

Analyzing Router Responsiveness to Active Measurement Probes

Mehmet H. Gunes¹ and Kamil Sarac²

¹ University of Nevada - Reno, Reno, NV 89557, mgunes@cse.unr.edu

² University of Texas at Dallas, Richardson, TX 75080, ksarac@utdallas.edu

Abstract. Active probing has increasingly been used to collect information about the topological and functional characteristics of the Internet. Given the need for active probing and the lack of a widely accepted mechanism to minimize the overhead of such probes, the traffic and processing overhead introduced on the routers are believed to become an important issue for network operators. In this paper, we conduct an experimental study to understand the responsiveness of routers to active probing both from a historical perspective and current practices. One main finding is that network operators are increasingly configuring their devices not to respond to active direct probes. In addition, ICMP based probes seem to elicit most responses and UDP based probes elicit the least.

Key words: Internet measurement, active measurements

1 Introduction

Internet has become one of the largest man made systems with a significant impact in many aspects of our daily life. Due to the tremendous growth in its size and importance, many groups, organizations, and governments have become interested in understanding various characteristics of the Internet for commercial, social, and technical reasons. In general, Internet measurement studies can be divided into two as (1) active measurement and (2) passive measurement studies. Active measurement studies can also be divided into two as (1) the ones that require participation from the network devices (i.e., routers) and (2) the ones that involve end systems only.

Active measurement studies that require router participation typically send measurement probes to routers and expect to receive responses from them. Naturally, such probes incur processing and traffic overhead on the routers. Upon receiving a probe message, a router is expected to create response packet and send it back to the probe originator. Most routers perform this processing on the slow forwarding path. This introduces a significant processing overhead as compared to simple packet forwarding on the fast forwarding path at the routers. In addition to network measurement studies, today many popular overlay and peer-to-peer network applications utilize active measurements to optimize their topology and/or routing performance.

In this paper, we conduct an investigation on the responsiveness of routers to active network measurements. Our goal is to quantify the responsiveness to measurement activities in two directions (1) historical perspective in terms of router participation in supporting active measurements and (2) today’s best practices employed by network providers in allowing different types of network measurements on their routers. For the historical study, we use router anonymity as observed in traceroute outputs and for today’s best practices we use different types of active probes and observe the responsiveness of routers to them.

For the historical study, we use path traces collected by skitter [11] and utilize one trace set for each year starting 1999. We look at the ratio of occurrences of ‘*’s in each path trace in data sets and use these occurrences to indicate router anonymity, i.e., lack of router participation to active probing. We study the data set both before and after processing the raw data to build the corresponding sample topology map (i.e., before and after resolving IP aliases and anonymous routers). Our results show that there has been an increase in the ratio of router anonymity in time and especially after the year of 2004. Our results also show some interesting trends on the locality of anonymous routers in path traces.

For the analysis of current best practices, we collected a set of IP addresses (536K+ of them) from a recent skitter and iPlane [10] measurement studies. These IP addresses are known to respond to indirect probes by skitter and iPlane systems. An indirect probe has a different destination than the routers that it elicit responses whereas a direct probe has a destination IP address of the intended router. In our work, we send different types of direct probe messages to each of these IP addresses and observed their responsiveness. One main observation we have is that routers are most responsive to ICMP based active probes. This is followed by TCP and then by UDP based probes. We also classified the IP addresses based on their top-level domain extensions and observed that routers in different type of organizations (e.g., .net, .com, .edu, .org, and .gov) present a similar behavior in response to direct active probes.

The rest of this paper is organized as follows. The next section presents the related work. Section 3 classifies anonymity types. Section 4 present our observations on the historical data set. Section 5 presents our observations on the current data set. Finally, Section 6 concludes the paper.

2 Related Work

Active probing has increasingly been used in various contexts to observe different characteristics of the underlying network. As an example, several research studies utilize active probing to monitor the routing and reachability behavior of the Internet in the inter-domain scale [9, 11, 13]. On the other hand, many overlay or peer to peer network applications depend on active probing to optimize the performance of their applications [12]. Based on the increasing need for active measurements, the research community has developed several large scale distributed measurement platforms (e.g., ark [1], Dimes [13], DipZoom [15],

iPlane [10], PlanetLab [2], skitter [11], rocketfuel [14] etc.) that are commonly used to conduct various measurement activities that include active probing.

As the volume of active measurement practices increased, several researchers pointed out the impact/overhead of active probing in the network and presented approaches to reduce the volume of redundant active probes in measurement studies. Nakao et al. pointed out the increase in active measurement traffic and proposed a *routing underlay* to unify all measurement activities [12]. In this approach, overlay networks query the routing underlay for measurement information and the routing underlay extracts and aggregates topology information from the underlying network and uses this information to answer the queries.

Within the specific context of traceroute based topology measurement studies, Donnet et al. presented one of the early work on minimizing the active probing overhead in traceroute based topology discovery [4, 5]. This work presented doubletree, a mechanism to avoid unnecessary probing of the same set of routers multiple times in a traceroute based topology discovery. A follow up work by Donnet et al. presented a modified approach called windowed doubletree that improves on doubletree by extracting missing links/nodes in discovering a targeted network topology [3]. More recently, Jin et al. considered scalable discovery of the underlying network topology among a group of nodes that participate in an overlay network [8]. Their main idea is to utilize a network coordinate system in identifying path traces to collect at each overlay node so as to discover the underlying topology while issuing the minimum number of path traces.

In summary, most of the related work in the area has been on reducing the unnecessary overhead in active probing based topology discovery studies. The main goal of the work presented in this paper is to quantify the responsiveness of routers to active probing based network measurement practices. The results presented in this paper provide a historical perspective and demonstrate the current practices of network operators to accommodate active network measurements.

3 Types of Router Anonymity

In our work, we measure router unresponsiveness (i.e., anonymity) to active probes. Active probes are divided into two types (1) direct probes and (2) indirect probes. A *direct probe* is the one where the destination IP address in the probe packet is the intended destination as in ICMP ping. An *indirect probe* is the one where the destination IP address in the probe packet is some other destination as in traceroute when it probes an intermediate router during the trace.

In both direct and indirect probing, eliciting a packet from probed node indicates the responsiveness of the node. On the other hand, not receiving a response to an active probe packet may have different interpretations. In the case of direct probing, the lack of a response message may not necessarily indicate node anonymity as it may be that the probed node may be unreachable; may be disconnected or turned off; or either probe or the response packet may be filtered out at some node in the network. In the case of indirect probing as in traceroute,

the lack of a response message, in general, indicates node anonymity especially if another responsive node appears later on within the same trace output.

We define several anonymity types for nodes/routers that we observed in our experiments. Note that, both Type 1 and Type 2 can be further classified into two types. However, as an observer there is no difference between them.

Type 1: A router may be configured to ignore certain probe packets causing it to be anonymous with such probing. In addition, a border router may be configured to filter out (i) certain types of packets (e.g., unsolicited UDP packets directed to a local host) or (ii) outgoing ICMP responses originating from nodes within its local domain. Filtering at the border causes the internal nodes to be seen as anonymous as well.

Type 2: A router may apply ICMP rate limiting and become anonymous if the rate of the incoming probes exceed the preset limit. Similarly, a router may ignore probe packets when it is congested but may respond to them when it is not congested. In either case, the router has changing responsiveness.

Type 3: A router may have a private (publicly unroutable) IP address. Such IP addresses cannot guarantee node uniqueness and hence their originators need to be considered as anonymous.

Finally, as the above discussion suggests, the classification presented in this paper are approximate in their nature as it is difficult to know the actual cause of a lack of response for an active probe packet.

4 Historical Perspective

In this section, we use traceroute collected historical data sets to study router reaction to indirect probe messages. We downloaded 10 sets of ICMP traceroute path traces from CAIDA's skitter web site which is the only source we are aware of publicly providing periodic historical topology data. These data sets were collected within the first collection cycle in January of the each year by the skitter system. The web site reports that they had several updates to destination IP address lists. They also had a major change in their topology collection system in mid 2004 where they utilized dynamic destination lists with increased probing frequency at skitter monitors.

In the first step, we look for a trend in the ratio of anonymous routers in the collected data set. We do this before and after processing the raw path traces. Our processing in this context includes IP alias resolution and anonymous router resolution tasks. For IP alias resolution, we use an analytical inference approach called AAR [6]. Note that due to the historic nature of the data (i.e., due to the fact that the underlying topology might have changed substantially in time), a probe based IP alias resolution approach is not considered. The implication of this limitation is that the resulting topology map may have inaccuracies due to the limitations in alias resolution. Especially, data sets in 2001-2003 had much fewer aliases which we think is due to poor alias resolution. For anonymous router resolution, we use a graph-based induction (GBI) approach [7]. Table 1 presents

Year	Initial					Anonym Type (%)			Final	
	#Srcs	#Traces	Reached	#Nodes	Anonym	1	2	3	#Nodes	Anonym
1999	5	3.5M	86.5 %	0.2 M	59.0 %	41.1	0.0	58.9	17 K	1.1 %
2000	14	14.8 M	83.5 %	0.7 M	80.6 %	19.8	0.0	80.2	18 K	1.8 %
2001	17	13.4 M	73.6 %	2.1 M	72.7 %	27.4	0.0	72.6	575 K	0.7 %
2002	20	19.1 M	50.4 %	1.5 M	51.2 %	48.9	0.0	51.1	369 K	0.8 %
2003	23	24.3 M	54.3 %	1.9 M	42.0 %	58.1	0.0	41.9	703 K	0.6 %
2004	23	22.9 M	53.0 %	2.4 M	64.1 %	36.0	0.0	64.0	45 K	1.0 %
2005	22	21.0 M	46.4 %	6.8 M	85.9 %	71.8	2.3	25.9	86 K	8.5 %
2006	19	18.4 M	37.2 %	6.4 M	87.4 %	61.3	2.2	36.5	73 K	9.7 %
2007	18	17.5 M	30.6 %	4.9 M	85.3 %	91.9	4.2	3.9	79 K	12.5 %
2008	11	10.7 M	23.2 %	2.8 M	77.2 %	90.9	7.1	2.0	61 K	9.4 %

Table 1. Analysis of historical responsiveness

the main characteristics of the results where “#Srcs” indicate the number of vantage points; “Reached” gives the percentage of traces that reached the final destination; “# Nodes” gives the number of nodes within the data set before processing (i.e., before IP alias and anonymous router resolutions); “Anonym” gives the percentage of anonymous nodes in the original data set. The next three columns give the classification of anonymous nodes as percentage values. The last two columns presents final topology size and percentage of anonymous routers after processing.

According to the table, the ratio of path traces reaching their final destination decreases in time. In addition, the ratio of anonymous nodes after processing has a big increase after 2004 (see the last column). The table also shows some increase in this ratio before processing but not as much. Another observation from the column “Type 2” is that the ratio of routers employing rate limiting on active probes has increased in time. These anonymous routers had IP addresses aliased to anonymous nodes in different path traces. On the other hand, “Type 3” anonymity seem to reduce significantly during the last two years increasing the rate of “Type 1” anonymity. This might be due to recent practices of dropping incoming packets with private source addresses. We also observed that in some cases a trace source was behind a NAT and there was a high number of “Type 3” anonymity close to the probing source. Finally, for 2001-2003 data sets after processing, the “# Nodes” is substantially larger than the other years. Our IP alias resolution process returned relatively smaller number of alias pairs for these data sets. The net affect of this outcome is that most IP addresses observed in the original data set appeared as unique nodes within the final data set. This then resulted in a final topology with a large number of nodes.

In the second step, we are interested in the length distribution of path segments formed by one or more consecutive “*”s in path traces. Note that in this analysis, we do not include Type 3 anonymity in order to observe behavior of the routers to active probe messages only. We call a path segment in the form of a $(IP_1, *_1, *_2, \dots, *_l, IP_2)$ a **-substring* of length l . We are then interested in the frequency distribution of *-substrings with respect to their length l . Although a

Year	Unique		*-substring length				
	*-substrings	Same AS	1	2	3	4	5
1999	1	100 %	1	-	-	-	-
2000	24	12.5 %	24	-	-	-	-
2001	57	22.8 %	57	-	-	-	-
2002	41	9.8 %	41	-	-	-	-
2003	79	20.3 %	79	-	-	-	-
2004	86	44.2 %	86	-	-	-	-
2005	225,456	12.6 %	151,133	63,662	6,360	4,301	-
2006	207,067	11.6 %	137,829	59,171	5,828	4,239	-
2007	305,331	14.4 %	212,263	73,263	14,019	5,779	7
2008	231,633	14.0 %	148,182	63,944	13,733	5,772	2

Table 2. *-Substring distribution

*-substring of length one may have different interpretations about the cause of router anonymity, the occurrences of *-substrings with large length values may be an indication of ISP policy in preventing active probing in its network. Table 2 on the right hand side presents the frequency distribution of *-substrings w.r.t. their length. In this table, we present the number of unique *-substrings in the original data set. For uniqueness, we represent a *-substring of length l as a triplet (IP_1, l, IP_2) and avoid counting the duplicate triples of this form in the table. The results in the table show a bimodal behavior where data sets prior to 2005 have only length $l = 1$ *-substrings. Starting from 2005, we observe *-substrings of larger lengths with the majority of *-substrings being of length $l = 1$ or $l = 2$. We partly attribute the bimodal behavior of the routers to changes in the skitter data collection process such as the increased probing rate and dynamically updating destination lists. Longer *-substrings might also be due to growth in networks where more hops of an autonomous system are traversed or due to increased use of MPLS. One general observation in this part is that within the last decade ISPs have become less cooperative to active probing and configure their routers to stay anonymous to such probes.

We also classify *-substrings into two based on their topological position. That is, for a given *-substring, say (IP_1, l, IP_2) , we look at the relation between IP_1 and IP_2 . If the IP address share the same 16-bit IP prefix, then we assume that these IP addresses belong to the same domain and therefore the anonymous nodes in between also belong to the same domain. We look at the ratio of this type of *-substrings in the data sets and present them in the “Same AS” column in Table 2. An alternative approach for this classification would be to map each such IP address (IP_1 and IP_2) to their corresponding Autonomous System (AS) number and do the check. But, since we did not have access to IP-to-AS mapping information for most of these historic data sets, we utilized the above mentioned heuristic. The results presented in the “Same AS” column suggest that most *-substrings after 2004 (e.g., more that 85%) were in the form that the two IP addresses IP_1 and IP_2 seemed to belong to different domains. The ratio prior 2005 did not show a consistent trend which we contribute to the relatively

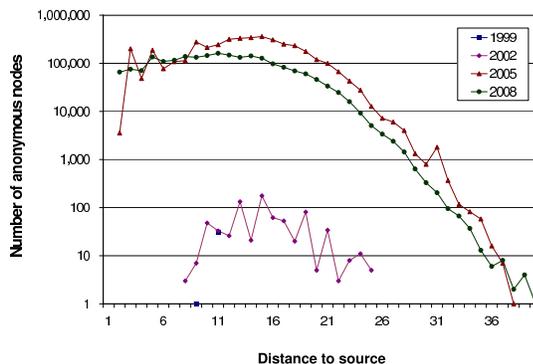


Fig. 1. Distance distribution

small number of occurrences of *-substrings during that time (see the “Unique *-substrings” column in the table). Given that most *-substrings are of length $l = 1$ or $l = 2$ from the right hand side of the table, we suspect that the majority of *-substrings that occurs in data sets after 2004 originated from routers at domain boundaries or exchange points between neighboring ASes. We also observed that even though the number of path traces used in our study decreased after 2004, the number of unique *-substrings increased during the same time period.

In the next step, we are interested in the position of anonymous nodes within path traces. In order to observe anonymous node positions in our path traces, we counted the number of anonymous nodes at each hop distance from trace sources in all path traces grouped by years. Figure 1 presents the distance distribution of anonymous nodes for four different years as samples. Similar to the previous case, this analysis excludes nodes due to Type 3 anonymity. Note that, the results presented in the figure do not consider the path length distribution of individual path traces in these data sets. According to the figure, early data sets (i.e., before 2005) contained small number of anonymous nodes that were mostly distributed 10 to 20 hops away from the source. On the other hand, recent data sets included much more anonymous nodes majority of which appeared 3 to 25 hops away from the source. The figure also shows a high number of anonymous nodes at a distance of 2 from the source for the 2008 data set. A close examination of the corresponding data set shows that this case is due to the existence of an anonymous router at a 2 hops distance to one of the vantage points.

5 Current Practices

In this section, we present our findings in an experimental study where we observe router responsiveness to direct and indirect probe messages. We use 536,743 IP addresses obtained from skitter and iPlane project web sites. These IP addresses were recently collected (between Apr 7-11, 2008) by running traceroute queries in the Internet, i.e., they belong to routers/systems that recently re-

Year	All	Router	End-Host	.net	.com	.edu	.org	.gov
# IPs	537 K	320 K	217 K	25.5 K	10.1 K	5 K	1.7 K	0.5 K
ICMP	81.9 %	84.6 %	77.9 %	92.3 %	86.4 %	88.9 %	95.5 %	92.9 %
TCP	67.3 %	70.4 %	62.8 %	76.7 %	72.6 %	83.2 %	77.3 %	83.0 %
UDP	59.9 %	64.7 %	50.3 %	63.5 %	61.7 %	57.3 %	64.4 %	62.8 %

Table 3. Responsiveness to direct probes

sponded to indirect probe messages (i.e., traceroute probes issued from skitter and/or iPlane systems).

In our study, we first issued UDP, TCP, and ICMP based direct probes to each of these IP addresses and recorded the response (or the lack of it). For each case, we issued three probes from a host in UT-Dallas network and expected to receive at least one response to consider the probed node as responsive. In general, UDP probes are expected to return ICMP Port Unreachable messages; TCP probes (TCP SYN packets) are expected to return TCP SYNACK or TCP RST messages; and ICMP probes (Echo Requests) are expected to return ICMP Echo Reply messages. Finally, we issued another set of ICMP probes with IP Route Record option set (by using the ping command available in Linux system). We observed a very small response rate (124 responses out of 536K+ probes) and therefore excluded them from discussion.

Table 3 presents the response rate that we observed during our direct probing. The first row gives the number of IP addresses (i.e., IP addresses responding to indirect probes by skitter and iPlane systems via traceroute). The following three columns give the percentage of the IP addresses (out of the absolute numbers given on the same column above) responding to direct probes by ICMP, TCP, or UDP based probes respectively. The table also groups IP addresses as belonging to routers or end systems; and classifies them based on their top level domain extensions (only 5 of them are presented).

According to the results presented in the table, ICMP based direct probes have the highest rate of responses followed by TCP and then UDP probes. The results suggest that TCP based probes are more welcomed than the UDP ones. However, during our TCP based active probing, we received several security alert e-mails from a national ISP indicating that our probe messages were detected by their monitoring system as possible network scan activity and we were asked to stop our probing. On the other hand, our UDP or ICMP based active probing did not raise any security alert (that we know) by this or any other ISPs. These results also indicate that many routers that respond to indirect query probes do not respond to direct query probes. The ratio of such routers ranges from around 18% for ICMP probes to 40% for UDP probes. This implies that in practice network operators are tolerating indirect active probing such as traceroute more than direct active probing such as ping.

In the next step, we issued DNS queries to obtain the host names corresponding to these IP addresses. Through DNS querying, we obtained host names of around 250K IP addresses (out of 536K IPs). We then use the host name exten-

Type	Initial				Final	
	#Traces	Reached	#Nodes	Anonymous	#Nodes	Anonymous
ICMP	306 K	93.1 %	1.0 M	68.7 %	45 K	9.7 %
TCP	306 K	73.4 %	1.0 M	72.3 %	35 K	12.5 %
UDP	306 K	45.0 %	1.5 M	86.0 %	41 K	9.4 %

Table 4. Responsiveness to indirect probes

sions to classify our IP addresses into several groups including .gov, .net, .org, .edu, and .com and look at the responsiveness of each group of nodes to active probes. Our data set had many host names with different extensions (e.g., .jp, .fr, .tr, etc.) which we did not include in the results. According to the results given in Table 3, the responsiveness ratio of routers in different types of institution networks present similarity with each other and also has a similar trend as the overall results presented earlier on the same table.

In the final step, we issued ICMP, TCP, and UDP based probes toward 306K of our IP addresses to observe router responsiveness to different indirect probes. For this, we used ICMP, TCP, and UDP based traceroute queries toward the IP addresses. Table 4 presents the results obtained in this experiment. According to the table, over 93% of our ICMP based probes reached their final destination whereas these ratios for TCP and UDP based probes were around 73% and 45% respectively. These results suggest that most network operators cooperate with ICMP based traceroute queries but more than half of the operators block UDP based traceroute queries in their domain. In addition, according to the “Anonymous” labeled columns in the Initial sections of the table (i.e., before IP alias and anonymous router processing), UDP based traceroute queries cause the highest anonymity percentage. This is also seen in the topology size difference under the “#Nodes” labeled column before processing. However, the trend in the results changes after processing indicating that IP alias and anonymous router resolution processing were effective in eliminating the large number of redundant nodes in the initial raw data. In the final data set, the differences in topology sizes are much smaller and the anonymity ratios are somehow close to the values reported for the historical data for 2005-2008 period in Table 1.

6 Conclusion

In this paper, we presented an experimental study on the responsiveness of routers to active probe messages. In our historical analysis, we observed that responsiveness reduced during the last decade. We also observed that network operators are increasingly using rate limiting to control the impact of such active probes in their network. Another observation from our study is that the destination reachability considerably reduced over time indicating that systems (i.e., routers and end hosts) are increasingly unwilling to respond to direct probes.

In the second part of our work, we observed that routers are less willing to respond to direct active probes as compared to indirect active probes. In

addition, our active direct and indirect probing based experiments showed that the responsiveness of routers changes with the type of the probes; ICMP based probes having the highest response rate and UDP based ones having the lowest response rate. Even though TCP based probes receive responses much better than UDP based ones, this type of probes sometimes raise security alerts at their destinations.

References

1. *Archipelago Measurement Infrastructure*. <http://www.caida.org/projects/ark>.
2. B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. PlanetLab: an overlay testbed for broad-coverage services. *SIGCOMM Comput. Commun. Rev.*, 33(3):3–12, 2003.
3. B. Donnet, B. Huffaker, T. Friedman, and K. Claffy. Increasing the coverage of a cooperative internet topology discovery algorithm. In *Proceedings of IFIP NETWORKING*, Atlanta, GA, USA, May 2007.
4. B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Efficient algorithms for large-scale topology discovery. In *Proceedings of ACM/SIGMETRICS*, pages 327–338, June 2005.
5. B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Deployment of an algorithm for large-scale topology discovery. *IEEE Journal on Selected Areas in Communications*, 24:2210–2220, December 2006.
6. M. Gunes and K. Sarac. Analytical IP alias resolution. In *IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, June 2006.
7. M. Gunes and K. Sarac. Resolving anonymous routers in Internet topology measurement studies. In *Proceedings of IEEE INFOCOM*, Phoenix, AZ, USA, April 2008.
8. X. Jin, W. Tu, and S.-H. Chan. Scalable and efficient end-to-end network topology inference. *IEEE Transactions on Parallel and Distributed Systems*, 19(6):837–850, June 2008.
9. E. Katz-Bassett, H. Madhyastha, J. John, A. Krishnamurthy, D. Wetherall, and T. Anderson. Studying black holes in the Internet with hubble. In *Proceedings of USENIX Symposium on Networked Systems Design and Implementation*, San Francisco, CA, USA, April 2008.
10. H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An information plane for distributed services. In *OSDI*, November 2006.
11. D. McRobb, K. Claffy, and T. Monk. *Skitter: CAIDA's macroscopic Internet topology discovery and tracking tool*, 1999. Available from <http://www.caida.org/tools/skitter/>.
12. A. Nakao, L. Peterson, and A. Bavier. A routing underlay for overlay networks. In *Proceedings of ACM SIGCOMM*, pages 11–18, Karlsruhe, Germany, August 2003.
13. Y. Shavitt and E. Shir. DIMES: Let the Internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35(5):71–74, October 2005.
14. N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP topologies using rocketfuel. *IEEE/ACM Transactions on Networking*, 12(1):2–16, February 2004.
15. S. Triukose, Z. Wen, A. Derewecki, and M. Rabinovich. Dipzoom: An open ecosystem for network measurements. In *Proceedings of IEEE INFOCOM*, Anchorage, AK, USA, May 2007.