DIMES: Let the Internet Measure Itself *

Yuval Shavitt and Eran Shir †
School of Electrical Engineering, Tel Aviv University, Israel
shavitt@eng.tau.ac.il, shire@eng.tau.ac.il

ABSTRACT

Today's Internet maps, which are all collected from a small number of vantage points, are falling short of being accurate. We suggest here a paradigm shift for this task. DIMES is a distributed measurement infrastructure for the Internet that is based on the deployment of thousands of light weight measurement agents around the globe. We describe the rationale behind DIMES deployment, discuss its design trade-offs and algorithmic challenges, and analyze the structure of the Internet as it seen with DIMES.

Categories and Subject Descriptors: C.4 [Performance of Systems]: Measurement techniques; C.2.3 [Network Operations]: Network monitoring.

General Terms: Measurements.

Keywords: Distributed Measurements, Internet topology.

1. INTRODUCTION

As the Internet evolved rapidly in the last decade, so has the interest in measuring and studying its structure. Numerous research projects [13, 9, 15, 4, 5, 14, 6] have ventured to capture the Internet's growing topology as well as other facets such as delay and bandwidth distributions, with varying levels of success. As the Internet continues to grow, especially far from its North American based core, measurement discrepancies are growing as well. A main handicap of current measurement projects is their rather limited number of measurement nodes (usually a few dozens up to a few hundreds) causing results to exhibit bias towards the core. In order to remedy this situation, a measurement infrastructure must grow several orders of magnitude in size and global dispersion.

We present DIMES, a highly distributed, global Internet measurement infrastructure, with the aim of measuring the structure and evolution of the Internet using a large set of interacting measurement agents. The key shift suggested in DIMES is the move from a small set of dedicated nodes, with measurements as their virtually sole objective, to a large community of host nodes, running light weight low signature measurement agents as a background process. Given the importance of location diversity in Internet measurements, this shift promises to enhance measurement results considerably.

Our goal is to map the Internet at several levels of granularity. At the coarse level, where each node is an AS, there are several mapping efforts, most notably are the active measurement Skitter project [6] and the passive collection of BGP data done by the RouteViews project [3], but also many of the studies mentioned above [13, 9, 15, 4, 5] examine the Internet (entirely or mostly) at this level. In the fine grain level, where each node represents a router, the mapping task is far more challenging, and the results achieved up to now [7, 14] are far from being satisfying. In addition, we believe neither of these granularities is enough. AS is too coarse a measure, where a node can represent a network that spans a continent, while the router level is too fine in many cases. Thus our goal is to generate, on top of the other two maps, a mid-level granularity map, namely, the PoP level map [17] where each node represents a group of routers working together, such as a small AS or a PoP of a large or medium size AS, which will be much more homogeneous.

*This research was supported in part by the EU 6th FP, IST Priority, Proactive Initiative "Complex Systems Research", as part of the EVERGROW integrated project; by a grant from the Israel Science Foundation (ISF) center of excellence program (grant number 8008/03); and by a grant from the Israel Internet Association.

†Corresponding author. Supported in part by Yeshaya Horowitz Association through the Center for Complexity Science.

2. MOTIVATION FOR DIMES

Measuring the structure of the Internet is a daunting task. The Internet is a highly complex, evolving system. Routing between ASes in the Internet is governed by the Border Gateway Protocol (BGP) and its characteristics dominate the ability to reveal details about the AS interconnection.

BGP is a path distance vector protocol, i.e., each AS announce to its neighbors not only the cost of its path to every destination but also the path itself. BGP is designed to enable Internet service providers (ISPs) to control the flow of data, thus an AS may choose not to announce some paths it knows due to *policy* which is determined by financial considerations. Thus, BGP allows two ISPs not to broadcast to their providers the link connecting them. As a result, a researcher collecting BGP announcements from a point outside of the two local ISPs cannot learn about the existence of the local connection. An attempt to learn about this local peer-to-peer connection using traceroute from an outside point will fail, as well, since the link is used only for local traffic. Only a presence in, at least, one of the two local ISPs will reveal the peer-to-peer link existence.

Previous studies [9, 5] show that indeed by adding more vantage points, new links are revealed, and that the marginal utility of adding new links decreases fairly fast. What escape these findings is the fact that while the marginal utility decreases, the mass of the tail is significant, thus if one is using a few vantage points, say up to a few tens, there is a small advantage to add a few more, but there is a significant advantage to add additional thousands of points as they will add a significant percentage of new links. Using only a few dozens vantage points gives a strong bias in the topology to customer-provider links and misses many of the peer-to-peer links.

These and other reasons make the case for a distributed, global, large scale measurement infrastructure. However, engineering a

dedicated infrastructure with thousands of measurement computers spread around the globe is a feat only the largest of corporations can accomplish. Thus, in order to accomplish such a task, one must move to a distributed hosting paradigm, where lightweight measurement software is hosted by volunteers on computers all over the globe. Recently, the effectiveness of this approach has been demonstrated by several projects [2, 1, 10, 8] in various contexts, most related to computation intensive tasks. For Internet measurements, the contribution of a distributed approach is in the location heterogeneity. Using this approach, one can envision gaining presence in thousands of ASes.

3. DIMES RESEARCH GOALS AND ALGORITHMS

DIMES has several research goals from which experimental goals are derived. The main goal is to be able to take full snapshots of the Internet graph, in the AS, PoP, and router level, annotated with delay and loss statistics, in fine time resolutions. For the AS level graph, our intention is to reach a time resolution of less than two hours, while for the router level, our initial aim is to acquire a daily snapshot of the entire IP level internet graph.

In order to achieve this goal, it is not suffice to merely achieve many measurement hosts. Rather, experiment design algorithms must be developed for optimally assigning measurement tasks to each of the agents according to its parameters and location. These algorithms can be divided into two separate groups: Discovery algorithms and Re-discovery algorithms. Discovery algorithms are realizations of heuristics which have a higher probability than random sampling to discover new edges and nodes that were not discovered to date. An example of such heuristic is the 2-neighborhood heuristic, where we find the 2-shell, i.e., the group of nodes which are exactly two hops from a certain node, of each AS from which we measure, and designate a set of measurements to destinations in that group. The underlying hypothesis being that the probability of two ASes to be connected is higher if they have a mutual neighbor. This realization stems from a popular structure appearing on the Internet where two ASes which are customers of a certain provider AS have a peering connection established between them.

Re-discovery algorithms are algorithms which aim to re-validate the existence and annotations of each of the edges already discovered using a minimal amount of measurements. One can think of this problem as an instance of a set cover problem, where the elements of the set are the edges and each traceroute measurement, defined by a source-destination couple is a different subset. Applying the set cover greedy algorithm, allows us to assign measurements to agents which will cover the discovered Internet graph with a measurement budget that is considerably smaller than the size of the network, and with a few thousands of well spread agent population in roughly one hour. Complementary smart algorithms [11, 12] were recently suggested to reduce the number of measurement traversing an edge.

4. ANALYZING DIMES PERFORMANCE

4.1 Building the case for DIMES

The underlying claim of the DIMES approach is that for accurately measuring the Internet's topology one must abundantly use distributed measurement nodes. To establish this claim, we need to compare DIMES results to results coming from traditional approaches, showing a significant qualitative difference, and to show that the agents' contribution distribution has a heavy tail, meaning that new agents added to the DIMES platform contribute a considerable amount of new information.

The Route Views project [3] gather BGP updates from about 70 BGP speakers around the world, which makes it the largest open passive measurement database. As such, AS topologies inferred from Route Views data are the best yardstick against which measurement projects should compare themselves, at least at the AS realm. Given the dynamical nature of the Internet, one should be careful in comparing topologies, making sure that the topologies relate to the same time period and scope. Thus, in order to appropriately compare the DIMES topology to BGP inferred topology, it was necessary to take an integration of BGP updates during the measurement period. Thus we sampled BGP updates from Route Views, choosing one BGP update per day.

Topology	N	E	$< k >$	γ	CC
DIMES	14697	61757	8.40	-2.10	0.67
BGP	20585	45720	4.44	-2.09	0.28
Complete	20691	82131	7.94	-2.10	0.59
BGPinDIMES	14583	33238	4.56	-2.27	0.30

Table 1:

4.2 Data Collection Methodology

Up to September 1^{st} 2005 we collected over 460 million measurements, roughly half of them are traceroutes and the rest pings, from over 5000 agents, spread in more than 570 ASes. These traceroutes can be integrated in time to produce periodic AS topologies. A first step in building the AS topology is to associate IP addresses to ASes. Our current approach for the association process is to mimic a router's decision making process using a longest prefix matching algorithm, which looks for the longest prefix in our database that matches the IP in question. The prefix database, in turn, is built from prefix announcements in BGP data available on the Internet. The resolution process is augmented with whois data resolution, which is performed for IP addresses for which the main process has failed. Typically about 2-3% of the IPs fail the longest prefix matching and are resolved using whois, and currently, between 1-1.5% of the IPs fail resolution entirely.

At the current measurement rate, we discover about 61000 AS edges in a month connecting over 15000 ASes. However, to be on the conservative side, we analyze a topology which contain an AS edge only if it was found by at least two separate measurements.

4.3 Comparing DIMES vs. BGP Topologies

In the following, we will compare four different topologies that were created from the set of measurements defined above: DIMES topology, which is the AS level topology inferred from the DIMES measurements during July 2005. BGP Topology, which is the topology inferred from BGP updates gathered from Route Views during July 2005. Complete Topology, which is the unification of the DIMES and BGP Topologies. BGPinDIMES Topology, which is the BGP Topology subgraph which spans only AS nodes that belong to the DIMES topology. Table 1 shows the main properties of the four topologies.

There are several conclusions that can be drawn from comparing these topologies. First, the degree distribution power exponent remains robust and hardly changes between the topologies, making it a poor characterizer of network differences. Thus, we should look for deeper topological characteristics to compare by [16]. Indeed, the clustering coefficients (CC) of the DIMES topology is almost double the CC of the BGP inferred topology, an immense difference. This cannot be attributed to the partial node population in our

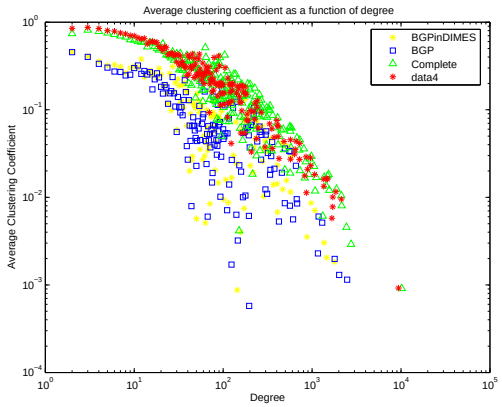

Figure 1:

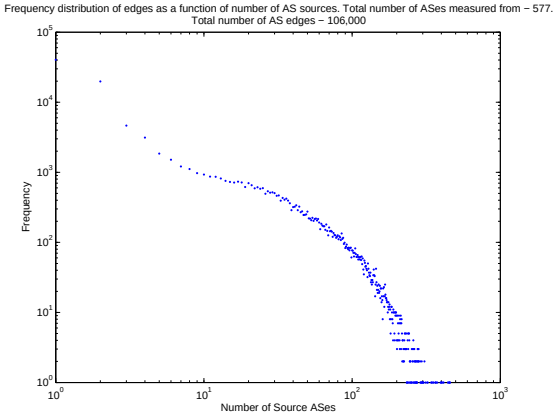

Figure 3:

topology since the CC of the BGPinDIMES topology is not much higher than the CC for the BGP topology. In Figure 1 we compare the clustering coefficient distribution of the topologies, showing a larger difference in clustering coefficients of low degree nodes as well as the apparent under-sampling of medium degrees clustering in BGP. This property shows that many of the new links found by DIMES are periphery peer links. Finding these rich structures in the periphery was one of the main motivations for constructing DIMES.

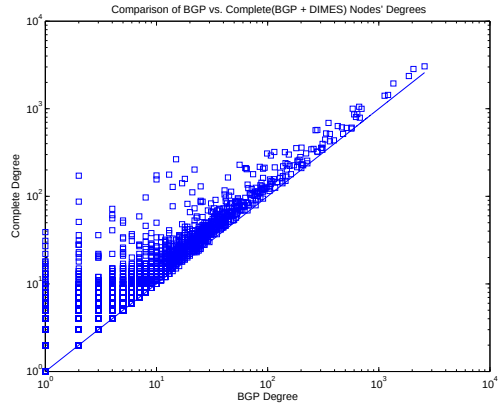

Figure 2:

In Figure 2 we compare the degree of nodes in the BGP topology vs. their degree in the Complete topology, namely with the DIMES contribution. We observe that about one third of the nodes has a higher degree than perceived by BGP data. The highest degree in the remaining two thirds of ASes is 238, which means that all high degree nodes are augmented with edges, many of them to a considerable amount. AS 7018 (AT&T), for example, has more than 900 DIMES edges which do not appear in the integrated BGP topology, increasing its degree by more than 40%. DIMES data has contributed more than 500 new AS links to UUNet, Sprint and Level3. However, focusing on the top 50 hubs of the network only tells portion of the story. The other 6400 ASes with BGP degree lower than 238 and DIMES edges not appearing in BGP constitute the bulk of the DIMES contribution. The lower degree nodes show sometime huge differences in degrees, which rise up to 30-fold and more.

An interesting question is how robust are the excess edges which do not appear in BGP, where by robust we refer to the number of ASes we see the edge from and the number of measurements that the edge was a part of. In Figures 3 and 4 we present the edge count distribution as a function of number of ASes from which it was measured and as a function of number of measurements it belonged to respectively.

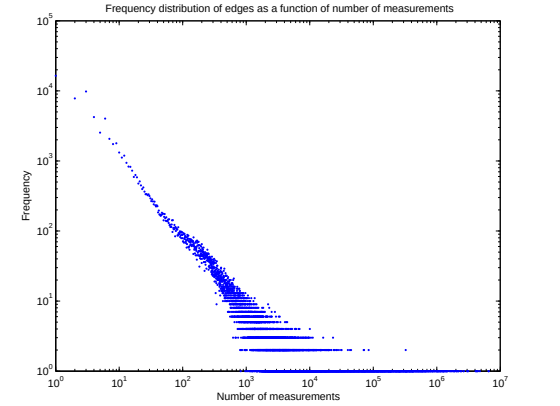

Figure 4:

As one can see, the BGP topology has about 25% more nodes than the DIMES topology. This difference is due to two main reasons. The first reason is that many ASes (for example military ASes and some corporations ASes) block active probes such as traceroute with various methods. To circumvent this issue to a certain degree the new version of the DIMES agent, which was just recently deployed, is augmenting the ICMP traceroute, which we used in the previous versions, with UDP based traceroute. We also plan the introduction of TCP SYN probes in the near future. The second reason for this gap is due to a lack of destination coverage, where we have not identified yet IP addresses which we can measure to in these ASes.

4.4 Agents contribution

The DIMES platform relies on volunteers enlisting into the system and installing the DIMES agent. As such, it is important to quantify their contribution, and specifically to quantify the con-

tribution of new agents joining in the presence of many existing agents. Several authors [5, 9] claimed that above a very low threshold (measured in few tens) additional agents' return will diminish and become unimportant. Looking into the contribution dynamics of the last year, one sees that the situation is far from it, as illustrated in Figure 5. In this figure, the X axis represents days since project initiation, and the Y axis represents the ordered rank of the agents (i.e., agent who was 38 to join will have index 38). A point is plotted for each AS edge discovered according to the agent who discovered it and the date in which it was found. As can be seen, even agents that registered after hundreds of millions of measurements were performed still contribute substantially to the AS graph.

An interesting observation from Figure 3 is that about 40% of the total DIMES edges have been seen only from a single AS, and additional 20% of the edges from two vantage points. Since we have measurements from only a few hundreds of ASes, we can assume that there are still many unknown edges we have not discovered, yet.

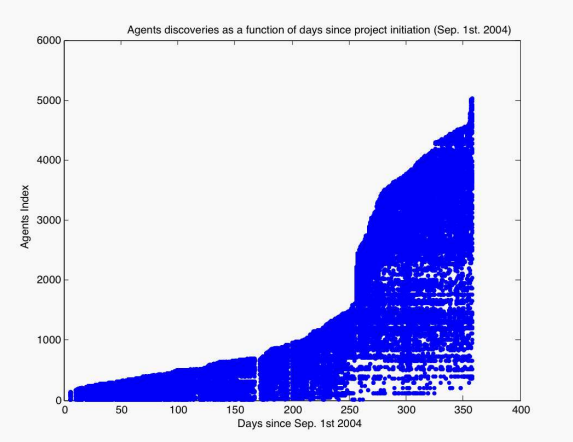

Figure 5:

5. DIMES STATUS AND INITIAL RESULTS

DIMES was launched on September 1st. 2004. Since then more than 5000 agents from 80 countries were registered at the DIMES website (www.netdimes.org). During its first year, measurement were performed from over 570 different ASes. Currently, we map about 61000 AS edges connecting over 15000 AS nodes. Out of these 61000 edges, almost half are edges that are not present in main BGP tables repositories such as the RouteView project [3], making the Internet 50% denser than previously thought.

In the IP level we have discovered about $1,200,000$ identifiable interfaces, connected with about $8,000,000$ edges. We are in the midst of the process of identifying interfaces that belong to the same router. Currently we managed to match over 20,000 interfaces which reduced the number of edges in the router level map to about $6,000,000$. In addition, we are working on identifying hosts that are not identifiable by traceroute like measurements (since they do not respond in any way) though we know they exist. In order to perform this task, we plan to use high dimension clustering algorithms, which will be able to identify groups of unknown hosts as a single host using path delay measurements from many vantage points. The underlying hypothesis of this approach is that if a certain router is situated in identical distances from many measurements points as another router, then they must be either the same router or situated very closely to each other.

6. DIMES FUTURE DIRECTIONS

Our main aim is to use the data collected by DIMES in order to develop a realistic, predictive model of Internet evolution and dynamics, and identify the fundamental properties of the Internet graphs. Though mapping the Internet bandwidth distribution is a very hard task, we aim to infer it by enforcing bandwidth-degree type trade-off constrains on the router level and PoP level graph. Eventually, we plan to embed the measurement results in a geographic metric, and develop measures which use the Internet evolution characteristics in various regions as an indicator of economic and social evolution.

7. REFERENCES

[1] Distributed.net. http://www.distributed.net/.

[2] SETI@Home. http://setiathome.berkeley.edu/.

[3] University of Oregon Route Views Project. http://www.antc.uoregon.edu/route-views/.

[4] S. Bar, M. Gonen, and A. Wool. An incremental super-linear preferential internet topology model. In *PAM '04*, Antibes Juan-les-Pins, France, Apr. 2004.

[5] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the marginal utility of network topology measurements. In *ACM SIGCOMM IMW '01*, San Francisco, CA, USA, Nov. 2001.

[6] A. Broido and K. Claffy. Internet topology: connectivity of IP graphs. In *SPIE International symposium on Convergence of IT and Communication '01*, Denver, CO, USA, Aug. 2001.

[7] H. Burch and B. Cheswick. Mapping the internet. *IEEE Computer*, 32(4):97–98, 1999.

[8] J. Charles Robert Simpson and G. F. Riley. Neti@home: A distributed approach to collecting end-to-end network performance measurements. In *PAM '04*, Antibes Juan-les-Pins, France, Apr. 2004.

[9] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power-laws in internet topologies revisited. In *IEEE Infocom 2002*, New-York, NY, USA, Apr. 2002.

[10] M. Dharsee and C. Hogue. Mobidick: A tool for distributed computing on the internet. In *Heterogeneous Computing Workshop '00*, Cancun, Mexico, May 2000.

[11] B. Donnet, T. Friedman, and M. Crovella. Improved algorithms for network topology discovery. In *PAM '05*, Boston, MA, USA, Mar./Apr. 2005.

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

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM 1999*, Boston, MA, USA, Aug./Sept. 1999.

[14] R. Govindan and H. Tangmunarunki. Heuristics for internet map discovery. In *IEEE Infocom 2000*, pages 1371–1380, Tel-Aviv, Israel, Mar. 2000.

[15] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in ip topology measurements. In *IEEE INFOCOM '03*, San Francisco, CA, USA, Apr. 2003.

[16] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet's router-level topology. In *SIGCOMM 2004*, 2004.

[17] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with rocketfuel. In *ACM SIGCOMM '02*, Pittsburgh, PA, USA, Aug. 2002.