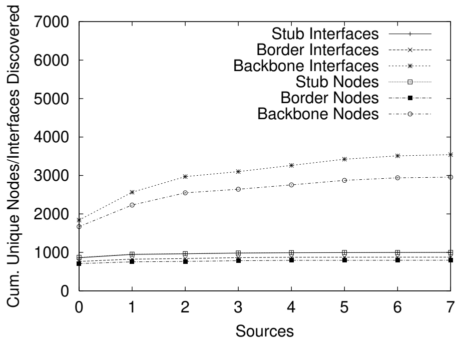

Fig. 8. Class of nodes and interfaces discovered as sources are added (greedily) when classification is known a priori.

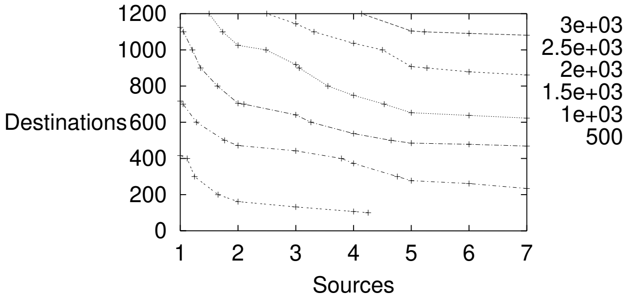

Fig. 9. Backbone node discovery as both sources and destinations are varied

C. Contour Plots

The following diagrams plot the scaling behavior of the subgraphs induced by IP routing for the topologies observed via the CAIDA trace data, assuming that each of the CAIDA sources and destinations reflects a randomly chosen vertex in the graph. In particular, we study the behavior of the function $v_G(k, m)$ as k and m vary. The values of k and m are plotted along the x and y-axes, respectively. Each labelled contour, or isoline, represents the discovery of a fixed constant number of nodes, such that all sets of measurements corresponding to a point (x, y) along a contour have an equal value of v_G. Our experiments were constrained by the fact that we have a limited number of sources, and a much larger set of destinations, so we are unable to plot a full square's worth of data. Another point regarding symmetry: if both sources and destinations are chosen uniformly at random from all locations in the Internet, then the labels of source and destination are arbitrary, which implies that $\forall i, j$, points (i, j) and (j, i) lie along the same contour.

The depiction shown in Figure 9 [4] gives preliminary evidence that the isolines do not follow hyperbolas of the form $x \times y = k$, which would hold in the event that each pair of measurements is equivalently useful. Instead, and pending further study on larger trace data, these contour plots indicate that striking a balance between sources and destinations is relatively less important than making use of a large number of sites overall, which can be done relatively easily by employing more passive targets, rather than requiring deployment of more active measurement infrastructure.

[4] We excluded the one anomalous source which only reached 184 destinations since its inclusion would dramatically alter the results displayed in this figure.

D. Estimating the distribution of node degree in the backbone

As the number of measurement sources increases, the distribution of node degree in the discovered portion of the backbone shown in Figures 10 and 11 (especially in the tail) changes. We calculated the root mean square difference to measure the differences in the distributions as we add nodes, which is shown in Figure 12. Surprisingly, the distribution on node degree in the backbone which we observe after taking measurements from a single site (forming a tree to the sources) is both visibly similar and similar with respect to the RMSE metric to the more refined distribution we identify with subsequent measurements. Quantifying the refinement in our measured distribution over time, in Figure 11, it appears that the weight in the tail may actually diminish somewhat as the number of measurements increase.[5] Another interesting point is that in the RMSE plot in Figure 12, the error actually increases after source 6 is added. Unlike node and edge coverage, which never decrease as additional measurements are conducted, the estimated node degree distribution may in fact become less accurate.

We conducted a similar analysis considering how the addition of destination nodes affects backbone node degree distributions. In Figures 13 and 14 we see the distribution of backbone node degree when all sources are used to trace to increasing numbers of destinations in groups of 100. The figures show that while the body of the dis-

[5] There are several explanations for why this may arise in our datasets which we are currently investigating, including a statistically insignificant sample size, effects from hosts with multiple interfaces, or issues inherent in the measurement set-up. We plan on re-running this experiment with other orderings of the sources as part of our further investigation.

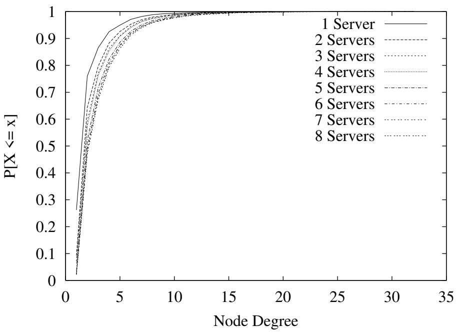

Fig. 10. CDF of backbone node degree as sources are added (randomly)

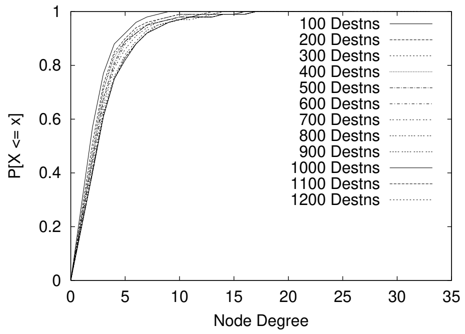

Fig. 13. CDF of backbone node degree as destinations are added (randomly)

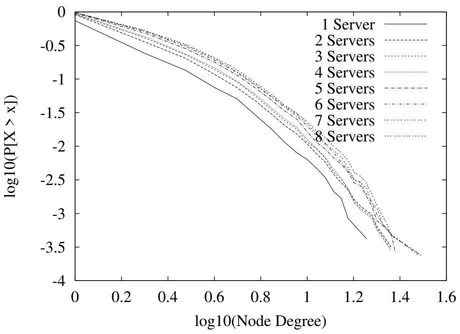

Fig. 11. Tail of CDF of backbone node degree as sources are added (randomly)

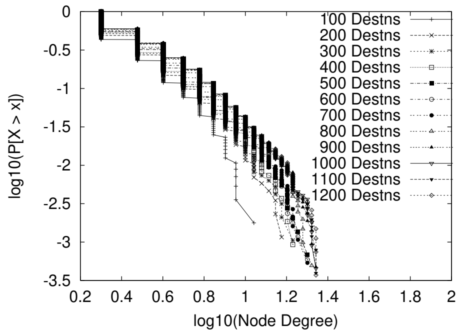

Fig. 14. Tail of CDF of backbone node degree as destinations are added (randomly)

tribution stays relatively constant as destination nodes are added, the tail weight increases as destination nodes are added.

E. Comparison to Larger Dataset

The results so far provide considerable insight into IP routing patterns but the limited size of the node set covered makes it hard to extend our conclusions to the Internet as a whole. To address this we examine a much larger dataset to see whether it shows similar patterns of diminishing returns when adding measurement sites.

The second data sets consists of 12 sources and 313,709 destinations; thus it is more than 10 times the size of the first data set. This dataset was gathered in October, 2000. Source locations for this dataset were Hamilton, NZ; San Jose, CA, USA; London, UK; Marina del Rey, CA, USA; Palo Alto, CA, USA; Tokyo, JP; Ottawa, CA; London, UK; Moffett Field, CA, USA; Washington, DC, USA; San Jose, CA, USA; and San Diego, CA, USA.

Unlike the first data set, in this case sources did not trace routes to a common set of destinations. In fact, no destination in this set is common to all sources. Furthermore, the considerable size of this node set makes it much more difficult to disambiguate interfaces, so our results are in terms of interfaces rather than nodes (routers).

In Figure 15 we show how the number of interfaces discovered grows as we add sources greedily. In this case, adding a source means that we add all the measurement paths originating from that source. The line labelled "interfaces" denotes the number of interfaces that would have been discovered had each source been used independently from the others. In Figure 16 we show how the number of interface-interface links discovered grows as we add sources. Presumably each individual interface-interface link corresponds to a router-router link, so for this plot the distinction between nodes and interfaces is less important.

These figures show a declining slope as sources are

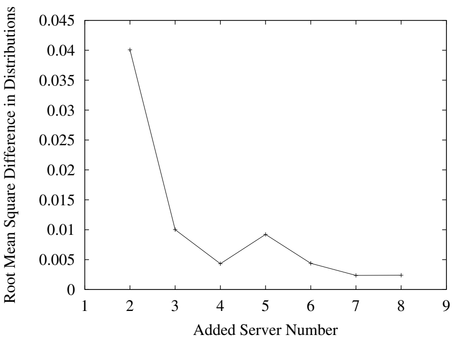

Fig. 12. Root mean squared error difference in backbone node degree distributions

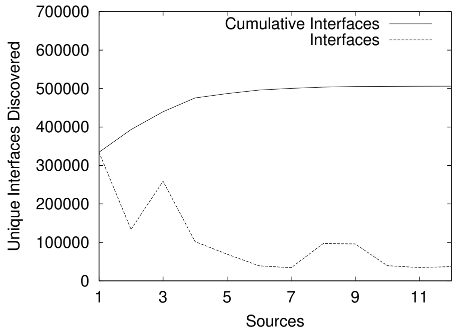

Fig. 15. Number of nodes discovered as sources are added (greedily)

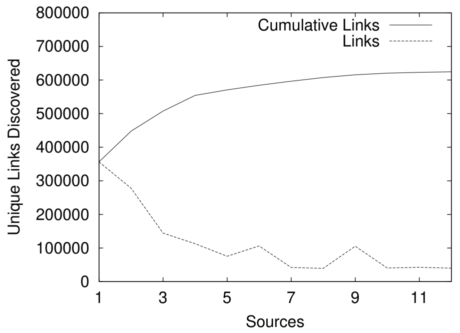

Fig. 16. Number of links discovered as sources are added (greedily)

added, similar in general shape to Figures 4 and 5. While the lack of identical experimental setup (*i.e.,* the absence of common destinations) makes it impossible to directly compare the two pairs of figures, the similarity is suggestive that a phenomenon of diminishing returns as measurement sites are added is present in the much larger dataset as well.

VI. An Information Theoretic Measure of Marginal Utility

Two elementary metrics which we defined earlier to compare a graph to one of its subgraphs are the node and edge coverage, and marginal utility of additional measurements reflects the increase in these metrics. We now return to another question closely to those posed in Section III: If we run additional traceroutes to provide further refinement to an existing topology snapshot, how can we quantitavely specify the information gained by these measurements. We provide a more precise formulation in information-theoretic terms.

A. Theory

The information content (measured in bits) revealed from the outcome s_i of an experiment S is defined as $-\log(Pr(s_i))$ [29]. For example, when there are two equally-likely outcomes of an experiment then the amount of information revealed by the outcome of the experiment is $-\log(0.5) = 1$ bit. The *expected* information content (measured in bits) gained as a result of conducting the experiment S is the *entropy* of S.

Definition 5: The *entropy* of S is given by

$$H(S) = -\sum_{\forall i} Pr(s_i) * \log(Pr(s_i))$$

The entropy of an experiment gives us a measure of the usefulness of that experiment, or equivalently, the *average amount of uncertainty* removed by the outcome of the experiment [30], [1].

Consider a sequence of n identical experiments $S^1, S^2, ..., S^n$. By identical experiments, we mean experiments that are aimed at discovering a common property. Without loss of generality, we assume that these experiments are conducted in sequential order, i.e. the results of experiment S^i are known *prior* to conducting experiment S^j, where $j > i$.

Intuitively, the marginal utility of experiment S^n can be measured in terms of the *reduction* in uncertainty that resulted from conducting this experiment. For experiment S^n, the reduction in uncertainty for outcome s_i is simply

$$-\log(Pr(s_i^{n-1})) + \log(Pr(s_i^n)) = \log\left(\frac{Pr(s_i^n)}{Pr(s_i^{n-1})}\right)$$

We define the *marginal utility* of experiment S^n as the *mean reduction* in uncertainty that resulted from conducting this experiment. This quantity can be estimated using the Kullback-Leibler (K-L) distance metric [15], which is a measure of the "relative entropy" of experiment S^n.

Definition 6: The *online marginal utility* of experiment S^n is defined to be $U(S^n)$, which is given by:

$$U(S^n) \quad = \quad \sum_{\forall i} Pr(s_i^n) \log\left(\frac{Pr(s_i^{n-1})}{Pr(s_i^n)}\right) \qquad (1)$$

where i ranges over all possible outcomes and $Pr(s_i^j)$ is the probability associated with outcome s_i after the conclusion of experiments $S^1, S^2, \ldots, S^j$.

Definition 6 quantifies the (multiplicative) gain in information (i.e. number of bits) as a result of additional experimentation. Clearly, the utility of additional experimentation diminishes as the average information gain decreases. This occurs when the additional experiments reveal no new surprises, in the sense that the probabilities of the various outcomes of an experiment converge to a fixed point.

The formulation of marginal utility given in equation 6 assumes that the evaluation of marginal utility is done in an online fashion. In other words, we evaluate the marginal utility of experiment S^n *before* conducting any additional experiments S^k, $k > n$.

An alternative formulation of marginal utility evaluates each experiment on an *ex post facto* basis, measuring each experiment's usefulness offline after all experiments have been conducted. While this evaluation cannot be performed online, it provides an estimate of marginal utility which is not biased by the ordering in which measurements are conducted .

Definition 7: The *offline marginal utility* of experiment S^n is defined to be $U^m(S^n)$, which is given by:

$$U^m(S^n) = \sum_{\forall i} Pr(s_i^m) \log(\frac{Pr(s_i^m)}{Pr(s_i^n)}) \qquad (2)$$

where i ranges over all possible outcomes and $Pr(s_i^j)$ is the probability associated with outcome s_i after the conclusion of experiments $S^1, S^2, \ldots, S^j$, and m is the total number of experiments conducted.

B. *Applications to Marginal Utility of Network Topology Measurements*

Starting from the definition provided above for gauging offline marginal utility, we focus on three network characterizations—namely, node coverage, link coverage, and the distribution of backbone node degrees. We then consider two types of experiments: one in which the set of destinations are fixed and sources are added one at a time, and one in which the set of sources are fixed and destinations are added one at a time. Considering the case of node coverage when each of our experiments adds an additional source, an *outcome* of the experiment is a subset of nodes covered. For simplicity, we express the outcome as a simple probability – the probability that a given node is covered, i.e. the node coverage. We consider these cases in detail next.

Utility of Adding New Traceroute Sources: Using the K-L distance metric as a gauge of marginal utility, we quantify the gain in information (bits) as a result of increasing the number of traceroute sources considered. Figure 17 shows the offline marginal utility for each successive experiment aimed at characterizing the probability that a node or link picked at random is discovered using traceroute experiments from i sources. This figure also shows the marginal utility of additional sources when characterizing the outdegree distribution of backbone nodes.

Pending further study on larger datasets, Figure 17 gives a preliminary indication that the marginal utility of adding new sources decreases rapidly for all three distributions, once a small constant number of sources are present.

Utility of Adding New Traceroute Destinations: Figure 18 shows the marginal utility for characterizing the probability that a link (node) picked at random is discovered using traceroute experiments from a constant number of sources to i destinations. The figure indicates that most of the information gain is achieved after considering the first 100 clients.

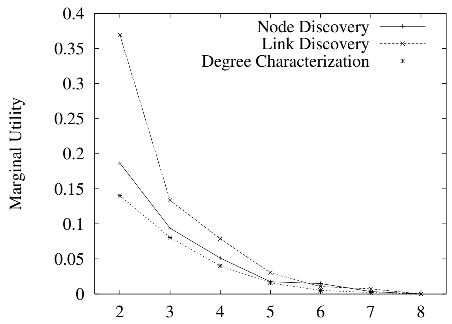

Fig. 17. Utility of additional sources (off-line)

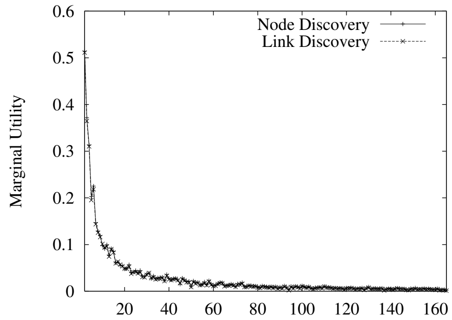

Fig. 18. Utility of additional destinations (off-line)

The above quantification of marginal utility assumed an offline approach (i.e. knowledge gained through experiment $i \leq m$ is gauged against the *cumulative* knowledge gained through all m experiments). Alternatively, one could use the online approach to incrementally quantify the utility of the last experiment performed and hence determine whether additional measurements are needed. Figure 19 shows the online marginal utility for characterizing the probability that a link (node) picked at random is discovered using traceroute experiments from a constant number of sources to i destinations. Unlike the offline K-L distance metric, the online K-L distance metric is not monotonically decreasing. An increase in the K-L distance metric for experiment i is indicative of an experiment with a "surprisingly" large information content (relative to the cumulative knowledge gained up to that experiment). For example, an added destination may result in the discovery of an unexpectedly large number of nodes/links since traceroute experiments to that clients may for example unveil a new AS. Despite this non-monotonicity, the *magnitude* of the "surprises" unveiled by the on-line K-L dis-

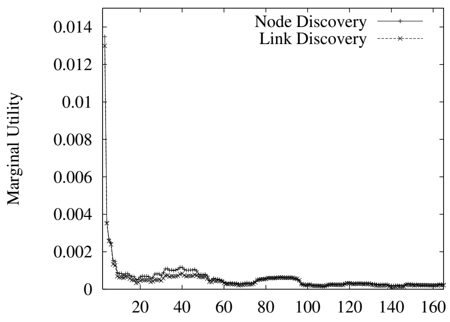

Fig. 19. Utility of additional destinations (on-line)

tance metric seem to decrease monotonically.

Comparative Utility of Adding Sources versus Adding Destinations: One of the attractive aspects of information theoretic measures of marginal utility is that they enable comparison of marginal utility (1) across multiple distributions (e.g. link vs node vs degree discovery as was done in Figure 17) and (2) across multiple experimental setups (e.g. adding new sources vs adding new destinations). To exemplify the latter of these cases, consider the question of comparing the utility of adding traceroute sources to the utility of adding traceroute destinations. Comparing the results illustrated in Figure 17 to those illustrated in Figure 18 reveals that doubling the number of destinations from 80 to 160 while holding the number of sources fixed at 8 (a total of 640 additional traceroute experiments) yields a marginal utility that is approximately equivalent to that resulting from increasing the number of sources from 7 to 8 while running traceroute to all 1277 destinations (a total of 1,277 additional traceroute experiments).

VII. Conclusions and Future Work

In principle, it should be possible to gain considerable insight into the conditions and configurations in the core of the Internet, given a sufficient array of measurement points located in end systems. This concept has been called "network tomography" because each measurement point sees a "projection" of the Internet's resources in a manner specific to its location.

While the concept of network tomography is attractive and in keeping with the design philosophy of keeping network-internal components as simple as possible, so far it has not been clear how extensive a measurement infrastructure is needed in order to see a large fraction of the network from its edges. In the absence of precise knowledge, the prevailing wisdom in Internet measurement has seemed to be "more is better." In this paper we have taken a step toward developing a more refined understanding of this problem. We have concentrated on the problem of discovery of basic Internet components — links and nodes (end systems and routers). We assumed the common measurement situation in which active measurement sites are scarce, but passive targets for measurement probes are relatively plentiful.

Our preliminary results indicate that the marginal utility of additional measurement sites declines rapidly even after the first two sites. This is evident in the discovery of nodes, of links, and of node degree distribution. We considered the aggregation of all datasets to be the most complete picture available; in each case (nodes, links, and node degree distribution) a vast majority of the information present in the aggregated dataset was present in the first two or three datasets alone. On the other hand, conducting additional measurements invariably provided a more complete picture of the entire topology.

Our conclusions are unavoidably sensitive to the particular choice of measurement sites to which we had access, and we believe that further measurements are warranted to reinforce our conclusions. However we believe that these results shed light on how typical IP routes pass through the Internet, and show that a majority of routes tend to pass through a relatively well-defined "switching core." We also note that `traceroute` measurements are just one technique for studying the marginal utility and scaling questions we pose here; numerous other datasets might also apply well, albeit with different pros and cons.

Finally, discovery of nodes and links in an internetwork provides only the most basic topographical information about the network. Questions about marginal utility could be framed in the context of richer network characteristics, such as studying the marginal utility of additional measurements to characterize the distribution of packet loss in the network. We hope that this paper, which we believe to be the first to rigorously quantify the marginal utility of network measurements, will eventually see broad application to a range of important problem domains in network measurement.

VIII. Acknowledgements

The data used in this research was collected as part of CAIDA's skitter initiative, http://www.caida.org. Support for skitter is provided by DARPA, NSF, and CAIDA membership. The authors would like to thank kc claffy, Amy Blanchard and Edouard Lagache from CAIDA for making Skitter trace data available for this study.

The authors would also like to thank Jennifer Rexford for her valuable help shepherding this paper and the anonymous IMW reviewers for their suggestions for improving the paper.

REFERENCES

[1] N. Abramson. *Information Theory and Coding*. McGraw-Hill, 1963.

[2] A. Adams, J. Mahdavi, M. Mathis, and V. Paxson. Creating a Scalable Architecture for Internet Measurement. In *Proceedings of INET '98*, 1998.

[3] P. Barford and M. Crovella. Measuring web performance in the wide area. *Performance Evaluation Review*, August 1999.

[4] H-W Braun and K.C. Claffy. Global ISP interconnectivity by AS number. http://moat.nlanr.net/AS/.

[5] A. Broido and K. Claffy. Connectivity of IP Graphs. In *Proceedings of SPIE ITCom '01, Scalability and Traffic Control in IP Networks*, August 2001.

[6] H. Chang, S. Jamin, and W. Willinger. On Inferring AS-Level Connectivity from BGP Routing Tables. Available at http://topology.eecs.umich.edu/archive/inferbgp.ps.

[7] H. Chang, S. Jamin, and W. Willinger. Inferring AS-level Internet Topology from Router-level Traceroutes. In *Proceedings of SPIE ITCom '01, Scalability and Traffic Control in IP Networks*, August 2001.

[8] B. Cheswick. Internet mapping project. http://www.cs.belllabs.com/who/ches/map/.

[9] J. Chuang and M. Sirbu. Pricing multicast communication: A cost-based approach. In *Proceedings of INET '98*, 1998.

[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In *ACM SIGCOMM*, pages 251–62, Cambridge, MA, September 1999.

[11] Cooperative Association for Internet Data Analysis (CAIDA). The Skitter project. http://www.caida.org/Tools/Skitter.

[12] P. Francis, S. Jamin, V. Paxson, L. Zhang, D. Gryniewicz, and Y. Jin. An Architecture for a Global Host Distance Estimation Service. In *Proceedings of IEEE INFOCOM '99*, March 1999.

[13] R. Govindan and A. Reddy. An Analysis of Internet Inter-Domain Routing and Route Stability. In *Proceedings of IEEE INFOCOM '97*, April 1997.

[14] R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery. In *Proceedings of IEEE INFOCOM '00*, April 2000.

[15] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, 1990.

[16] T. Griffin and G. Wilfong. An Analysis of BGP Convergence Properties. In *ACM SIGCOMM*, pages 277–88, Cambridge, MA, September 1999.

[17] V. Jacobson. traceroute. ftp://ftp.ee.lbl.gov/traceroute.tar.Z, 1989.

[18] S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. On the Placement of Internet Instrumentation. In *Proceedings of IEEE INFOCOM 2000*, March 2000.

[19] J.-J. Pansiot and D. Grad. On Routes and Multicast Trees in the Internet. *ACM Computer Communication Review*, 28(1):41–50, January 1998.

[20] V. Paxson. End-to-End Internet Packet Dynamics. In *Proceedings of ACM SIGCOMM '97*, Cannes, France, September 1997.

[21] V. Paxson. *Measurements and Analysis of End-to-End Internet Dynamics*. PhD thesis, University of California Berkeley, 1997.

[22] V. Paxson. End-to-End Routing Behavior in the Internet. *IEEE/ACM Transactions on Networking*, pages 601–615, December 1998.

[23] G. Phillips, S. Shenker, and H. Tangmunarunkit. Scaling of multicast trees: Comments on the Chuang-Sirbu scaling law. In *Proceedings of ACM SIGCOMM '99*, September 1999.

[24] The SCAN Project. http://www.isi.edu/scan/.

[25] The Surveyor Project. http://www.advanced.org/, 1998.

[26] Internet Traffic Report. http://www.internettrafficreport.com/.

[27] The Internet Weather Report. http://www3.mids.org/weather/index.html, 2000.

[28] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson. End-to-end Effects of Internet Path Selection. In *ACM SIGCOMM*, pages 289–300, Cambridge, MA, September 1999.

[29] C. Shannon. A Mathematical Theory of Communication. *Bell Systems Technical Journal*, 47:143–157, 1948.

[30] C. E. Shannon and W. Weaver. *Mathematical Theory of Communication*. University of Illinois Press, 1949.

[31] R. Siamwalla, R. Sharma, and S. Keshav. Discovering Internet Topology. Technical report, Cornell University Computer Science Department, July 1998. http://www.cs.cornell.edu/skeshav/papers/discovery.pdf.

[32] H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin. The Impact of Policy on Internet Paths. In *Proceedings of IEEE INFOCOM '01*, April 2001.

[33] E. W. Zegura, K.L. Calvert, and M. J. Donahoo. A Quantitative Comparison of Graph-based Models for Internetworks. *IEEE/ACM Transactions on Networking*, pages 770–783, December 1997.