On the Marginal Utility of Deploying Measurement Infrastructure*

Paul Barford barford@cs.bu.edu
Azer Bestavros best@cs.bu.edu
John Byers byers@cs.bu.edu
Mark Crovella crovella@cs.bu.edu

Computer Science Department
Boston University
Boston, MA 02215

July 2000

Abstract

An emerging strategy to gain insight into the conditions and configurations in the core of the Internet is the use of end-to-end measurements from a set of distributed measurement points. As the number of measurement points increases, so does the number of paths through the network that can be studied, which is an argument for a "more is better" approach to wide-area measurement. However, basic questions about the *utility* of increasing the number of measurements and measurement sites have not yet been addressed.

In this paper, we present a more refined and quantifiable understanding of the marginal utility of performing wide-area measurements. We focus on problems in Internet topology discovery, namely, discovering the set of nodes and links which comprise the Internet backbone, discovering the degree distribution of these nodes, and classifying nodes according to their role. We provide a model for how the discovery process scales as the number of measurements and measurement sites increase and validate this model against a set of *traceroutes* run across the Internet from 17 measurement sites run to 60,000 destinations. We characterize the topology in terms of nodes, links, node degree distribution, and distribution of end-to-end flows using statistical and information-theoretic techniques. We show that the utility of adding sources beyond the second source goes down significantly from the perspective of interface, node, link and node degree discovery. We 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.

1 Introduction

An emerging strategy to gain insight into the conditions and configurations in the core of the Internet is the use of end-to-end measurements from a set of distributed measurement points. This approach has been dubbed "network tomography" because each measurement point observes a "projection" of the Internet specific to its location. Typically, increasing the number of vantage points has been viewed as beneficial, since this increases the number of potential paths along which measurements can be performed. However, the cost and complexity of deploying wide-area measurement infrastructure as well as performing data collection and analysis is considerable, and the basic question of the *utility* of

*This work was partially supported by NSF research grants CCR-9706685 and ANIR-9986397.

performing additional measurements has not yet been addressed. In this paper, we provide an analytic and information-theoretic framework in which to study this question.

The particular problem domain within which we discuss wide-area measurement utility is that of mapping the Internet's physical interconnection topology. We discuss topology from the perspectives of nodes, links, and node degree. We focus on three general aspects of the topology mapping problem, which are all closely related. The first is the effects on topology mapping and measurement utility before and after routers which advertise multiple interfaces are resolved to a single node. The second is the effects on marginal utility of node and link discovery when nodes are classified into different groups based on their location in the Internet. The third is a characterization of link utilization on all routes traced through the backbone.

Performing experiments to exhaustively map out the Internet backbone is an ambitious and tedious process which becomes increasingly challenging as the size of the backbone continues to scale. One technique for generating approximate maps is to run route tracing tools such as `traceroute` [13] across a large number of end-to-end paths through the network, and cross-referencing the results to build up a view of the topology. However, it is difficult to quantify the effectiveness of such an experiment, as one cannot be certain how many nodes and links were *not* traversed in the course of a set of measurements. Moreover, the set of nodes and links actually traversed may not be representative of the entire topology. In this paper, we study the effectiveness of this approach. Since we assume that our route traces provide correct information, i.e. never reveal nodes and links which are not actually present, the approximation of the exact map that we build is *monotone* in the sense that the quality of our approximation never decreases as we add more measurements. However, as our experiments demonstrate, monotonicity is not necessarily present in addressing the question of how well we can characterize the structure of the Internet (such as the distribution of node degree). In this situation, as additional measurements are taken it may be the case that those additional measurements present us with misleading information, degrading the quality of our characterization.

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 [7]. 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. This technique is discussed in detail in Section 3. 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 familiar picture of the Internet 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.

Our results for marginal utility are presented using both statistical and information-theoretic techniques. The statistical results are sufficient to evaluate the marginal utility of adding measurement points in an Internet topology study. We show that the marginal utility of adding additional source nodes beyond 2 decreases significantly. We show that the marginal utility of adding destination nodes beyond 200 is roughly constant and yields approximately 3 new nodes and 4 new links. The information-theoretic methods we develop in this paper allow us to specify distance metrics between the measured topology and the actual topology of the Internet. Our results show that the marginal utility for node,

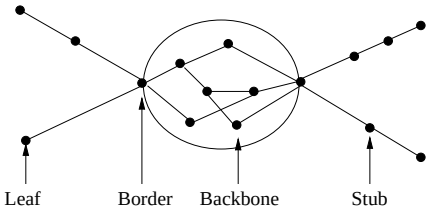

Figure 1: Classification of Internet nodes

link and node degree discovery essentially goes to zero when more than 5 sources are used (although the discovery of "new" AS's through the addition of sources or destinations can cause fluctuations in this metric).

In addition to the marginal utility results we also analyze the paths taken through the backbone nodes in our graph of the Internet. We focus specifically on the link utilization of paths through the backbone. We count the number of times specific backbone links appear on each unique path through the backbone. We show that the distribution of link utilization in the backbone approximately follows a power law indicating that some links in the backbone are very heavily utilized while the majority are not.

The rest of the paper is organized as follows. In Section 2 we describe related analytical work in evaluating the effectiveness of deploying wide-area measurement infrastructure with a focus on topology mapping. In Section 3, 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 4. We define the information-theoretic tools and results from their application to the data in Section 5. We present backbone link utilization analysis in Section 7. We summarize, conclude and discuss future work in Section 8.

2 Related Work

A number of research groups have focused on generating maps of the Internet using route tracing tools including [25, 7]. There are also a number of repositories of Internet mapping information including [17, 16, 18] which list projects, data sets and topological visualization tools. Recent studies have used logs collected in the wide-area by BGP-capable routers to perform post-hoc analysis on the quality of routes produced using BGP. These studies include [12, 33, 30]. At a slightly higher level, AS (Autonomous System) connectivity has been studied in [9, 4] which gives a different perspective on the notion of Internet topology.

Work by Govindan [26, 10] outlines heuristic techniques for generating complete domain maps. One of the challenges in this area go far beyond the capabilities of traceroute, and lie in mapping out the nooks and crannies of regions in 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 interfaces to a single node.

Jamin et al [14] 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 [8]. 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 [6]). However, they do not attempt to quantify the marginal value of information gained as measurements are added, nor do they qualitatively discuss the characterization of the topology they can obtain with a small number of measurements.

Paxson [20, 21] 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 [22] 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 [27]. Another piece of generally related work are the Internet weather reports such as [29, 28]. These are general compilations of the packet loss and round trip time measurements from Internet monitoring boxes deployed in the wide-area.

3 Experimental Design

3.1 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. The Skitter infrastructure consists of 16 source nodes deployed around the world; we received data from 8 of those nodes. The source nodes for our data were located in New Zealand, Japan, Singapore, San Jose (2), Canada (Ottawa), London, and Washington DC. Each source node sends traceroute-like probes to destination nodes located all over the world. Destination nodes are all Web servers. Our data set contains results from traces run to 1277 destinations. On average, probes are sent to each destination once every 30 minutes. While it is not clear how sites for sources are selected in Skitter, they state that destinations are selected at random from a "crawl of IP address space" [7].

One note on our data set: one of the sources (one of the San Jose nodes) was not functioning correctly during the time that our traces were taken. The result is that this source traced routes to only 184 of the destinations, hence the number of paths traced from this source is much smaller than those traced from the other sources.

3.2 The Classification Procedure

We model the Internet as a directed graph $G = (V, E)$, vertices and edges of which we do not know in advance. Using the terminology of Zegura et al [34] 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 – but we emphasize that correctness of classification is not an important aspect of our study – we strive only to provide a reasonable differentiating mechanism so that we can focus on how well our measurements uncover a central region of the topology. In the CAIDA study, the evidence is the result of the following set of measurements: a fixed set of $k = 8$ measurement sites (sources) $M \subseteq V$ ran *traceroute* to a fixed set of $S \subseteq V$ destinations once a day for the course of several weeks[1]. 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. From this collection of traces, we observe a set of paths through G for all pairs of endpoints (m, s) such that $m \in M, s \in S$. This collection of paths forms a graph G' which is a subgraph of G. Even using a relatively small number of measurement sites, a clear distinction between backbone links and stub links in 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 and 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 our backbone, and we classify it as such. Unlabelled nodes which adjoin at least one stub link are now 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.

3.3 Coverage

For a given graph $G = (V, E)$ and a subgraph $G' = (V', E')$ of G, we define the *node coverage* of G by G' to be the ratio $\frac{|V'|}{|V|}$ and similarly, the *edge coverage* of G by G' as the ratio $\frac{|E'|}{|E|}$. 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. However, this approach is infeasible, as the exact makeup of the graph which comprises this backbone is not known a posteriori. As a quantifiable alternative, 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. Again, our most significant findings are with respect to the coverage of the backbone.

[1]These traces are actually run every day from more sites, but we do not have access to the full datasets

3.4 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 is similar to the one 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 we suspect that a router has two interfaces I_1 and I_2, we 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 respond 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 immediately be rendered useless. We have no way of estimating the likelihood of this event; however, the frequency of success of detecting duplicate interfaces gives us some confidence that routers do in fact behave according to specification.

3.5 Accuracy of Classification

One central aspect of node classification is the speed and 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. As we increase the number of sources in Figure 2, our classification procedure increases in accuracy. In particular, 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 classification becomes highly reliable after only three or four measurement sites (vantage points) are used.

3.6 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 not be unbiased, since the behavior of the routing algorithms in the Internet are non-deterministic. In particular, the prevalent route from point A to point B may not be along the shortest path (in hops) from A to B. Diverse factors such as hop-normalized routing metrics, variability in network load, and economic agreements between autonomous systems (AS's)

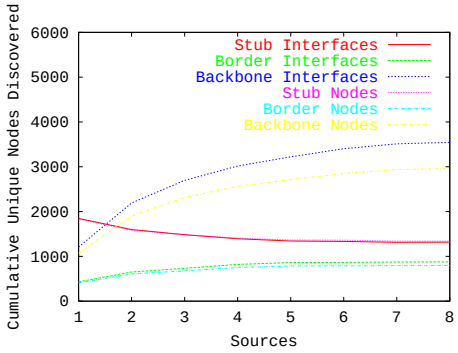

Figure 2: Class of nodes and interfaces discovered as sources are added (greedily) when classification is not known a priori.

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

Figure 3: Distribution of observed interface density across routers. (Rtrs)

cause the routes chosen to be quite different than an observer with access only to topology information might expect.

4 Results

The results in this section are divided into five parts: (1) the results of interface disambiguaton run on all nodes in the 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.

4.1 Results of running the disambigution 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 on the return path to B.U. 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

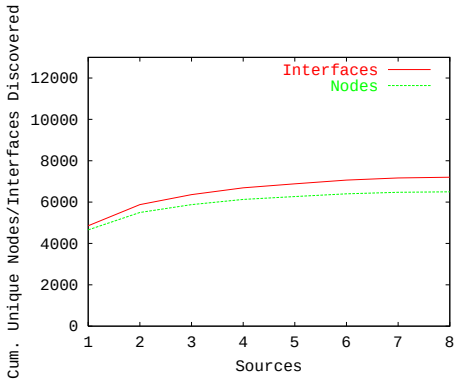

Figure 4: Number of nodes discovered as sources are added (greedily)

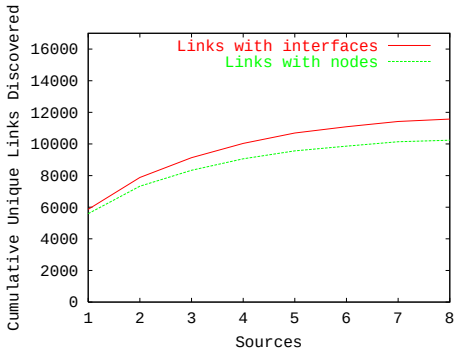

Figure 5: Number of links discovered as sources are added (greedily)

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 $\frac{871}{5763} = 15.1\%$.

4.2 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 sharply 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 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

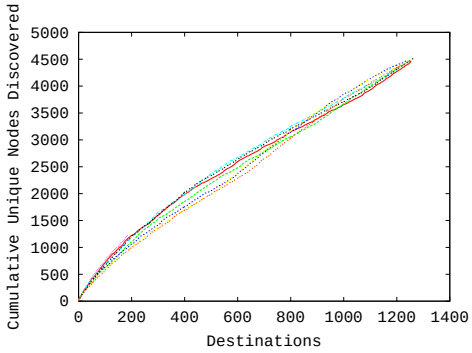

Figure 6: Number of nodes discovered as destinations are added (randomly). Each line is for a single source

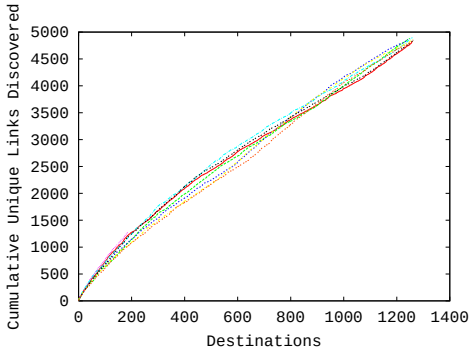

Figure 7: Number of links discovered as destinations are added (randomly). Each line is for a single source

added.

4.2.1 Contour Plot

Recall the example motivated in the introduction, in which you have the goal of taking measurements over a set of paths which cover at least l distinct links in the backbone. Then suppose that you have the capability to deploy k sources which can each run *traceroute* to n common destinations. Then, for which values of n and k can you achieve your objective, and how many links are covered as a function

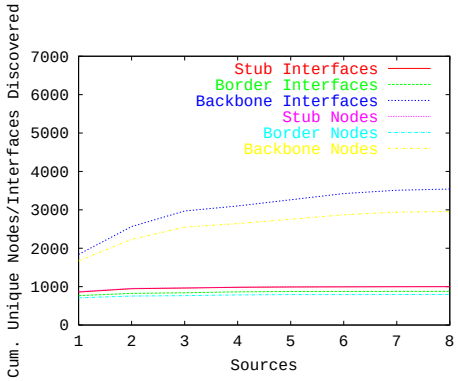

Figure 8: Class of nodes and interfaces discovered as sources are added (greedily) when classification is known a priori.

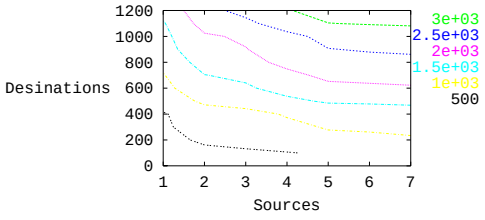

Figure 9: Backbone node discovery as both sources and destinations are varied

of n and k? In the contour plot diagrams below, we plot the answer to that question. The values of n and k 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 equal utility. 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 regards 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. While this assumption may not be realized in general, it is certainly quite close to the truth for our datasets in the context of observing only the Internet backbone.

The surprising result shown in Figure 9 [2] is that our initial results indicate that the contours are more similar to lines of the form $x+y = k$ than to hyperbolas of the form $x \times y = k$. This further implies 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 using destinations, rather than sources). The bottom line is that taking more measurements is better, as is deploying more infrastructure, but the relative benefit is much more comparable than we thought (while the relative cost differs by orders of magnitude).

4.3 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 heaviness in the tail may actually diminish somewhat as the number of measurements increase [3] Another interesting point is that in the RMSE plot in Figure 12, the error actually increases after source 6 is added. This exemplifies a characteristic of the node degree distribution characterization experiments – the quality of the characterization over time may be non-monotone.

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

[3] 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

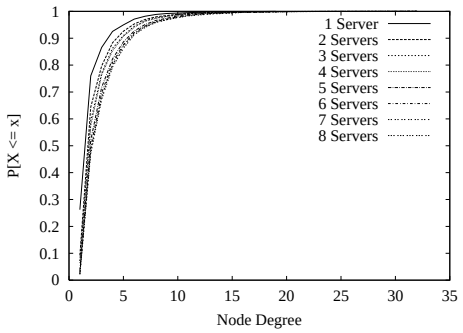

Figure 10: CDF of backbone node degree as sources are added

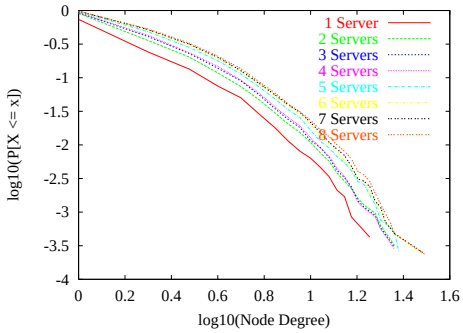

Figure 11: Tail of CDF of backbone node degree as sources are added

We conducted a similar analysis considering how the addition of destination nodes effects 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 distribution stays relatively constant as destination nodes are added, the tail weight increases as destination nodes are added.

set-up. We plan on re-running this experiment with other orderings of the sources as part of our further investigation.

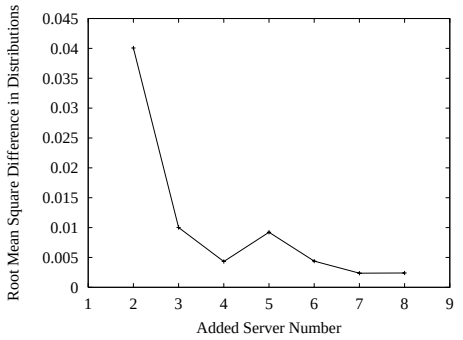

Figure 12: Root mean squared error difference in backbone node degree distributions

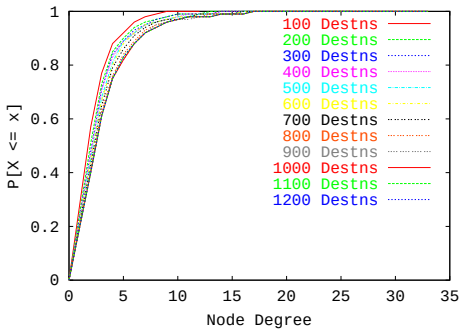

Figure 13: CDF of backbone node degree as destinations are added

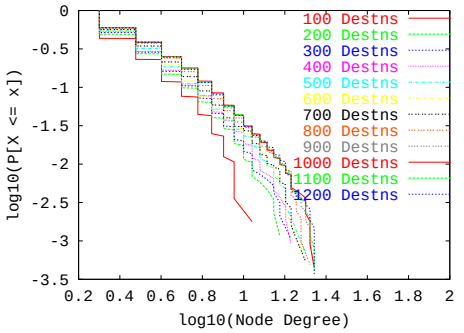

Figure 14: Tail of CDF of backbone node degree as destinations are added

5 An Information Theoretic Measure of Marginal Utility

We now consider another question closely related to the one studied in the previous section: If we run traceroute between a set of k sources and a set of n destinations, and if we are given the topology G of the underlying backbone in advance, how can we specify appropriate distance metrics between our measured topology and the actual topology?

Two elementary metrics which we study when comparing our measured graph to the actual graph are the node distance and link distance between them, where the node distance is simply the size of the set $V_G \setminus V_M$. and the link distance is the size of the set $E_G \setminus E_M$. Intuitively, the *marginal utility* of an additional measurement with respect to either of these metrics will reflect the reduction in distance to the goal. We state this more precisely in information-theoretic terms below.

The information content (measured in bits) revealed from the outcome s_i of an experiment S is defined as $-\log(Pr(s_i))$ [31]. For example, if there are two equally-likely outcomes of an experiment (e.g. coin toss), then the amount of information revealed by the outcome of the experiment (e.g. heads or tails) is $-\log(0.5) = 1$ bit. Moreover, the *expected* information content (measured in bits) gained as a result of conducting the experiment S is defined as the *Entropy* of S and 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* associated with the outcome of the experiment [32, 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 (i.e. reducing a common uncertainty).

Without loss of generality, we assume that these experiments were conducted in a sequential order (i.e. the results of experiment S^i were 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 (KL) distance metric [11], which is a measure of the "relative entropy" of experiment S^n. Formally, we denote the *marginal utility* of experiment S^n by $U(S^n)$, which is given by:

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

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 $S^1, S^2, ...S^j$.

Equation 1 quantifies the (multiplicative) gain in information (i.e. number of bits) as a result of additional experimentation. Obviously, the utility of additional experimentation will diminish as the average number of bits gained decreases. This occurs when the additional experiments reveal no "new surprises", in the sense that the probabilities of the various outcomes $s_1, s_2, ...$ do not change much. In particular, $U(S^n)$ approaches zero when $Pr(s_i^{n-1})$ approaches $Pr(s_i^n)$ for all values of i.

The formulation of marginal utility given in equation 1 assumes that the evaluation of marginal utility is done in an "on-line" fashion. In other words, we evaluate the marginal utility of experiment S^n *before* conducting any additional experiments S^n, $k > n$.

An alternative formulation of marginal utility would "revise" the estimated marginal utility of experiment S^n based on the results of later experiments S^k, $k > n$. This leads us to the definition of the "off-line" estimated marginal utility, which we denote by $U^m(S^s)$.

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

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 $S^1, S^2, ...S^j$, and m is the total number of experiments conducted.

One of the attractive aspects of the information theoretic marginal utility metric given by Equation 2 is that it enables comparison of marginal utility (1) across multiple distributions (e.g. link vs node discovery) and (2) across multiple experimental setups (e.g. adding new sources vs adding new destinations). We illustrate this below.

Utility of Adding New Traceroute Sources: We use 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. We focus on three network characterizations—namely, node coverage, link coverage, and the distribution of backbone node degrees. Figure 15 shows the marginal utility function for each successive experiment aimed at characterizing the probability that a link (node) picked at random is discovered using traceroute experiments from i sources. Figure 15 also shows the marginal utility of additional sources when characterizing the outdegree distribution of backbone nodes.

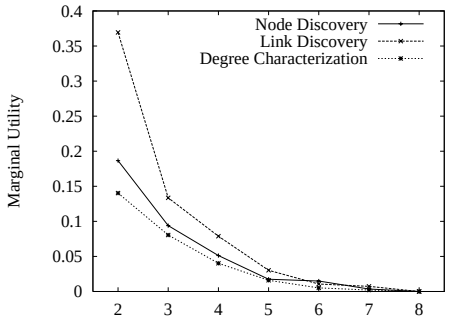

Figure 15: Marginal utility of additional sources

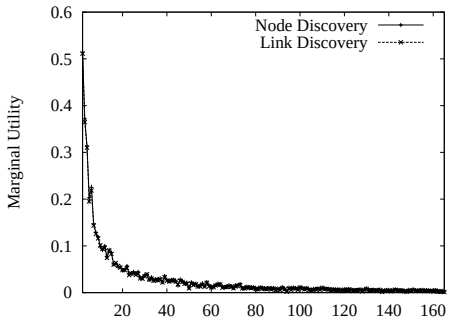

Figure 16: Marginal utility of additional destinations

Figure 15 indicates that the marginal utility of adding new sources decreases rapidly for all three distributions. Specifically, for backbone outdegree characterization, the marginal utility of increasing the number of sources beyond 5 approaches zero. For node and link discovery, the marginal utility of increasing the number of sources beyond 6 approaches zero.

Utility of Adding New Traceroute Destinations: Figure 16 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 (namely 8) to i destinations. The figure indicates that most of the information gain is achieved after considering the first 100 clients.

The above quantification of marginal utility assumed an "off-line" 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 "on-line" approach to incrementally quantify the utility of the last experiment performed and hence the need for additional experiments (or lack thereof). Figure 17 shows the "on-line" 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 "off-line" K-L distance metric, the "on-line" 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 distance metric seem to decrease monotonically.

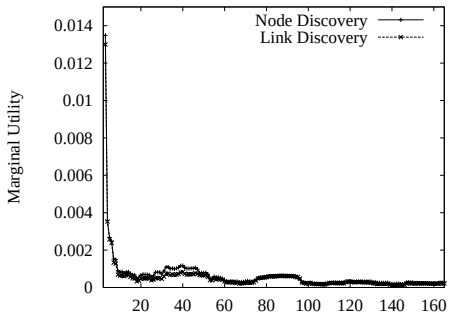

Figure 17: Marginal utility of additional destinations

6 Scaling Laws for Internet Tomography

The network discovery problems we consider fall into a general class of problems of which some special cases have been extensively studied by graph theorists and networking researchers. Consider a network topology represented by an undirected graph $G = (V, E)$ in which $|V| = n$. Suppose that k sources and m destinations are chosen uniformly at random from the vertex set of this graph. We can then consider the fraction of the total graph that is covered by the set of shortest paths from the sources to each of the destinations.

In general, let $v_G(k, m)$ denote the expected number of nodes covered by paths spanning k sources and m destinations and let $e_G(k, m)$ denote the expected number of edges covered by such a set of paths. The rate at which these functions scale with respect to k and m give insight into the (expected) marginal utility of additional measurements and measurement sites for a given graph G. Considerable attention has been devoted to such scaling laws for the random graph $G_p(N)$, where edges are chosen uniformly and equiprobably with probability p. However, it has been widely documented [34, 15] that random graphs do not capture essential features of Internet topologies.

For the special case of $k = 1$, Chuang and Sirbu [5] considered the scaling law in the context of multicast trees. Their work, and subsequent work by Phillips, Shenker and Tangmunarunkit [24], 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 H_G is a constant reflecting the average path length in G. We consider cases in which $k > 1$ and where m can be moderately large (note that an interesting special case arises when $m = n$). This corresponds to "overlays" of "projections" from multiple vantage points.

7 Backbone Link Utilization Analysis

Routing protocols and policies between AS's determine the paths that packets take through the Internet. We are interested in the amount of link sharing that these protocols and policies induce between paths taken through the backbone nodes in our graph of the Internet. We focused specifically on the paths taken through the *backbone* (as we define it using our classification method i.e. paths taken between border nodes) since clearly there will be a high degree of sharing of stub links - especially between sources and border nodes. Our analysis consists of counting the number of times specific backbone links appear on each unique path taken through the backbone between border nodes. We exclude links between backbone nodes and border nodes since these could well be biased by the nature of the routes from the sources to a majority of the destinations.

There were 8,859 unique paths taken through the backbone of our graph. We show that the distribution of link utilization in Figures 18 and 19. The figures show that backbone link utilization

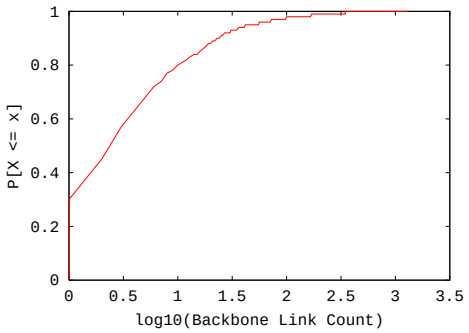

Figure 18: CDF of backbone link utilization for backbone nodes.

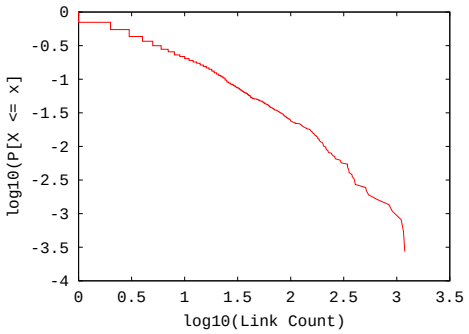

Figure 19: Tail of the CDF of backbone link utilization for backbone nodes.

approximately follows a power law indicating that some links are very heavily utilized in the backbone while the majority are not. In fact 57% of the 3,674 links in our backbone appear on 3 or fewer paths while the most frequently used link appears on 1,279 paths.

Our interest in this question stems from the many implications of link sharing in the backbone. If we assume that the paths taken between sources and destinations in our data set are representative of paths taken by ordinary clients in the Internet, then this result indicates that many paths utilize a small number of links in the backbone. Paxson shows in [23] that paths through the Internet tend to be fairly stable thus this condition of a small number of links being used by many clients should be fairly persistent (although we have done no such temporal analysis to confirm this conjecture). This condition means that many routes are vulnerable to failures of a small number of links. It also has implications for congestion control since bursts of traffic from clients can lead to loss conditions for a large number of flows.

8 Conclusion

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 a more refined answer to this question. We have concentrated on the 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 most surprising result is that the marginal utility of additional measurement sites declines rapidly 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.

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 our results hold out hope that although the Internet's structure is vast and dynamic, it is possible nonetheless to obtain useful pictures of its internal structure via measurements from a limited number of locations.

9 Acknowledgements

The authors would like to thank KC Claffy, Amy Blanchard and Edouard Lagache from CAIDA for making Skitter trace data available for this study.

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

[6] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On Power-Law Relationships of the Internet Topology. In *ACM SIGCOMM*, pages 251–62, Cambridge, MA, September 1999.

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

[8] 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.

[9] R. Govindan and A. Reddy. An analysis of internet inter-domain routing and route stability. In *Proceedings of IEEE INFOCOM '97*, April 1997.

[10] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *Proceedings of IEEE INFOCOM '00*, April 2000.

[11] Robert M. Gray. *Entropy and Information Theory.* Springer-Verlag, 1990.

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

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

[14] 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.

[15] A. Medina, I. Matta, and J. Byers. On the Origin of Power Laws in Internet Topologies. *ACM SIGCOMM Computer Communication Review (CCR)*, pages 18–28, April 2000.

[16] The Geography of Cyberspace Directory. http://www.geog.ucl.ac.uk/casa/martin/, 2000.

[17] An Atlas of Cyberspaces. http://www.cybergeography.com/atlas/atlas.html, 2000.

[18] Russ Haynal's ISP Page. http://www.navigators.com/isp.html.

[19] Jean-Jacques Pansiot and Dominique 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 Routing Behavior in the Internet. In *Proceedings of ACM SIGCOMM '96*, Palo Alto, CA, August 1996.

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

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

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

[24] 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.

[25] Internet Mapping Project. http://www.cs.belllabs.com/who/ches/map/, 1999.

[26] The SCAN Project. http://www.isi.edu/scan/, 1999.

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

[28] Internet Traffic Report. http://www.internettrafficreport.com/, 2000.

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

[30] 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.

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

[32] Claude E. Shannon and Warren Weaver. *Mathematical Theory of Communication.* University of Illinois Press, 1949.

[33] 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.

[34] 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.