vrfinder: Finding Outbound Addresses in Traceroute

Alexander Marder
amarder@caida.org
CAIDA / UC San Diego
La Jolla, California, USA

Bradley Huffaker
bradley@caida.org
CAIDA / UC San Diego
La Jolla, California, USA

Matthew Luckie
mjl@wand.net.nz
University of Waikato
Hamilton, New Zealand

kc claffy
kc@caida.org
CAIDA / UC San Diego
La Jolla, California, USA

ABSTRACT

Current methods to analyze the Internet’s router-level topology with paths collected using traceroute assume that the source address for each router in the path is either an inbound or off-path address on each router. In this work, we show that outbound addresses are common in our Internet-wide traceroute dataset collected by CAIDA’s Ark vantage points in January 2020, accounting for 1.7% – 5.8% of the addresses seen at some point before the end of a traceroute. This phenomenon can lead to mistakes in Internet topology analysis, such as inferring router ownership and identifying interdomain links. We hypothesize that the primary contributor to outbound addresses is Layer 3 Virtual Private Networks (L3VPNs), and propose vrfinder, a technique for identifying L3VPN outbound addresses in traceroute collections. We validate vrfinder against ground truth from two large research and education networks, demonstrating high precision (100.0%) and recall (82.1% – 95.3%). We also show the benefit of accounting for L3VPNs in traceroute analysis through extensions to bdrmapIT, increasing the accuracy of its router ownership inferences for L3VPN outbound addresses from 61.5% – 79.4% to 88.9% – 95.5%.

CCS CONCEPTS

• **Networks** → **Network measurement**; *Public Internet*.

ACM Reference Format:

Alexander Marder, Matthew Luckie, Bradley Huffaker, and kc claffy. 2020. vrfinder: Finding Outbound Addresses in Traceroute. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '20 Abstracts)*, June 8–12, 2020, Boston, MA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3393691.3394227>

1 INTRODUCTION

Traceroute is the primary method available for uncovering the Internet’s router-level topology, but it presents insidious epistemological challenges. Traceroute is essentially a hack that exploits common router behavior. When a router discards a packet due to the expiration of the TTL field in the IP packet header, it sends

Figure 1: Traceroute observes a sequence of router interfaces between a Vantage Point (VP) and a Destination (DST). Logically, the inbound interfaces face the VP, while the outbound interfaces face the destination.

an ICMP Time Exceeded response packet to the originator of the dropped packet. When routers respond to traceroute probes, the ICMP response contains a router interface IP address in the source field, exposing an address used on the router. The traceroute tool exploits this behavior by sending probes with incrementing TTL values toward a single destination, attempting to induce responses from each router on the path from the source to the destination.

Most routers report the address of the interface that received the traceroute probe, known as the *inbound address*, in accordance with common network operator expectations [11], even if the router sends the response using a different interface. Other routers adhere to the specification in RFC 1812, which mandates that routers report the address of the interface that transmits the ICMP response. If the route back to the source is not via the inbound interface, the router will report an *off-path address*, as illustrated in Fig. 1. We do not expect, however, that a source address in a traceroute response represents an interface that would have been used to forward the probe toward its destination, known as the *outbound interface*. The conventional interpretation, that routers report inbound or off-path interface addresses in response to TTL-expiring probes, functions as an axiom of traceroute analysis. Reliance on this axiom pervades the literature [1–6, 8–10].

We summarize the results of our effort [7] to develop and validate a method to infer the presence of outbound interfaces in traceroute paths primarily caused by L3VPNs. Outbound addresses potentially confound any traceroute analysis that relies on conventional traceroute assumptions. This paper makes the following contributions: (1) we present vrfinder, a novel technique for identifying L3VPN outbound addresses in traceroutes; (2) we validate our inferences for IPv4 and IPv6 traceroute collections against ground truth from two R&E networks, with 100.0% PPV and 82.1% – 95.3% TPR; and (3) we extend bdrmapIT [6] to account for outbound interfaces, increasing its accuracy on L3VPN networks by 9.5% – 29.1%.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '20 Abstracts, June 8–12, 2020, Boston, MA, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7985-4/20/06.

<https://doi.org/10.1145/3393691.3394227>

Figure 2: Intuition for inferring outbound addresses in traceroute paths. We infer an outbound address on R_2 when we observe an inbound address in the same subnet on adjacent router R_3 (Path C).

2 INTUITION

The Internet is assembled by connecting routers that speak IP. These connections are either point-to-point between two routers, or point-to-multipoint between many routers that exchange traffic using a shared fabric, such as at an Internet Exchange Point (IXP). Because IPv4 addresses are scarce resources, point-to-point connections in IPv4 are typically numbered out of a /30 or /31. In IPv6, operators most commonly address point-to-point links from /64, /126, and /127 subnets, although they may use any subnet size between 64 and 127. Host addresses, i.e., addresses usable as interface addresses, are consecutive addresses for /30, /31, /126, and /127 subnets. While there is no such guarantee for /64 subnets, current best practice is to assign consecutive addresses to the point-to-point link interfaces.

As discussed in §1, current methodologies that use traceroute paths typically assume that traceroute observes mostly inbound addresses, with some off-path addresses. Fig. 2 illustrates two expected traceroute observations in these cases (Paths A and B). Because traceroute paths mostly observe inbound addresses, our intuition is that we may infer the presence of an outbound address when it is followed by an inbound address, as these addresses will be part of the same subnet. In Fig. 2 Path C, R_2 replies with outbound address 192.0.2.13 and R_3 replies with inbound address 192.0.2.14 – consecutive addresses in 192.0.2.12/30. This technique is the core of the *vrfinder* methodology.

vrfinder requires that the router following an outbound address replies with its inbound address, but it might reply with an off-path address instead. If R_3 replies with the off-path address 192.0.2.25 (Fig. 2 Path D), then *vrfinder* will not infer that R_2 used an outbound address when it replied with 192.0.2.13 because the addresses are not consecutive (a false negative). If R_3 does not reply at all, as illustrated by Path E in Fig. 2, then *vrfinder* will also not infer R_2 used an outbound address. In [7] we also discuss apparent outbound addresses that result from forwarding loops, and /31 and /127 subnets.

3 RESULTS

We evaluated *vrfinder* on the traceroute collections used for the January 2020 CAIDA ITDK. Of the addresses in the middle of traceroutes in the collection, the inferred L3VPN outbound addresses accounted for 5.8% of the IPv4 addresses and 1.7% of the IPv6 addresses. These results demonstrate the prevalence of L3VPN outbound addresses in the IPv4 traceroutes, and suggest that L3VPN outbound addresses appear less frequently in the IPv6 traceroutes.

We validated our inferences against ground truth from Internet2 and REANNZ, two large research and education networks. Overall, *vrfinder* achieved 100.0% PPV and 82.1% – 95.3% TPR for our two ground truth datasets. The missing inferences, especially for REANNZ, highlight that *vrfinder* cannot identify an outbound address when the subsequent router does not respond to traceroute.

We also spoke with network operators at five large commercial ISPs and a European R&E network, asking them about specific outbound address inferences from their IPv4 address space. One operator tested our L3VPN outbound address assumption on the ISP’s network, confirming that the L3VPN egress router always reported the outbound address, and two other operators confirmed all but one of our outbound inferences in their networks. However, operators at the other three ISPs told us that none of their L3VPNs should appear in our traceroute collection and informed us that many of our mistaken inferences resulted from forwarding loops. These conversations suggest that *vrfinder*’s outbound inferences might not be as accurate as the evaluation against Internet2 and REANNZ indicates.

Finally, we extended *bdrmapIT* to account for outbound addresses when inferring AS ownership for the IP addresses in traceroute. Our extensions improved the AS ownership accuracy for the L3VPN outbound addresses in our ground truth by 9.5% – 29.1%, and produced nearly identical results for other ground truth datasets. Overall, these results show that detecting and accounting for L3VPN outbound addresses can improve the accuracy of AS ownership inferences for the outbound addresses.

REFERENCES

- [1] Lisa D Amini, Anees Shaikh, and Henning G Schulzrinne. 2002. Issues with Inferring Internet Topological Attributes. In *Internet Performance and Control of Network Systems III*.
- [2] Brice Augustin, Timur Friedman, and Renata Teixeira. 2007. Measuring load-balanced paths in the Internet. In *IMC*. San Diego, CA, USA, 149–160.
- [3] Mehmet H Gunes and Kamil Sarac. 2006. Analytical IP Alias Resolution. In *Communications, 2006. ICC’06. IEEE International Conference on*, Vol. 1.
- [4] Matthew Luckie, Amogh Dhamdhare, Bradley Huffaker, David Clark, and kc claffy. 2016. *bdrmap*: Inference of Borders Between IP Networks. In *IMC*.
- [5] Zhuoqing Morley Mao, Jennifer Rexford, Jia Wang, and Randy H Katz. 2003. Towards an Accurate AS-Level Traceroute Tool. In *Conference on Applications, technologies, architectures, and protocols for computer communications*.
- [6] Alexander Marder, Matthew Luckie, Amogh Dhamdhare, Bradley Huffaker, kc claffy, and Jonathan M. Smith. 2018. Pushing the Boundaries with *bdrmapIT*: Mapping Router Ownership at Internet Scale. In *IMC*.
- [7] Alexander Marder, Matthew Luckie, Bradley Huffaker, and kc claffy. 2020. *vrfinder*: Finding Outbound Addresses in Traceroute. In *SIGMETRICS*.
- [8] Vern Paxson. 1997. End-to-End Internet Packet Dynamics. In *ACM SIGCOMM CCR*, Vol. 27.
- [9] Rob Sherwood, Adam Bender, and Neil Spring. 2008. DisCarte: A Disjunctive Internet Cartographer. In *SIGCOMM*. 303–314.
- [10] Neil Spring, Mira Dontcheva, Maya Rodrig, and David Wetherall. 2004. *How to Resolve IP Aliases*. Technical Report UW-CSE-TR 04–05–04. University of Washington.
- [11] Richard A Steenberg. 2009. A Practical Guide to (Correctly) Troubleshooting with Traceroute. *NANOG* (2009).