Learning Network Structure from Passive Measurements

Brian Eriksson
UW-Madison
bceriksson@wisc.edu

Paul Barford
UW-Madison
pb@cs.wisc.edu

Robert Nowak
UW-Madison
nowak@ece.wisc.edu

Mark Crovella
Boston University
crovella@cs.bu.edu

ABSTRACT

The ability to discover network organization, whether in the form of explicit topology reconstruction or as embeddings that approximate topological distance, is a valuable tool. To date, network discovery has been based on active measurements. However, it is feasible to envision passive discovery of network topology and distance, simply by monitoring packet traffic. Unfortunately, the lack of explicit control over the choices of which endpoints are measured means that passive network discovery must deal with the problem of missing information. We consider one such example, namely reconstructing embeddings and some network structure information from unwanted network traffic captured at a set of honeypots. We develop a number of algorithms for reconstruction of missing measurements. Our algorithms use insights derived from the known topology of the Internet as well as local imputation techniques from approximation theory. We characterize the degree to which missing information can be reconstructed and show that a limited but useful amount of reconstruction is possible, allowing the recovery of network embeddings and some topological relationships from passively collected data.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network Monitoring*

Keywords

Topology, Embedding, Measurement, Inference, Imputation

1. INTRODUCTION

Discovering network topology is useful for many purposes. Knowledge of the interconnection pattern of routers can be used to improve replica placement, overlay configuration, troubleshooting, model construction, and routing analysis, among other activities. However topology measurement has classically been approached using active techniques. Most topology measurement uses `traceroute` or one of its variants to actively probe intermediate nodes along the path from a probe source to a destination.

Active probing has strengths and weaknesses. One strength with respect to topology measurement is that it allows some control over the choice of paths that are measured. A weakness of active measurement, especially for topology discovery, is that it puts a significant traffic load on the measured paths. In fact, research has focused on the question of how to reduce the traffic load placed on networks by active measurement [6].

A number of advances in Internet measurement mean that it is now conceivable to capture information about Internet topology in a passive manner. First of all, honeypots and other passive traffic collectors can obtain large quantities of traffic at low expense. Furthermore, it is possible to infer the number of hops between a sender and receiver using only information in a packet header. Using the technique described in [11], one can infer the number of hops between the honeypot monitor and the host. This inference is made based on the fact that (i) there are only a few initial TTL values used in popular operating systems (*e.g.,* 64 for most UNIX variants, 128 for most Microsoft variants and 255 for several others), and (ii) typical hop counts for end-to-end paths are far less than the differences between the standard TTL values. Thus, hop count can be inferred by rounding the TTL up to the next highest initial TTL value and then subtracting the initial TTL.

Putting together passive traffic collection with hop count inference provides a rich source of topology-related information about the Internet. However there are a number of hurdles to be cleared before such information can be reliably converted into an actual topology. A key hurdle arises from the intrinsic nature of passive traffic measurement. When hop count data is collected passively there is little control over which paths are measured, and so invariably some data is missing.

For example, our collection setup is based on the use of honeypots. Honeypots monitor routed but otherwise unused address space, so all traffic directed to these monitors is unwanted and almost always malicious. When a single source sends traffic to multiple honeypots, we can obtain useful information. Consider a set of N honeypots. If packets from a source S_1 are observed at all N honeypots, we obtain an N-dimensional vector of hop counts. Such a vector can be useful for constructing a network embedding that provides information about topological "closeness." However if another source S_2 is observed at only $N-1$ honeypots, we

IMC'07, October 24-26, 2007, San Diego, California, USA.

obtain a hop count vector with a missing component. A direct embedding of this vector in the same space as that of S_1 is problematic. It would be useful to be able to infer or replace the missing hop count data so as to place both nodes in the same N dimensional space.

In this paper we address two key questions that need answers before this kind of passive topology measurement can be considered practical. First, how can we infer or replace missing hop count data? And second, given a large collection of hop count data, how can we extract topological information about the structure of the Internet?

Our results provide initial, albeit partial answers to each of these questions. First, we describe and evaluate methods whereby missing hop count information can be replaced. To solve this we adopt an interpolation approach. We develop a distance metric based on an understanding of Internet topology, and we show that it performs better than traditional distance metrics as the basis for interpolating hop counts.

Second, we show how our distance metric can be used to infer a partial topology. In particular, we are able to infer the fact that certain nodes share a common router along the paths to all measurement points. This allows us to cluster nodes topologically. Although conclusive validation is difficult, we show evidence that this clustering corresponds to real topological information about the Internet.

2. RELATED WORK

Internet structure and topology have been widely studied over the years (*e.g.,* [1, 3, 9, 18]). Prior studies of router level topologies have been conducted exclusively on data sets derived from active probe-based measurements. For example, a well known project using active data collection for topology measurement is Skitter [4]. While the end goals of our work are similar to past efforts on topology analysis, our methodology differs since it is based only on passive measurements at honeypots.

Related to Internet topology is the notion of Internet distance. Most often, distance estimation focuses on minimum packet latency and is important to end-to-end performance evaluation. However Internet distance estimation can also be concerned with geography, congestion, or routing. Early work on infrastructure support to estimate Internet distance from partial measurements can be found in [7, 8]. Another related idea is IP geolocation *i.e.,* finding the geographic location of an Internet host [10, 12]. In our work we focus on topology measurement which can give insight into other kinds of Internet distance.

Another aspect of our work concerns the use of hop count vectors as a kind of network embedding or coordinate system. Network embeddings have been most often proposed as mechanisms for latency estimation [14, 19, 5]. The basic idea is to use latency measurements between a set of landmark nodes to create an embedding in a high dimensional space. Hosts can then use estimates of their latency to points in the coordinate space to predict the latency to those hosts in the Internet. The challenges in creating coordinate systems are in making them scalable, robust and accurate. While some reports have identified difficulties with some of the basic assumptions of some coordinate systems [21] more recent work has shown them to perform quite well in practice [13]. More recently, coordinate systems have been proposed as a mechanism for topological inference as well, which is how we use them in this work. In this case latency estimation is replaced by estimation of the hop count between nodes. One study taking this approach is [17], in which a hyperbolic embedding is used to embed nodes using hop counts as distance.

Num. Honeypots	Num. Sources
2	8680
3	4051
4	2816
5	2156
6	1570
7	1583
8	1574
9	55
10	4

Table 1: Counts of occurrences of common unique source IP addresses in multiple honeypots

Our work also contributes to the large and growing literature analyzing data gathered in network honeypots [2, 16, 15]. Honeypots do not solicit traffic; however low interaction sensors such as those used to collect data for our study will respond to incoming connection requests in order to distinguish spoofed addresses. In this way they are not completely passive. However, monitors of large address segments can receive millions of connections per day from systems all over the world and therefore offer a unique and valuable perspective [20]. Our work is based on data collected over a 24 hour period starting at 00:00 on December 22, 2006 from 15 topologically diverse honeypot sensors. These sensors are located in 11 distinct /8 prefixes that are managed by 10 different organizations. The segments of IP address space monitored by the honeypots varied in size from /25 to /21, along with one /16. Over 37,000,000 total packets were collected and evaluated in our study. The packets do not contain spoofed source IP addresses since they were the responses to SYN/ACKs from the honeynet [2]. Details of the data set can be found in Table 2. In order to preserve the integrity of the honeypots, we cannot disclose their locations in IPv4 address space.

The specific data provided by each honeypot was the source IP address and TTL extracted from the header of each packet. As already described, we can infer the number of hops between the source IP address and the honeypot using the packet's TTL value. Of particular interest and importance in our evaluation are the occurrences of the same source IP address in multiple honeypots. We found that 93.5% of the unique IP addresses in our data set appear in only one of the honeypots. This is most likely due to the diverse locations of the sensors coupled with the fact that different instances of malware limit their scans to smaller segments of address space. Nevertheless, this left us with over 22,000 unique IP addresses from which we conducted our analysis. Details of the instances of multiple occurrences of unique IP addresses are listed in Table 1 (note that there were virtually no addresses seen in more than 10 monitors).

3. NETWORK EMBEDDING FROM INCOMPLETE DATA

The main idea of our approach is to use the honeypots as landmarks and passively collected hop counts from IP

Node	Total Pkts.	Uniq. IPs	Mean Hops	Hop Variance
1	22,586,386	217,505	13.16	87.48
2	10,533,700	9,554	7.95	54.00
3	100,689	8,431	12.11	59.59
4	4,446	738	11.45	253.29
5	25,062	474	12.07	53.65
6	128,158	6,423	6.87	58.85
7	110,621	11,942	17.23	81.13
8	49,253	6,456	15.49	105.33
9	42,226	6,534	14.31	115.00
10	45,334	6,223	13.93	94.70
11	75,522	8,645	12.45	124.12
12	523,907	8,714	9.82	96.54
13	1,955,100	6,195	19.24	71.68
14	332,986	5,364	8.16	75.44
15	917,894	107,632	12.79	41.18

Table 2: Details of honeypot data sets used in our study. All data was collected over a one day period on December 22, 2006.

sources to the honeypots as distance measures. Ideally, each IP source is observed at each of the M honeypots, yielding an M-dimensional vector of hop counts from the IP source to the honeypots. The M-dimensional hop count vector places or *embeds* the IP source in an M-dimensional space. This is an instance of the so-called *network embedding* problem, discussed in the introduction. Using dimensionality reduction techniques (e.g., principle component analysis) it may be possible to reduce the M-dimensional embedding space to a lower dimensional space (see [19] for a discussion of such techniques in the context of network embedding based on latency measurements).

Mathematically, the dataset can be considered as the set of hop counts between a set of sources indexed by set $\mathcal{I}_S = [1, 2, ..., S]^T$ and a set of measurement (honeypot) nodes indexed by set $\mathcal{I}_M = [1, 2, ..., M]^T$. If the data is *complete*, this implies that we have knowledge of every hop count $h_{i,j}$, such that $h_{i,j}$ = the number of routers between source S_i and measurement node (honeypot) M_j, for all $i \in \mathcal{I}_S$ and $j \in \mathcal{I}_M$. For ease of notation, we will state the vector of hop counts as $h_i = [h_{i,1}, h_{i,2}, ..., h_{i,M}]^T$.

The major challenge associated with the use of honeypots is that, in general, *each IP source is not observed at each of the honeypots*. Thus, rather than obtaining a complete M-dimensional hop count vector for each IP source, typically only a subset of the honeypots observe each source (see Equation 1, 2). Moreover, different IP sources are observed by different subsets of the honeypots, making direct distance comparisons between IP sources impossible in most cases. The problem at hand can be viewed as a *missing data problem*.

$$h_j = [h_{j,1}, h_{j,2}, \ldots, h_{j,i-1}, h_{j,i}, \ldots, h_{j,M}]^T \quad (1)$$

$$h'_j = [h_{j,1}, -, \ldots, h_{j,i-1}, -, \ldots, h_{j,M}]^T \quad (2)$$

Missing data problems are routinely encountered in statistical inference problems, and a rich set of theories and tools have been developed to address them. At the heart of most of these is a data interpolation or *imputation*, wherein the missing data are estimated from the observations. In our problem, the missing data are missing hop counts between IP sources and honeypots. Once these missing data have been imputed, the completed dataset can be analyzed using standard network embedding techniques.

4. DATA IMPUTATION

In this section we propose and discuss several approaches to the missing data imputation problem. First we discuss fairly standard approaches such as imputation using the mean value and imputation based on ℓ_p-norm nearest neighbors. Then we propose a novel imputation scheme based on a network-centric uniformity measure motivated by topological considerations. Our experimental analysis in Section 5 indicates that the uniformity measure is more effective than conventional imputation strategies.

4.1 Mean Imputation

The most straightforward, albeit rather crude, estimator for the missing hop counts is the mean value of the observed data. That is, we compute the average hop count to a certain honeypot (over all IP sources observed at that honeypot), and use this average for all IP sources that were not observed by the honeypot. Suppose that hop count value $h_{k,j}$ is missing, we first need to find the indices for IP sources that have known hop values for measurement node j. Defining source index subset as $\mathcal{I}_j = \{i : h_{i,j} \text{ is known}\}$. Suppose that hop count value $h_{k,j}$ is missing, we then compute the estimated value:

$$\widehat{h_{k,j}} = \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} h_{i,j}$$

4.2 ℓ_p Imputation

Another method of estimating missing hop data would be to find the set of IP sources closest to S_k in the ℓ_p norm sense and then compute the average hop count using only the closest IP sources (rather than all the sources, as in the mean imputation above). Recall that the ℓ_p norm between two M vectors h and g is defined as $\|h - g\|_p = \left(\sum_{j=1}^{M} (h(j) - g(j))^p\right)^{1/p}$ (with $\|h - g\|_\infty = \max |h - g|$). Again, supposing that hop count value $h_{k,j}$ is missing, we first define source index subset $\mathcal{I}_{\ell_p}$ as the set of indices of

sources that are closest to S_k in the ℓ_p norm sense (as a subset of the set of indices ($\mathcal{I}_j$) for sources that have known values for measurement node j). This is mathematically defined as

$$\mathcal{I}_{\ell_p} = \{i : \underset{i \in \mathcal{I}_j}{\text{argmin}} (\|h_i - h_k\|_p)\}.$$

Given the set of IP sources in the complete data set closest to S_k in the ℓ_p norm sense ($\mathcal{I}_{\ell_p}$), the imputed values are computed as follows:

$$\widehat{h_{k,j}} = \frac{1}{|\mathcal{I}_{\ell_p}|} \sum_{i \in \mathcal{I}_{\ell_p}} h_{i,j}$$

While we only report the results of experiments with the ℓ_1 and ℓ_∞ norms, the same methodology can be applied for the ℓ_2 norm. Results from imputation experiments with the ℓ_2 norm were nearly identical to the ℓ_1 norm, which is why they are omitted in Section 5.

4.3 Network-centric Imputation

One problem with both the mean imputation and the ℓ_p imputation methods are that neither methodology takes advantage of the topology of the network. From [3], we can state that the network structure will resemble the network in Figure 1-(left). Each source IP address will enter the densely connected network core through a border router. Given this structure of the network, we can divide the total hop count values into the number of routers from the IP source to the core, and the number of routers from the core border node to the measurement node. We define the variables $\{x_i =$ the number of routers along the path from source S_i and the core$\}$, and $\{w_{i,j} =$ the hop between the core border node of source i and $M_j\}$. We state the hop count values: $h_{i,j} = x_i + w_{i,j}$ are the given number of routers between source S_i and measurement node M_j. Now consider the situation where two sources (S_i, S_j) are connected at the same border node (see Figure 1-(right)). Given that these two IP sources will share a path through the core to each measurement node, we can state that $w_{i,k} = w_{j,k}$ for all measurement nodes M_k. Therefore, $h_{i,k} - h_{j,k} = x_i - x_j = C$ for all measurement nodes M_k, where C is some constant integer.

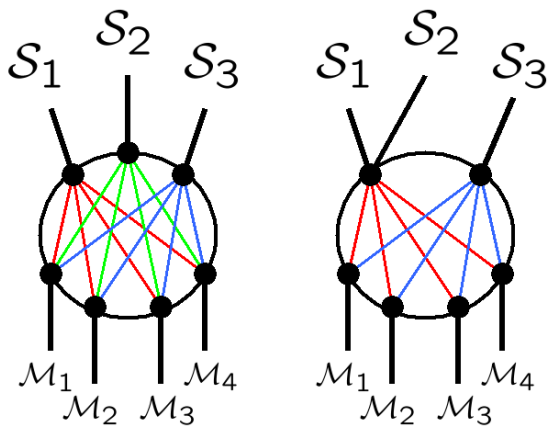

Figure 1: (Left) Example Network Topology, (Right) Example network w/ shared border node

We now develop a measure that exploits this property of two sources that have the same entry point into the core.

4.3.1 *Uniformity Measure Imputation*

Consider some subset of the indices of measurement nodes ($\mathcal{I}_{M(D)} \subseteq \mathcal{I}_M$). We can then define the uniformity measure of two sources (S_i, S_j) on the measurement node subset $\mathcal{I}_{M(D)}$ as:

$$\Delta_{i,j}\left(\mathcal{I}_{M(D)}\right) = \max_{k \in \mathcal{I}_{M(D)}} (h_{i,k} - h_{j,k}) - \min_{k \in \mathcal{I}_{M(D)}} (h_{i,k} - h_{j,k})$$

Proposition 1. If $h_{i,k} = h_{j,k} + C$ for all $k \in \mathcal{I}_M$ and some constant integer C then $\Delta_{i,j}(\mathcal{I}_M) = 0$

Proof. If $h_{i,k} = h_{j,k} + C$, then $\max_{k \in \mathcal{I}_M} (h_{i,k} - h_{j,k}) = C$ and $\min_{k \in \mathcal{I}_M} (h_{i,k} - h_{j,k}) = C$. Then by definition, $\Delta_{i,j}(\mathcal{I}_M) = C - C = 0$ □

Therefore, by Proposition 1, if two sources have hop counts that consistently differ by an integer constant across all measurement nodes, then the uniformity measure between the two sources will be zero. We can use this as an insight into the potential topology of the network, as a uniformity measure of zero is a necessary (but not sufficient) condition for two sources sharing a border node at the entrance to the network core.

Now consider a source S_k missing hop count data to a subset of the measurement nodes, indexed by subset $\mathcal{I}_{M(unknown)}$, and with known hop count data to measurement node subset $\mathcal{I}_{M(known)}$. We first need to find the subset of IP sources :

$$\mathcal{I}_{S(known)} = \{i : h_{i,j} \text{ is known, for all} j \in \mathcal{I}_{M(known)}\}$$

We then find the subset of source indices representing sources with the smallest uniformity measure with respect to source S_k and the known measurement node hop counts values:

$$\mathcal{I}_{S(uni)} = \{i : \underset{i \in \mathcal{I}_{S(known)}}{\text{argmin}} \left(\Delta_{i,j}\left(\mathcal{I}_{M(known)}\right)\right)\}$$

Then calculate the average difference between the known hop counts for IP source S_k and known hop counts for the uniformity-close subset of IP sources indexed by $\mathcal{I}_{S(uni)}$:

$$\widehat{C} = \frac{1}{|\mathcal{I}_{M(known)}|\,|\mathcal{I}_{S(uni)}|} \cdot \sum_{i \in \mathcal{I}_{S(uni)}} \sum_{j \in \mathcal{I}_{M(known)}} (h_{i,j} - h_{k,j})$$

We can then estimate all unknown hop counts $h_{k,j}$ for source S_k:

$$\widehat{h_{k,j}} = \left(\frac{1}{|\mathcal{I}_{S(uni)}|} \sum_{i \in \mathcal{I}_{S(uni)}} h_{i,j}\right) - \widehat{C}$$

5. EXPERIMENTAL RESULTS

The efficacy of the data imputation methods for missing data estimation was assessed as follows. Using the data acquisition methodology described in Section 2, we identified a set of over 1500 IP sources that have recorded hop counts to

the same 8 distinct honeypots. Thus, this dataset provides a "complete" 8-dimensional dataset. From this dataset we can synthetically generate missing data examples by knocking out (eliminating) certain hop count measurements. For example, knocking out $k < 8$ of the hop counts in a given hop count vector mimics the case in which k of the 8 honeypots did not observe the corresponding IP source. After generating synthetic missing data examples, we can apply the imputation methods and measure their effectiveness at "filling-in" the missing hop-counts in various ways (to be explained below). Using 10-fold cross-validation, in which 1/10 of the complete data is used to generate missing data examples and the remaining 9/10 is used for imputation of the missing hop counts, we can assess the performance of the methods.

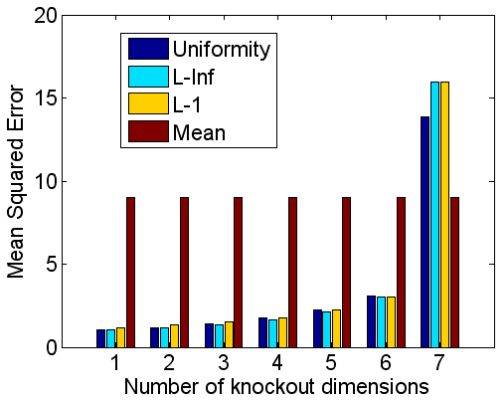

Figure 2: Imputation MSE Results

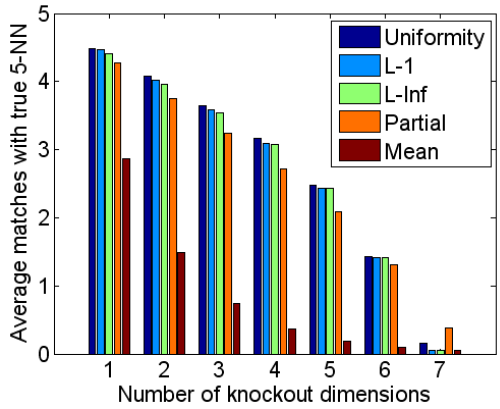

Figure 3: Imputation NN Results

5.1 Mean Squared Error Analysis

First we consider the mean squared error of the imputed values, estimated using the cross-validation procedure. As one can observe from the graph in Figure 2, the uniformity imputation technique performs on par with the best ℓ_p imputation technique. The poor performance of the mean technique shows that both the uniformity and ℓ_p imputation techniques are performing relatively well at estimating the missing data in up to six of the eight dimensions.

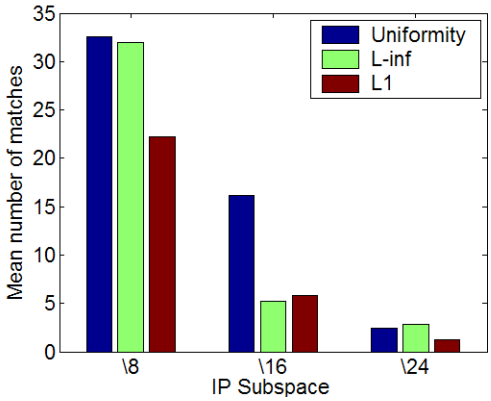

Figure 4: IP Subspace Cluster Results

5.2 Nearest Neighbor Analysis

A second measure of the effectiveness of the data imputation techniques, which captures the degree to which the imputed data preserves network structure/topology, is to compare the 5 Nearest Neighbors for a source S_k given complete data, with the 5 Nearest Neighbors for source S_k after deliberately eliminated hop counts were imputed. In addition to the imputation methodology described above, we also consider the 5 Nearest Neighbors based on only the remaining dimensions without imputation of the missing data (marked as *partial* in the graph/table). As one can observe from Figure 3, the imputation method based on the uniformity measure performs better than the other imputation methods except when all but one of the dimensions are missing. This results shows that the uniformity measure will recover the structure/topology of the network better than all other methods tested.

5.3 IP Cluster Analysis

The final analysis performed on the dataset looked at clusters of IP sources in the network using the hop count data (using all eight measurement node hop counts from the dataset). As stated previously, the uniformity measure will be zero for any two IP sources that share the same core border router, therefore the clustering procedure for the uniformity measure grouped together sources with uniformity measure values of zero for intra-cluster IP sources. For the purpose of comparison, cluster sets need to be created using the imputation methods. For the ℓ_p-norm, two sources were considered in the same cluster if the ℓ_p-norm for the hop count difference between two intra-cluster IP sources was less than a parameter τ (where $\tau > 0$). In the experiments, $\tau = 1$ so that the total number of clusters for each metric is roughly the same.

To verify the effectiveness of the clustering procedures, the five largest clusters were considered for each imputation method with respect to the entire 8-dimensional dataset. The IP addresses relating to the clusters were examined and the most frequently occurring /8, /16 and /24 addresses were found. As a measure of the topological relevance of the clusters, we then examined how often these dominant IP address subspaces occur in each cluster. The results in Figure 4) show that the uniformity measure clusters slightly more IP addresses in the /8 subspace with respect to the ℓ_1 and ℓ_∞ metrics. Meanwhile, for the /16 subspace, the

uniformity measure clusters a significantly larger number of addresses located in the dominant /16 subspace as compared to the ℓ_1 and ℓ_∞ metrics.

6. CONCLUSIONS

In this paper we take initial steps toward the possibility of discovering network topology in a passive manner. By taking passively collected traffic from honeypots and inferring the hop count from the honeypot to the source for each packet, we obtain raw data that hold topological information. Extracting useful topology from the raw data entails addressing two issues: how to replace missing measurements, and how to infer topology from a collection of complete measurements. We develop a network-informed distance metric for these problems, and show that it outperforms traditional distance metrics. As a result we are able to identify, from passive measurements alone, clusters of nodes that are likely to be sharing a common router on their paths to multiple honeypots. While our initial results are encouraging, there are more challenges ahead before network topology can be fully discovered in a passive manner, if indeed it ever can. In particular, it is an open question whether the set of hop counts alone contains enough information to recover detailed topological information. We leave this question for future work.

Acknowledgments

The authors would like to thank Farnam Jahanian and Michael Bailey from the IMS project at University of Michigan for providing the data used in this study. This work was supported in part by NSF grants CNS-0347252, CNS-0646256, CNS-0627102, CCR-0350213, CCF-0353079, and CCR-0325701, and support from Intel. Any opinions, findings, conclusions, or recommendations expressed in this material do not necessarily reflect the views of the NSF.

7. REFERENCES

[1] D. Alderson, L. Li, W. Willinger, and J. Doyle. Understanding Internet Topology: Principles, Models and Validation. *IEEE/ACM Transactions on Networking*, 13(6), December 2005.

[2] M. Bailey, E. Cooke, F. Jahanian, J. Nazario, and D. Watson. The Internet Motion Sensor: A Distributed Blackhole Monitoring System. In *Proceedings of The Network and Distributed Security Symposium (NDSS '05)*, San Diego, CA, January 2005.

[3] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the marginal utility of network topology measurements. In *Proceedings of ACM Internet Measurement Workshop (IMW '01)*, San Francisco, CA, October 2001.

[4] CAIDA. The Skitter Project. http://www.caida.org/tools/measurement/skitter/, 2007.

[5] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: A Decentralized Network Coordinate System. In *Proceedings of ACM SIGCOMM '04*, Portland, OR, August 2004.

[6] B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Efficient algorithms for large-scale topology discovery. In *Proceedings of ACM SIGMETRICS*, June 2005.

[7] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. IDMaps: A Global Internet Host Distance Estimation Service. *IEEE/ACM Transactions on Networking*, 9(5), 2001.

[8] P. Francis, S. Jamin, V. Paxson, D. Bryniewicz, and Y. Jin. An Architecture for a Global Internet Host Distance Estimation Service. In *Proceedings of IEEE INFOCOM '99*, New York, NY, April 1999.

[9] R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery. In *Proceedings of IEEE INFOCOM '00*, Tel Aviv, Israel, March 2000.

[10] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based Geolocation of Interenet Hosts. In *Proceedings of ACM Internet Measurement Conference (IMC '04)*, Taormina, Italy, October 2004.

[11] C. Jin, H. Wang, and K. Shin. Hop-Count Filtering: An Effective Defense Against Spoofed Traffic. In *Proceedings of IEEE INFOCOM '03*, San Francisco, CA, April 2003.

[12] E. Katz-Bassett, J. John, A. Krishnamurthy, D. Weatherall, T. Anderson, and Y. Chawathe. Towards IP Geolocation Using Delay and Topology Measurements. In *Proceedings of ACM Internet Measurement Conference (IMC '06)*, Rio de Janerio, Brazil, October 2006.

[13] J. Ledlie, P. Gardner, and M. Seltzer. Network Coordinates in the Wild. In *Proceedings of USSENIX Network Systems Design and Implementation (NSDI '07)*, San Jose, CA, April 2007.

[14] E. Ng and H. Zhang. Predicting Internet Network Distance with Coordinate-baseed Approaches. In *Proceedings of IEEE INFOCOM '02*, New York, NY, April 2002.

[15] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet Background Radiation. In *Proceedings of ACM Internet Measurement Conference (IMC '04)*, Taormina, Italy, October 2004.

[16] N. Provos. A Virtual Honeypot Framework. In *Proceedings of the USENIX Security Symposium '04*, San Dieog, CA, August 2004.

[17] Y. Shavitt and T. Tankel. Hyperbolic Embedding of Internet Graphs for Distance Estimation and Overlay Construction. *IEEE/ACM Transactions on Networking*, To Appear.

[18] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP Topologies with Rocketfuel. In *Proceedings of ACM SIGCOMM '02*, Pittsburgh, PA, August 2002.

[19] L. Tang and M. Crovella. Virtual Landmarks for the Internet. In *Proceedings of ACM Internet Measurement Conference (IMC '03)*, Miami, FL, October 2003.

[20] V. Yegneswaran, P. Barford, and D. Plonka. On the Design and Use of Internet Sinks for Network Abuse Monitoring. In *Proceedings of Recent Advances on Intrusion Detection (RAID '04)*, Sophia, France, September 2004.

[21] H. Zheng, E. Lua, M. Pias, and T. Griffin. Internet Routing Policies and Round Trip Times. In *Proceedings of The Passive and Active Measurement Workshop*, Boston, MA, April 2005.