Detecting Third-party Addresses in Traceroute IP Paths

Pietro Marchetta, Walter de Donato, Antonio Pescapé
University of Napoli Federico II
Napoli, Italy
{pietro.marchetta,walter.dedonato,pescape}@unina.it

ABSTRACT

Traceroute is probably the most famous computer networks diagnostic tool, widely adopted for both performance troubleshooting and research. Unfortunately, traceroute is not free of inaccuracies.

In this poster, we present our ongoing work to address the inaccuracy caused by *third-party addresses*. We discuss the impact of third-party addresses on traceroute applications and present a novel active probing technique able to identify such addresses in traceroute traces. Finally, we detail preliminary results suggesting how this phenomenon has been largely underestimated.

Categories and Subject Descriptors

C.2.1 [**Computer-communication networks**]: Network Architecture and Design—*Network topology*

General Terms

Measurement

Keywords

Traceroute, Internet topology, AS-level path.

1. MOTIVATION AND SUMMARY

In the last decade many research works have used traceroute (or its variants) to infer network topological properties [5,8], and, more in general, in active monitoring approaches for anomaly detection, performance analysis, and geolocation. Traceroute is known to be affected by several inaccuracies that can produce errors or wrong assumptions when using its results to infer other information [7]: third-party (TP) addresses have been pointed out as an uncommon case mostly appearing at the border of multi-homed Autonomous Systems (ASes) [6].

To the best of our knowledge, the occurrence of TP addresses in IP paths has not been properly quantified and we believe that its impact has been largely underestimated. In this poster, to shed light on this topic, we propose a novel active probing technique based on the IP prespecified timestamp option: it allows to directly detect the presence of TP addresses in traceroute IP paths. We present a preliminary analysis conducted on about 13 K traceroute traces collected from a single vantage point. The analysis confirms the potentiality of our technique and unveils an unexpected amount of TP addresses. Finally, by performing IP-to-AS mapping on our dataset, we quantify the amount of potentially fake AS links due to TP addresses.

SIGCOMM'12, August 13–17, 2012, Helsinki, Finland.
ACM 978-1-4503-1419-0/12/08.

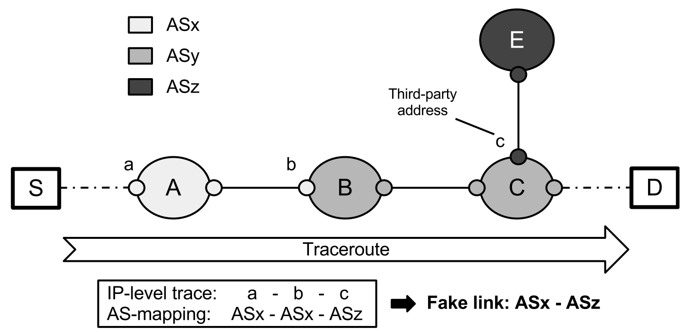


Figure 1: TP addresses inducing the inference of false AS links

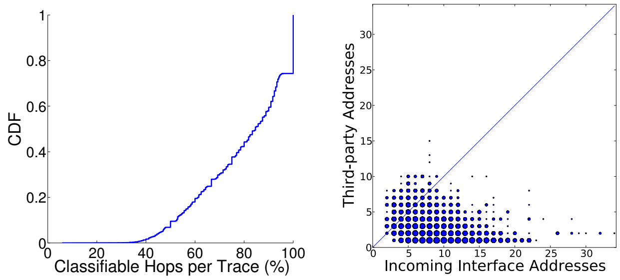
2. TP ADDRESSES AND THEIR IMPACT

The RFC1812 [2] states that the source address of an ICMP reply packet should correspond to the outgoing interface of the reply, rather than the interface on which the packet triggering the reply was received. This behavior can cause a traceroute IP path to include addresses associated to interfaces not included in the actual traversed path. Let's see an example (Fig. 1): in the trace from S to D containing the sequence (a, b, c) of IP addresses (hereafter IPs), where a and b represent the incoming interfaces of routers A and B respectively, and c is the interface used by router C to send ICMP replies to the traceroute originator, the last IP is a TP address (being associated - in this specific trace - to an interface not effectively traversed by packets from S to D).

The occurrence of TP addresses can have a significant impact on some traceroute applications. A first example are the issues occurring when mapping traced IP addresses to AS numbers, in order to discover AS level links. In this case, as shown in previous works [6], TP addresses may cause the inference of false AS links. Considering again Fig. 1: if the IP address associated to interface b belongs to ASx, and the one associated to c belongs to the ASz addressing space, then the traceroute conversion will produce a false AS link, i.e. $ASx - ASz$. Another example is related to the identification of the subnets traversed by packets along the path toward a destination, as performed by the Tracenet tool [10]: the presence of TP addresses forces the tool to exploit several heuristics to state if a detected subnet is actually on the traversed path.

3. PROPOSED APPROACH

Our approach exploits the IP prespecified Timestamp (TS) option to identify TP addresses. The TS option allows to prespecify in a single packet up to four IP addresses from which a timestamp is requested. When processing this option, most devices insert their own timestamp only if the packet passes through the interface as-



(a) Approach applicability. (b) Per-path classification.

Figure 2: Preliminary results.

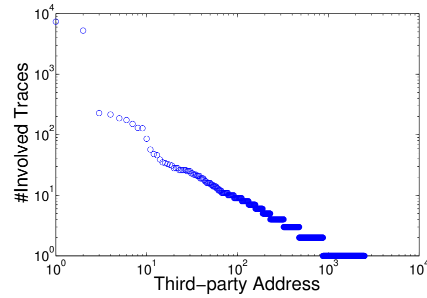


Figure 3: Third-party addresses over traces.

sociated to the prespecified address [4]. Suppose IPx owned by one of these devices and belonging to a traceroute path. In order to understand if IPx is a TP address, we target the traceroute destination with an UDP probe[1] equipped with a TS option, in which IPx is prespecified four times: if the TS option brought back into the payload of the ICMP PORT UNREACH message contains at least one timestamp, according to the behavior described above, IPx belongs to the actual IP path[2], otherwise IPx is a TP address. To be able to distinguish between routers reporting a TP address and ones not implementing the TS option, after collecting a traceroute path, we first directly probe each discovered hop IPy with an ICMP ECHO REQUEST message in which IPy is prespecified: this step allows to identify the set of IP addresses classifiable by our approach. The proposed approach would allow to avoid complex heuristics for subnet positioning in Tracenet [10] and to take into account path fluctuations affecting the Palmtree [9] alias resolution technique.

4. PRELIMINARY RESULTS

To evaluate the proposed approach, we randomly selected 50 K destinations among the ones showing stable responsiveness according to the PREDICT project [1]. About 13 K destinations replied to UDP probes carrying the TS option. Focusing our attention on this subset, we launched a traceroute campaign, from our laboratory at University of Napoli, collecting a final dataset of 32 K IPs. Fig. 2(a) shows the percentage of hops classifiable in each trace: all the hops resulted classifiable in 26% of traces. In general, the proposed approach showed a promising level of applicability: on average and in the worst case, we were able to classify 80% and 38% of hops per trace, respectively. Furthermore, it classified 2.5 K IPs (7.8%) as TP addresses in at least one trace. Such value appears surprisingly high, suggesting that TP addresses are not so uncommon as previously hypothesized [6]. Fig. 2(b) compares, for each trace, the number of addresses classified as TP or associated to incoming interfaces (IN) and shows how in 324 traces (2.43% of the total) we found more TP addresses than IN IPs.

As shown in Fig. 3, a single TP address may affect several traces. The two most common TP addresses appeared respectively in 7.3 K and 5.2 K traces, essentially because they were located close to the vantage point: on average, a TP address appeared in 7.8 traces.

Finally, mapping each address to the owner AS [3], we found that our dataset covers 1.9 K distinct ASes. By considering two consecutive hops in a traceroute trace mapped to distinct ASes as a potential AS link, we extracted about 2 K AS links. Among these, 527 (26%) involved TP addresses, thus being potential source of inaccuracies according to the scenario depicted in Fig. 1. Although not necessarily all these links were false, such high percentage suggests that TP addresses may represent a significant cause of AS maps distortion. Our approach allows to identify the subset of links to carefully consider when the objective is to draw the AS level map exploiting the information obtained with traceroute.

[1] UDP probes allow to avoid ambiguities caused by the reverse path.

[2] In this case, IPx is probably associated to an incoming interface.

5. ONGOING AND FUTURE WORK

This analysis represents a first snapshot of an ongoing research work. We are extending our study in different ways: (i) adopting different probes to obtain a wider support; (ii) using a larger dataset, to better evaluate the applicability of our technique; (iii) exploiting multiple vantage points, to detect when IPs act as incoming and TP addresses. (iv) deepening the TP addresses impact on AS level topology exploiting ground truth. We plan to implement and release to the research community a modified traceroute version able to automatically discover the IP path and label IPs as TP address.

6. REFERENCES

[1] Scrambled internet trace measurement dataset. *PREDICT ID USC-LANDER / internet-address-hitlist, 2009-12-23 2012-02-14*.

[2] F. Baker. Ietf rfc1812: Requirements for ip version 4 routers. *http://ds.internic.net/rfc/rfc1812.txt*.

[3] T. Cymru. Ip to asn mapping. *http://www.team-cymru.org/Services/ip-to-asn.html*, 2012.

[4] W. de Donato, P. Marchetta, and A. Pescapé. A hands-on look at active probing using the ip prespecified timestamp option. In *PAM'12, Vienna, Austria*, 2012.

[5] B. Donnet and T. Friedman. Internet topology discovery: a survey. *IEEE Communications Surveys and Tutorials*, 9(4), 2007.

[6] Y. Hyun, A. Broido, and K. Claffy. On third-party addresses in traceroute paths. In *PAM'03, San Diego, CA, USA*, 2003.

[7] Y. Hyun, A. Broido, and K. Claffy. Traceroute and bgp as path incongruities. Technical report, CAIDA, University of California, San Diego, 2003.

[8] P. Marchetta, P. Mérindol, B. Donnet, A. Pescapè, and J.-J. Pansiot. Topology discovery at the router level: A new hybrid tool targeting isp networks. *JSAC*, 29(9):1776–1787, 2011.

[9] M. Tozal and K. Sarac. Palmtree: An ip alias resolution algorithm with linear probing complexity. *Computer Communications*, 2010.

[10] M. Tozal and K. Sarac. Tracenet: an internet topology data collector. In *Proc. of IMC'10*. ACM, 2010.