A Second Look at Detecting Third-Party Addresses in Traceroute Traces with the IP Timestamp Option

Matthew Luckie and kc claffy
{mjl,kc}@caida.org

CAIDA, UC San Diego, USA

Abstract. Artifacts in traceroute measurement output can lead to false inferences of AS-level links and paths when used to deduce AS topology. One traceroute artifact is caused by routers that respond to traceroute probes with a source address not in the path towards the destination, i.e. an off-path address. The most well-known traceroute artifact, the third-party address, is caused by off-path addresses that map to ASes not in the corresponding BGP path. In PAM 2013, Marchetta *et al.* proposed a technique to detect off-path addresses in traceroute paths [14]. Their technique assumed that a router IP address reported in a traceroute path towards a destination was off-path if, in a subsequent probe towards the same destination, the router did not insert a timestamp into a pre-specified timestamp option in the probe's IP header. However, no standard precisely defines how routers should handle the pre-specified timestamp option, and implementations are inconsistent. Marchetta *et al.* claimed that most IP addresses in a traceroute path are off-path, and that consecutive off-path addresses are common. They reported no validation of their results. We cross-validate their approach with a first-principles approach, rooted in the assumption that subnets between connected routers are often /30 or /31 because routers are often connected with point-to-point links. We infer if an address in a traceroute path corresponds to the interface on a router that received the packet (the in-bound interface) by attempting to infer if its /30 or /31 subnet mate is an alias of the previous hop. We traceroute from 8 Ark monitors to 80K randomly chosen destinations, and find that most observed addresses are configured on the in-bound interface on a point-to-point link connecting two routers, i.e. are on-path. Because the technique from [14] reports 70.9%–74.9% of these addresses as being off-path, we conclude it is not reliable at inferring which addresses are off-path or third-party.

1 Introduction

The AS-level view of the Internet afforded by public BGP data is severely limited by a well-known visibility issue: peer-to-peer links between ASes are observable only if one of the ASes or their downstream customer provides a public view [15], which few ASes do. Traffic data collected at IXPs [2], although typically proprietary, can reveal many AS peering links established at the IXP. IXP

route-servers used to establish multilateral peering [6] may also support a query interface that reveals peering activity at the IXP. But many important peerings are established bilaterally using the IXP fabric, or at private exchange points, so traceroute retains an important role in uncovering AS-level topology [4].

Using traceroute to infer AS links and paths involves many recognized challenges [9, 16, 20]. Inferring AS paths from traceroute IP paths relies on an accurate AS inference for each IP address in traceroute, i.e. an IP2AS mapping. The most widely used IP2AS mapping technique is to associate each IP address in a path with the origin AS in a BGP path for the longest matching prefix. However, real-world practices such as (1) operators not announcing IP prefixes used to number their routers, (2) multiple ASes announcing the same prefix, and (3) organizations which own multiple ASes announcing different prefixes with different ASNs, all complicate IP2AS mapping. A further complication is routers which respond to traceroute probes using an *off-path* address; i.e. an address that does not represent the path through the router that the packet would have taken towards the destination. Off-path addresses are derived when (1) a router sets the source address of ICMP response packets to the outgoing interface used to send the response packet, and (2) that interface is not the in-bound or out-bound interface the router would have used to receive or transmit the packet if the router had forwarded the packet. A *third-party* address is an off-path address that resolves to a third-party AS that is not in the corresponding BGP path.

There has been considerable debate about the prevalence of third-party addresses. In PAM 2003, Hyun *et al.* [9] reported that third-party addresses were rare, often observed close to the destination probed, and caused by multi-homing and stale configurations. In PAM 2010, Zhang *et al.* [20] reported that the majority of false links in AS topology data derived from traceroute were due to third-party addresses. In PAM 2013, Marchetta *et al.* [14] proposed a technique to detect third-party addresses in traceroute paths using the pre-specified IP timestamp option. This option allows a host to request a timestamped response from a specific IP address (i.e., the associated router) in the path. RFC 791 [17] does not describe precisely how to implement this option, in particular whether the IP packet must actually traverse the IP interface configured with the pre-specified IP address in order to trigger the timestamp recording.

Figure 1 illustrates the technique from [14]. Using the same notation as Sherry *et al.* [18], a probe to destination G that requests B, C, D, and E include timestamps is denoted as G|BCDE. The technique from [14] assumes the behavior of a router with address B can be inferred from the response to an ICMP echo probe B|BBBB. If B embeds between one and three timestamps in the ICMP echo packet, [14] infers that the router embeds timestamps as the packet arrives or departs on the interface with address B; if B embeds four timestamps, [14] infers the router with B will insert timestamps regardless of the interface it arrived or departed from and therefore B cannot be classified as an on-path or off-path address. The technique from [14] also cannot classify routers that embed zero timestamps, remove the option, or do not reply to the ICMP probe. Because only destinations that quote the IP options in ICMP responses can be

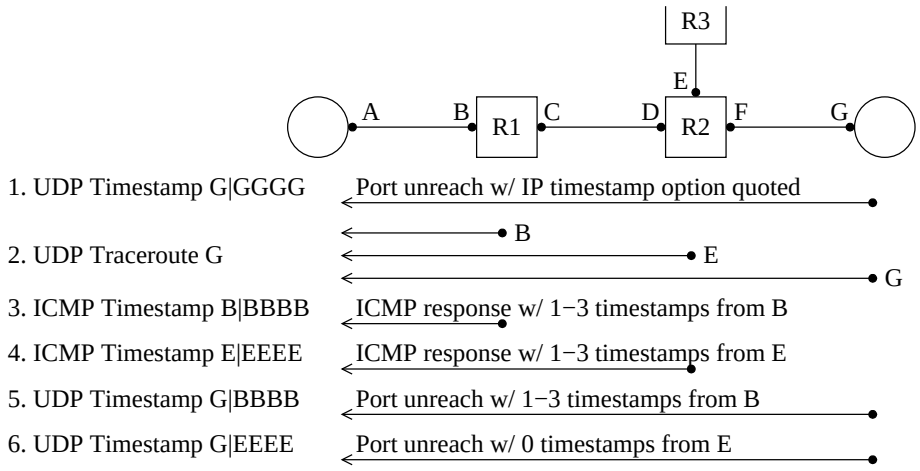

Fig. 1. Using pre-specified IP timestamps to infer third party addresses with the technique described in [14]. If G returns probes with the IP timestamp option quoted (1), then [14] evaluates the traceroute path B-E-G (2) for third-party addresses. First, [14] determines if routers will set timestamps for their IP address when a packet is sent directly to them (3, 4). For the routers that set 1-3 timestamps (i.e. set timestamps when the packet arrives and/or departs), [14] sends probes to the destination which also request those interfaces to embed a timestamp. [14] infers interface B is on-path because it does embed a timestamp (5), and infers E is a third-party address because it does not (6). However, RFC791 [17] is under-specified and it is not safe to assume E is a third-party address because it did not insert a timestamp.

evaluated for third-party addresses, the first step shown in figure 1 is to ensure a destination will respond to probes containing IP options and also quote the IP option in responses. The technique from [14] uses UDP probes for traceroute (step 2) and determining whether or not an address is on-path or off-path (steps 5, 6) because G quotes the timestamp option as the option was when G received the packet. Therefore, if B is observed in a traceroute path to G, and B embeds 1-3 timestamps to a probe G|BBBB, then the technique from [14] infers the interface with B is on-path toward G; if no timestamps are embedded by B then B is inferred by [14] to be off-path and could lead to a third-party address.

Marchetta *et al.* used their technique to estimate the prevalence of third-party IP addresses in traceroute paths. They used 53 PlanetLab nodes to obtain 12M traces towards 327K destinations among 14K ASes. They reported that most classified IP addresses in their data are off-path, and that consecutive off-path addresses are common [14]; Hyun *et al.* considered this to be a remote possibility. Further, they inferred that 17% of AS links in their dataset were inferred using third-party addresses. However, they reported no validation of their results. We revisit the effectiveness of their technique by attempting to determine which addresses in a traceroute path are likely to be the in-bound interface and thus on-path, and then examining the classification made using the technique from [14]

for these in-bound interfaces. We find most in-bound interfaces are incorrectly classified by the technique from [14] to be off-path. Further, most addresses observed in our traceroute paths are assigned by operators to the in-bound interface. We believe that the results reported in [14] are not robust because their technique is unreliable; RFC791 under-specifies how the option should be implemented and there is considerable heterogeneity in how it is implemented.

2 Method and Data

In this section we describe the method and data collected to evaluate the utility of pre-specified timestamps for inferring third party addresses. Our cross-validation of [14] involves two steps. First, we infer which addresses in a traceroute path represent the in-bound interface on the router receiving the packet, and therefore are not off-path addresses. Then, we evaluate the classification made by the technique from [14] using the pre-specified timestamp IP option for the interfaces we infer to be in-bound interfaces.

We use the *prefixscan* method implemented in scamper and described in [12] to infer which addresses in a traceroute path are the in-bound address on a router. An address B is the in-bound interface of a router in a traceroute path if we find an alias A' of the address A returned for the previous hop and A' is a /31 or /30 mate of B, i.e. the link between A and B is a point-to-point (pt2pt) link. The prefixscan method infers A and A' are aliases if (1) the IPIDs in responses to five alternating probes sent one second apart monotonically increase and differ by no more than 5,000, or (2) probes to A and A' elicit responses with a common source address. The first technique is a pairwise comparison similar to Ally [19], and the second is the Mercator technique [7]. A threshold of 5,000 allows aliases to be inferred for routers with fast moving IPID counters and has a 7.6% chance of falsely inferring aliases, in the worst case, between two routers with fast moving but overlapping counters.

Because we may falsely infer aliases when two independent counters happen to overlap when we probe them [5], or when two routers randomly generated IPID values that happened to fall within the threshold, we probe A and A' six further times approximately ten minutes apart, with five probes per round. We do not classify a link as pt2pt if any of these subsequent probes do not solicit a monotonically increasing sequence or if the IPID distance falls outside of the threshold. For each hop in a traceroute path, we prefixscan with ICMP-echo, TCP-ack, and UDP probes (in that order) to maximize our potential to infer pt2pt links. We use this ordering because in previous cross-validation efforts, we found this order to produce the most accurate inferences [13]. We believe our pt2pt inferences are robust because other researchers have previously validated IPID-based alias inferences [19, 5, 10].

Table 1 lists the eight CAIDA Archipelago (Ark) vantage points (VPs) we use for our study. We chose the eight VPs that were operational on 2 September 2013 that also provide a complete BGP view publicly. We chose these VPs because we could also evaluate traceroute-inferred and BGP-observed AS path

Ark VP	Hosting Network (AS)	Public BGP view (peer IP)
ams3-nl	RIPE NCC (3333)	RIPE rrc00 (193.0.0.56)
gva-ch	IP-Max SA (25091)	RIPE rrc04 (192.65.185.244)
nrt-jp	APAN (7660)	Routeviews 2 (203.181.248.168)
per-au	AARnet (7575)	Routeviews ISC (198.32.176.177)
sin-sg	DCS1 Pte Ltd (37989)	RIPE rrc00 (203.123.48.6)
syd-au	AARnet (7575)	Routeviews ISC (198.32.176.177)
sql-us	ISC (1280)	RIPE rrc14 (198.32.176.3)
zrh2-ch	Kantonsschule Zug (34288)	RIPE rrc12 (80.81.194.119)

Table 1. To support future study of traceroute and BGP incongruities, we chose for our measurement study 8 Ark VPs that also provide a complete BGP view publicly.

incongruities on pt2pt links, shedding further light on how incongruities from on-path addresses arise in practice. We leave this analysis for future work and invite the research community to study this problem further using these VPs.

From each VP, we randomly chose 10,000 destinations that quoted a probe's IP options in an ICMP destination unreachable message; these destinations were *useful* because they quote the IP options (step 1 in figure 1). Each VP randomly selected a different set of destinations to probe. To maximize our chances of selecting useful destinations, we selected the 2.5M of the 14.5M addresses in the ISI hitlist [1] with a score of at least 80 (where 99 represents an address that has always responded to ISI's ICMP echo probes [8]). Despite selecting destinations that were generally responsive to ping, we found that only 15.1% to 18.8% (depending on the VP) responded and echoed the pre-specified timestamp option; i.e. we tried between 53K and 66K addresses to obtain 10,000 useful destinations. Of the destinations that were not useful because we did not receive a response with a quote of the IP options, only 3.5% to 5.9% did not quote the timestamp option; another 94.1% to 96.5% of them did not respond at all. When we probed the same destinations without the timestamp option, 36.4% to 36.8% responded, implying that including a timestamp option in a UDP probe reduced the fraction of responsive destinations by at least half.

Overall, we obtained 80,004 traces containing 150,188 IP addresses, inferring 197,335 IP-level links between 7,401 ASes. Many IP interfaces are observed from multiple VPs; in our dataset, we observed 28.0% of interfaces (IP addresses) from at least two VPs even though we only used eight VPs total. We received responses to ICMP B|BBBB probes from 119,594 interfaces. For the 30,594 (25.6%) interfaces for which we received responses to ICMP B|BBBB probes from more than one VP, all VPs observed the same timestamp behavior except for 538 interfaces (1.8%). 324 of these (60.2%) were in one AS, suggesting routers in our data behaved the same regardless of probing location for ICMP B|BBBB probes. For each VP, the technique from [14] classified between 42.5% and 47.3% of interfaces in our data as appearing as on-path or off-path because they responded with 1-3 timestamps when we probed them with ICMP B|BBBB packets. In total, 77,348 of the 150,188 (51.5%) interfaces observed in our data embedded 1-3 timestamps when we probed them with ICMP B|BBBB probes.

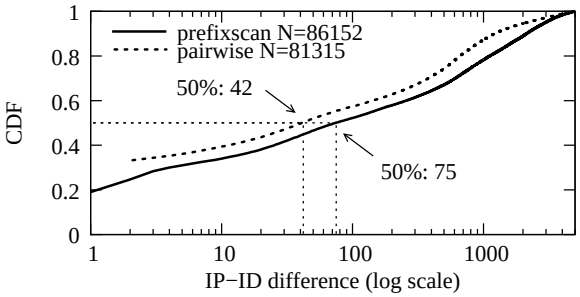

Fig. 2. IPID differences between inferred aliases A and A′. The smaller the difference, the more likely the alias inference is reliable. In our pairwise measurements, we rejected 4,837 initial pt2pt link inferences because either the IPIDs did not monotonically increase, or the counters increased too fast to reliably infer aliases. Of the 81K links remaining, 33% of the IPID values were strictly incremented between alias pairs.

We resolved the IP-level links to AS-level links using the longest matching prefix observed by peers at RouteViews; 31K (15.8%) have addresses that map to two different ASes (are inter-AS links), 153K (77.8%) have addresses that map to a single AS (are links internal to an AS), and 13K (6.4%) have addresses that are either not announced publicly, or whose longest matching prefix is originated by multiple ASes. In total, we infer 10,175 AS-level links from these traces. Of the 197,335 IP-level links, we inferred a /30 or /31 link for 86,152 links with an initial prefixscan; our followup pairwise measurements discarded 5K links because the returned IPID sequence did not meet our requirements, leaving us with 81,315 pt2pt links. Figure 2 shows the IPID differences where aliases were inferred between A and A′. The solid line corresponds to the initial prefixscan measurement that inferred a /30 or /31 mate A′ as an alias of A; we plotted the maximum IPID difference between any two samples in the sequence of five probes. 50% of the samples had a maximum difference of 75. The dashed line corresponds to the subsequent pairwise measurements; for each inferred alias, we plotted the minimum IPID difference between responses from the same probed address; i.e. the alias' IPID had to fit within the two IPID values. 50% of the samples had a minimum difference of 42, and 33.2% had a minimum difference of 2; in these latter cases the IPID of the alias fell immediately between a monotonic sequence. We therefore believe most of our pt2pt inferences are robust.

3 Results

In this section, we focus on addresses in traceroute paths that we inferred to be the address of the in-bound interface on a router and visited across a pt2pt link (i.e. were on-path). Of the 197,335 IP-level links, we inferred 81,315 (41.2%) to be pt2pt. Figure 3 plots the distribution of the fraction of in-bound interfaces in traceroutes observed by each VP. In our data, we inferred that more than half

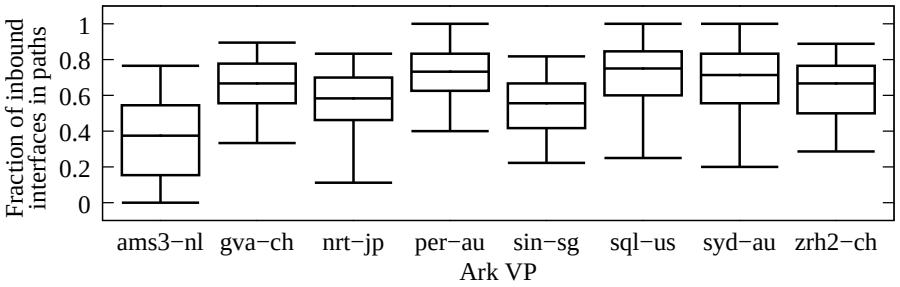

Fig. 3. Distribution of the fraction of in-bound interfaces observed by each VP at the 2nd, 25th, 50th, 75th, and 98th percentiles. For 7 of the 8 VPs, more than half of the interfaces in paths represent the in-bound interface for at least half of their traceroutes.

VP	1-3 TS (on-path)	Zero TS (off-path)	4 TS (juniper)	mixed
ams3-nl	1631 (26.1%)	4550 (72.8%)	1 (0%)	64 (1.0%)
gva-ch	1678 (26.4%)	4600 (72.3%)	0 (0%)	83 (1.3%)
nrt-jp	1543 (27.6%)	3958 (70.9%)	1 (0%)	84 (1.5%)
per-au	1547 (24.8%)	4610 (73.8%)	2 (0%)	89 (1.4%)
sin-sg	1649 (25.8%)	4657 (72.9%)	0 (0%)	80 (1.3%)
sql-us	1583 (24.8%)	4698 (73.7%)	1 (0%)	90 (1.4%)
syd-au	1524 (24.0%)	4731 (74.6%)	0 (0%)	91 (1.4%)
zrh2-ch	1404 (26.1%)	3900 (72.5%)	0 (0%)	74 (1.4%)

Table 2. Consistency of timestamps embedded by interfaces that we infer to be the in-bound interface on a pt2pt link. Between 70.9% and 74.6% of interfaces do not insert a timestamp despite being on-path. Between 1.0% and 1.5% of interfaces behaved differently depending on the destination probed (mixed column).

of the interfaces in each path were the in-bound interface for at least half of the VP's traceroutes for 7 of the 8 VPs. This is a lower bound on the actual fraction of in-bound interfaces for these VPs because some routers do not respond to our prefixscan probes that we use to infer pt2pt links. In particular, many paths from the Ark node in Amsterdam (ams3-nl) traverse AS7018 (AT&T), but none of AT&T's routers will respond to probes.

We next examine the classification made using the technique from [14] for the addresses we inferred to represent the in-bound interface on the router for paths that traversed a pt2pt link. Of the 77,348 interfaces that embedded 1-3 timestamps in the pre-specified timestamp option in response to ICMP B|BBBB probes, we inferred 29,930 (38.7%) of these to represent the in-bound interface on the router on at least one pt2pt link. However, the majority of UDP G|BBBB probes across these pt2pt links obtain zero timestamps and would be classified by [14] as off-path. In our data, between 77.1% and 90.0% of interfaces visited embedded zero timestamps, depending on the VP. Techniques relying on pre-specified timestamps to infer off-path addresses are unreliable.

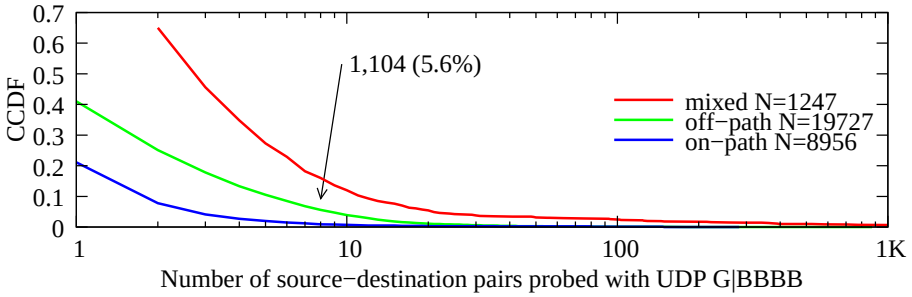

Fig. 4. CCDF of the number of source-destination pairs an interface was observed in a traceroute path, grouped by classifications made with UDP G|BBBB probes and the technique from [14]. 1,104 (5.6%) of interfaces always inferred to be off-path using the technique from [14] that we infer to represent the in-bound interface of the router were traversed in at least eight source-destination pairs in our data.

Because hop B might be observed in multiple traceroutes to destinations G_1, G_2, G_N, and therefore be counted N times, particularly for interfaces close to our VP [11], we next examine the variability of classifications made using pre-specified timestamps. Table 2 reports the number of interfaces that behave consistently regardless of the destination probed; 70.9% – 74.5% do not insert a timestamp when the in-bound interface is visited regardless of the destination probed. This result partly explains Marchetta *et al.*'s surprising result that most classified addresses in traceroute paths are off-path: it seems routers that insert 1-3 timestamps when probes are addressed to them with ICMP B|BBBB probes often do not insert timestamps when they forward UDP G|BBBB probes.

We next investigate the possibility that the prevalence of off-path inferences are due to load-balancing routers. As with Marchetta *et al.* [14], the UDP probes for the traceroute towards G contain no IP options, while the UDP probes G|BBBB to infer on- and off- path interfaces do. Routers that per-flow load-balance IPv4 packets using bytes 20-23 (where the transport header would be located if the IP header contained no options) may forward probes based on the first four bytes of a pre-specified timestamp option rather than on the first four bytes of the UDP header (source and destination ports). However, this explanation is unlikely to explain the prevalence of off-path inferences for two reasons. First, per-destination load-balancers are the most common form of load balancer, i.e. they do not consider bytes 20-23 when forwarding a packet. Augustin *et al.* reported that 70% of source-destination pairs traversed such a load balancer in their data, while 39% traversed a per-flow load balancer [3]. Second, figure 4 presents a CCDF of the number of source-destination pairs an interface was observed in a traceroute, grouped by the classifications made using the technique from [14]. 5.6% of interfaces were consistently inferred to be off-path despite being traversed by at least eight source-destination pairs with UDP G|BBBB probes. We are at least 99% confident the hop prior to B did not per-flow load balance these probes on a path avoiding B. Figure 5 shows a scatter-plot of in-

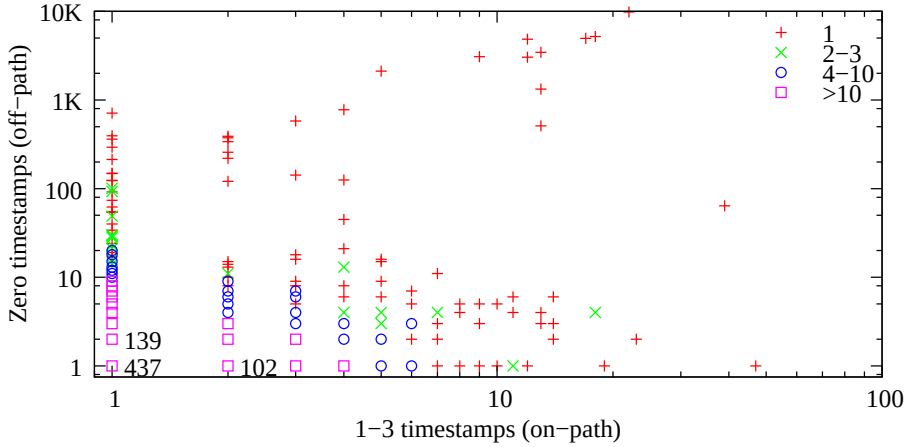

Fig. 5. Scatter plot of in-bound interfaces inferred to be on-path toward some destinations and off-path toward others (i.e. have mixed timestamp behavior). The symbol shape reflects the frequency of the on-path:off-path ratio in our data. Most interfaces are inferred to be on-path for just one source-destination pair using the technique from [14] despite being the in-bound interface of a pt2pt link.

terfaces that we inferred to be received on the in-bound interface on a pt2pt link, but which were inferred to be on-path for some source-destination pairs and off-path for others using the technique from [14]; that technique infers the majority of interfaces to be on-path for a few destinations, and off-path for most. We attempted to traverse some interfaces with hundreds of UDP G|BBBB probes; the technique from [14] inferred these interfaces to be on-path only a few times.

4 Conclusion and Future Work

Traceroute has an important role in overcoming the visibility issue of AS topology data because we have no other way of uncovering some peerings. However, researchers must first overcome traceroute artifacts such as third-party addresses which cause us to deduce false AS links and paths. Using traceroutes from eight Ark monitors to 80K randomly chosen destinations and a method derived from first principles, we showed (counter to the result in [14]) that the majority of IP addresses in traceroute paths are the in-bound interface on a pt2pt link, and that current techniques using pre-specified timestamps to infer third-party addresses are not reliable. We also release our code used to collect these measurements so others can reproduce our work. In future work, we plan to use these eight Ark VPs with public BGP data available to investigate incongruities between BGP and traceroute paths where the incongruity is inferred on a pt2pt link. Deriving a technique that accurately infers AS links from traceroute paths remains an important and currently unsolved problem.

Acknowledgments

The work was supported by U.S. NSF grant CNS-0958547, DHS S&T Cyber Security Division (DHS S&T/CSD) BAA 11-02 and SPAWAR Systems Center Pacific via N66001-12-C-0130, and by Defence Research and Development Canada (DRDC) pursuant to an Agreement between the U.S. and Canadian governments for Cooperation in Science and Technology for Critical Infrastructure Protection and Border Security. This material represents the position of the author and not of NSF, DHS, or DRDC.

References

1. IP address hitlist, PREDICT ID USC-LANDER/internet_address_hitlist_it52w 2nd Jan 2013, `http://www.isi.edu/ant/lander`
2. Ager, B., Chatzis, N., Feldmann, A., Sarrar, N., Uhlig, S., Willinger, W.: Anatomy of a large European IXP. In: SIGCOMM 2012
3. Augustin, B., Friedman, T., Teixeira, R.: Measuring load-balanced paths in the Internet. In: IMC 2007
4. Augustin, B., Krishnamurthy, B., Willinger, W.: IXPs: Mapped? In: IMC 2009
5. Bender, A., Sherwood, R., Spring, N.: Fixing Ally's growing pains with velocity modeling. In: IMC 2008
6. Giotsas, V., Zhou, S., Luckie, M., Claffy, K.: Inferring multilateral peering. In: CoNEXT 2013
7. Govindan, R., Tangmunarunkit, H.: Heuristics for Internet map discovery. In: INFOCOM 2000
8. Heidemann, J., Pradkin, Y., Govindan, R., Papadopoulos, C., Bartlett, G., Bannister, J.: Census and survey of the visible Internet. In: IMC 2008
9. Hyun, Y., Broido, A., Claffy, K.: On third-party addresses in traceroute paths. In: PAM 2003
10. Keys, K., Hyun, Y., Luckie, M., Claffy, K.: Internet-scale IPv4 alias resolution with MIDAR. IEEE/ACM Transactions on Networking 21(2) (Apr 2013)
11. Lakhina, A., Byers, J.W., Crovella, M., Xie, P.: Sampling biases in IP topology measurements. In: INFOCOM 2003
12. Luckie, M.: Scamper: a scalable and extensible packet prober for active measurement of the Internet. In: IMC 2010
13. Luckie, M., Dhamdhere, A., Claffy, K., Murrell, D.: Measured impact of crooked traceroute. CCR 14(1) (Jan 2011)
14. Marchetta, P., de Donato, W., Pescapé, A.: Detecting third-party addresses in traceroute traces with IP timestamp option. In: PAM 2013
15. Oliveira, R., Pei, D., Willinger, W., Zhang, B., Zhang, L.: In search of the elusive ground truth: the Internet's AS-level connectivity structure. In: SIGMETRICS 2008
16. Oliveira, R., Zhang, B., Zhang, L.: Observing the Evolution of Internet AS Topology. In: SIGCOMM 2007
17. Postel, J.: Internet protocol (Sep 1981)
18. Sherry, J., Katz-Bassett, E., Pimenova, M., Madhyastha, H.V., Anderson, T., Krishnamurthy, A.: Resolving IP aliases with prespecified timestamps. In: IMC 2010
19. Spring, N., Mahajan, R., Wetherall, D.: Measuring ISP topologies with Rocketfuel. In: SIGCOMM 2002. Pittsburgh, PA, USA
20. Zhang, Y., Oliveira, R., Zhang, H., Zhang, L.: Quantifying the pitfalls of traceroute in AS connectivity inference. In: PAM 2010