Trust, But Verify, Operator-Reported Geolocation

Katherine Izhikevich
UC San Diego
katherine@ucsd.edu

Ben Du
UC San Diego
bendu@ucsd.edu

Manda Tran
UC Los Angeles
mandat@ucla.edu

Sumanth Rao
UC San Diego
svrao@ucsd.edu

Alisha Ukani
UC San Diego
aukani@ucsd.edu

Liz Izhikevich
UC Los Angeles
lizhikev@ucla.edu

Abstract

To geolocate an IP address, current methodologies often rely upon a global, human-reported set of vantage points (i.e., probes) to measure from and verify accuracy against. In this work, we analyze how often and where human-reported geolocations are incorrect.

1 Introduction

IP address geolocation plays a central role in networking and security: understanding the Internet's topology to improve last mile latency [2, 10], filtering traffic to prevent cyber attacks [24], and identifying malicious activity [18, 22], all rely on geolocation. To geolocate an IP address, methodologies often rely upon a global, human-reported set of vantage points (i.e., probes) to measure from and verify accuracy against [14, 16, 20, 25].

While necessary, relying on human operators to report and maintain accurate probe geolocation can introduce human-error to the geolocation process. For example, if an operator moves their probe, but fails to update its location metadata, geolocation methodologies relying on the outdated information risk inaccurate conclusions. To account for incorrect metadata, a few geolocation studies take an additional precaution to filter out probes reporting unlikely locations [7, 11]. However, probes with misreported geolocations are still not well understood, including where they are most likely to be located, what their magnitude of error is, and how likely they are to eventually be corrected. As operator-reported geolocations continue to serve as a foundation for many studies [1, 4, 6, 8, 9, 12, 15, 17, 23], it is critical that we understand the successes and pitfalls of our community-contributed datasets and work towards an accurate ground truth.

In this work, we provide an in-depth analysis of operator misreported geolocation. We apply a simple methodology that conservatively infers the set of probes that do not respond from their operator-reported geolocation. The methodology (1) uses speed of Internet calculations to derive a minimum bound of expected latency between two vantage points—the target vantage point and a vantage point we control—and (2) applies the minimum bound to infer whether operator-reported geolocations are physically possible.

Between May 2024 and March 2025, we deploy our methodology from 294 globally distributed probes, to measure the validity RIPE Atlas [21]. Our methodology infers a lower bound of 470 probes (3.96%) are likely not responding from their reported location between May 2024 and March 2025, as doing so would require exceeding the physical limits of Internet speed (e.g., the speed of light in fiber). We find errors in misreported geolocation are substantial; roughly 50% of violating probes likely responded at least 1,600 kilometers away from their reported location. Finally, we release a list of 664 likely-inaccurately-geolocated RIPE Atlas probes from the past five years and maintain a weekly-updated list of currently-violating probes at `https://github.com/kizhikevich/violating_ripe_probes`.

2 Methodology

Our objective is to identify whether a probe responding from its operator-reported geolocation is physically impossible. We rely on simple physics to infer whether measured latency between our measurement source and the destination's reported location exceeds the speed at which light travels in either a fiber optic cable or an inter-satellite laser. We measure latency using a network of probes operated by collaborators (Ark) and historical RIPE Atlas measurement data.

Vantage Points. Our methodology requires measuring latency to and from vantage points with validated locations. First, we use Caida's Ark [3], a set of 294 globally distributed vantage points whose exact latitude and longitude we verify directly with Caida's operators. Ark nodes are in 194 autonomous systems, across 218 cities, in 68 countries, on 6 continents. A map of Ark nodes can be found in [3]. Ark is open to researchers upon request. Additionally, we use seven RIPE operated "central" servers, located in Fremont, California, USA; Newark, New Jersey, USA; Singapore; two in Amsterdam, Netherlands; and two in Nuremberg, Germany.

Latency Measurement. First, we measure the latency from all of our vantage points (VPs) to all 12K RIPE Atlas probes. We use the Ark nodes to ICMP ping every RIPE probe

twice a month from May 2024 to March 2025. Since RIPE Atlas is volunteer-run and intended for network measurements, ICMP filtering is unlikely to occur. Between every <VP, probe> pair, we save the minimum round trip time (RTT) of all pings as the representative RTT because the speed of light in optical fiber or air imposes a hard lower bound. Unlike average RTTs, which can be inflated due to queuing or congestion, the minimum RTT reflects the fastest observed path and cannot be artificially reduced. For RIPE central servers, we use their API to collect historical ping data from each RIPE Atlas probe to each server until September 2024, when RIPE stopped running measurements to their central servers. We collect the minimum RTTs three times a day (6h,12h,18h UTC), on one day a week, over the past five years of data. Continuously unresponsive probes are filtered out.

Estimating Distance Between Source and Destination. Second, to help derive the theoretical minimum latency between a vantage point and probe, we estimate the minimum (direct) distance. We calculate the direct distance between the reported coordinates of every RIPE Atlas probe to every RIPE central server. RIPE does not disclose the exact locations of the seven central servers, so we calculate a maximum error radius around each city center. We subtract the error radius from the direct distance. We also calculate the direct distance between the locations reported by the probes on the day we send pings to the location of our Ark vantage points. For Ark nodes we do not need an error radius, since we verify the exact location of the vantage points with our collaborators.

Deriving Speed of Internet Theoretical RTTs. Once we estimate the minimum distance between each probe and server, we convert it to a minimum theoretical RTT. To the best of our knowledge, Starlink satellites are the theoretically fastest networking infrastructure due to their use of inter-satellite lasers, which operate at the speed of light in a vacuum [5]. To find probes hosted over Starlink, we look for probes whose IPv4 or IPv6 addresses are announced by Starlink's autonomous system (ASN 14593) [19]. For Starlink-hosted probes, we conservatively estimate the theoretical *absolute minimum* RTT as $direct\_distance/c$, where c is the speed of light in a vacuum, as an extra precaution for being conservative and estimating the lower bound. Otherwise, we calculate the theoretical *absolute minimum* RTT if the pings were sent at the speed of light in optical fiber, $direct\_distance/\frac{2}{3}c$ [5]. In the remainder of the text we refer to the speed of Internet as SOI.

Identifying Violating Probes. We identify an RIPE Atlas probe as "violating" (i.e., we believe it does not respond from its operator-reported geolocation) if any RIPE Atlas probe measured latency is lower than the theoretical minimum latency from the self-reported geolocation to any of the seven RIPE central servers or any of the 294 Ark servers. For RIPE central server historical data, we take note of when within the past 5 years the RTT violation happened, if not presently occurring. Furthermore, since RIPE halted their built-in measurements to their central servers, we continuously check whether the last set of violating probes from September 2024 have updated their location or disconnected from the platform. If not, we label them as still violating.

Country	# Violations	% VPs
Germany	51	2.69%
USA	27	1.48%
Netherlands	9	1.32%
France	8	0.75%
Russia	7	1.34%
...	...	...
UAE	3	11.11%
Zambia	1	25.00%
Mozambique	1	33.33%
Lesotho	1	100.00%
Botswana	1	100.00%

Table 1: Locations with Misreported Probes—As of March 2025, Violating probes (VPs) are most likely to occur in countries with a large number of RIPE Atlas probes, including Germany and the USA. Countries with the largest fraction of misreported probes are near southern Africa.

Precautions to Minimize False Positives. We take precautions to avoid overclaiming and false positives, at the expense of increasing false negatives. For example, it is possible that a probe's operator-reported geolocation is accurate, but a middlebox located away from the operator-reported geolocation might be responding on the probe's behalf. To assess the extent of this issue, we run ping measurements from present-day violating RIPE Atlas probes to Ark servers and find that 67/159 probes still violated the SOI constraint. These 67 probes should be excluded from use as either vantage points or measurement targets. The remaining 92 probes, while not suitable as measurement targets, may still be acceptable for use as vantage points.

We never attempt to identify the true geolocation of a probe, which is an open problem [7]. We only identify when the probe's response from an operator-reported geolocation is physically impossible.

3 Results

Between May 2024 and March 2025, we conservatively infer that at least 470 unique probes violate speed of Internet. Of these, we primarily analyze the 159 probes that remain in violation as of March 2025.

3.1 Distribution of Violating Probes

Violating probes exist all over the world. In Table 1, we list the countries with the most violating probes, nearly half of which originate in Germany (32%) or the USA (17%). Within the US, violating probes are broadly distributed. We infer that violating probes in Kansas are reported to be more than 200 km from the geographic center of the US [13], suggesting that the geolocation is likely misreported, not that RIPE automatically used the default US location.

Countries in southern Africa are most likely to host at least one violating probe. Table 1 lists the top 5 countries with the highest proportion of probes violating the speed of Internet. Many of the top 5 countries are in southern Africa: Botswana (100% of probes violated), Lesotho (100%), Mozambique (33.33%), and Zambia (25%). Notably, the southern African countries have extremely low vantage point coverage to begin with (i.e., only one to four probes each), thereby exacerbating the effects of a violating probe.

The majority of violating probes respond from at least 1,600 km from their operator-reported geolocation. For every <probe, vantage point> pair that violates the SOI constraint, we calculate the minimum error of distance by (1) selecting all VPs that violated SOI, (2) multiplying the minimum observed RTT by SOI, to derive the furthest distance, d, a probe could be if it traveled at exactly SOI, (3) calculate the theoretical distance, d_{theory}, between our VP and the operator-reported probe location, (4) subtract d from d_{theory}, recording the minimum distance error.

We analyze the full distribution of distance errors: 80% exceed 160 km, and 20% are greater than 4,800 km. The large distance errors suggest violating probes are likely on different continents, as we discuss in Section 3.2.1.

3.2 Understanding Contributing Factors

Through a manual analysis of reported geolocations and round trip times, we discover that tardiness in updating geolocation after a move, and initial misreporting of geolocation are likely contributors to SOI violations.

3.2.1 Tardiness. At times, probes reporting RTTs that violate SOI are simply late to report a move. Figure 1 shows longitudinal ping measurements from probe 822 whose operator reported a location in Germany. From 2019 to 2022, its RTTs to the Nuremberg RIPE servers were a minimum of 17.2 ms, indicating the probe was likely within 1,700 km of Nuremberg. After a disconnect between November 2021 and February 2022, the probe reported a minimum RTT of 23 ms to the Singapore RIPE server, indicating it was likely within 2,400 km of Singapore. However, until April 2023, the probe reported a location in Germany, which is not within 2,400 km of Singapore. In April 2023, the operator updated their metadata to Brunei, which is within 2,400 km of Singapore. The probe no longer reports RTTs in violation of the SOI.

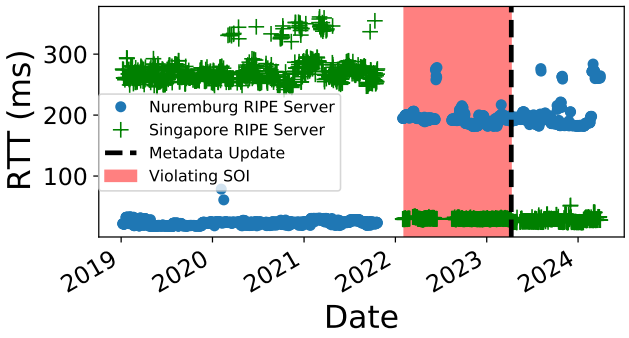

Figure 1: Moving from Germany to Brunei Case Study— The operator's probe reports a SOI violation between February 2022 and April 2023, before updating their location metadata from Germany to Brunei in April 2023.

3.2.2 Initial Misreporting. Some long-term violations stem from probes configured with incorrect geolocation metadata. We examine Probe 1000011, which violated speed of Internet from October 2019 to June 2023. It reported a location in Marina Del Ray, CA until June 2023, when it switched to Arlington, VA. However, its minimum RTT to all RIPE servers remained stable—unexpected for a cross-country move. Moreover, RTTs to both Fremont and Newark RIPE servers violated speed of Internet under the Marina Del Ray location, but no longer did once the location was updated. This suggests the probe was likely near New Jersey all along from 2019 until 2023. The location change coincides with an upstream provider switch from Cogent to Lumen, hinting that a network change may have prompted the operator to verify the probe's configuration and update the geolocation.

4 Future Work

Our work shows that operator-reported geolocation is not always trustworthy. Future work could verify how the prevalence of violating probes affects vantage point platforms other than RIPE Atlas. Other methods should consider how to distinguish between a probe's true location and that of a device responding on its behalf (e.g., a middlebox).

References

[1] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. 2015. Lessons learned from using the ripe atlas platform for measurement research. *ACM SIGCOMM Computer Communication Review* 45, 3 (2015), 35–42.

[2] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. 2017. Dissecting Last-mile Latency Characteristics. *SIGCOMM Comput.*

Commun. Rev. 47, 5 (oct 2017), 25–34. https://doi.org/10.1145/3155055.3155059

[3] CAIDA. 2024. Archipelago (Ark) Measurement Infrastructure. https://www.caida.org/projects/ark/.

[4] Massimo Candela, Enrico Gregori, Valerio Luconi, and Alessio Vecchio. 2019. Using RIPE atlas for geolocating IP infrastructure. *IEEE Access* 7 (2019), 48816–48829.

[5] Aizaz U Chaudhry and Halim Yanikomeroglu. 2022. Optical Wireless Satellite Networks versus Optical Fiber Terrestrial Networks: The Latency Perspective: Invited Chapter. In *30th Biennial Symposium on Communications 2021*. Springer, 225–234.

[6] Lorenzo Corneo, Maximilian Eder, Nitinder Mohan, Aleksandr Zavodovski, Suzan Bayhan, Walter Wong, Per Gunningberg, Jussi Kangasharju, and Jörg Ott. 2021. Surrounded by the clouds: A comprehensive cloud reachability study. In *Proceedings of the Web Conference 2021*. 295–304.

[7] Omar Darwich, Hugo Rimlinger, Milo Dreyfus, Matthieu Gouel, and Kevin Vermeulen. 2023. Replication: Towards a Publicly Available Internet scale IP Geolocation Dataset. In *Proceedings of the 2023 ACM on Internet Measurement Conference*. 1–15.

[8] Lily Davisson, Joakim Jakovleski, Nhiem Ngo, Chau Pham, and Joel Sommers. 2021. Reassessing the constancy of end-to-end internet latency. In *IFIP Network Traffic Measurement and Analysis Conference*.

[9] Rodérick Fanou, Pierre Francois, and Emile Aben. 2015. On the diversity of interdomain routing in africa. In *Passive and Active Measurement: 16th International Conference, PAM 2015, New York, NY, USA, March 19-20, 2015, Proceedings 16*. Springer, 41–54.

[10] Romain Fontugne, Anant Shah, and Kenjiro Cho. 2020. Persistent Last-mile Congestion: Not so Uncommon. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 420–427. https://doi.org/10.1145/3419394.3423648

[11] Manaf Gharaibeh, Anant Shah, Bradley Huffaker, Han Zhang, Roya Ensafi, and Christos Papadopoulos. 2017. A look at router geolocation in public and commercial databases. In *Proceedings of the 2017 Internet Measurement Conference*. 463–469.

[12] Petros Gigis, Vasileios Kotronis, Emile Aben, Stephen D Strowes, and Xenofontas Dimitropoulos. 2017. Characterizing user-to-user connectivity with RIPE Atlas. In *Proceedings of the 2017 Applied Networking Research Workshop*. 4–6.

[13] Google. 2024. countries.csv. https://developers.google.com/public-data/docs/canonical/countries_csv.

[14] Bamba Gueye, Artur Ziviani, Mark Crovella, and Serge Fdida. 2004. Constraint-based geolocation of internet hosts. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. 288–293.

[15] Thomas Holterbach, Cristel Pelsser, Randy Bush, and Laurent Vanbever. 2015. Quantifying interference between measurements on the RIPE Atlas platform. In *Proceedings of the 2015 Internet Measurement Conference*. 437–443.

[16] Zi Hu, John Heidemann, and Yuri Pradkin. 2012. Towards geolocation of millions of IP addresses. In *Proceedings of the 2012 Internet Measurement Conference*. 123–130.

[17] Mattia Iodice, Massimo Candela, and Giuseppe Di Battista. 2019. Periodic path changes in RIPE atlas. *IEEE Access* 7 (2019), 65518–65526.

[18] Katherine Izhikevich, Geoffrey M Voelker, Stefan Savage, and Liz Izhikevich. 2024. Using Honeybuckets to Characterize Cloud Storage Scanning in the Wild. (2024).

[19] Liz Izhikevich, Manda Tran, Katherine Izhikevich, Gautam Akiwate, and Zakir Durumeric. 2024. Democratizing LEO Satellite Network Measurement. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8, 1 (2024), 1–26.

[20] Ethan Katz-Bassett, John P John, Arvind Krishnamurthy, David Wetherall, Thomas Anderson, and Yatin Chawathe. 2006. Towards IP geolocation using delay and topology measurements. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. 71–84.

[21] RIPE NCC. 2024. RIPE Atlas API. https://atlas.ripe.net/coverage/.

[22] Audrey Randall, Enze Liu, Ramakrishna Padmanabhan, Gautam Akiwate, Geoffrey M. Voelker, Stefan Savage, and Aaron Schulman. 2021. Home is where the hijacking is: understanding DNS interception by residential routers. In *Proceedings of the 21st ACM Internet Measurement Conference* (Virtual Event) *(IMC '21)*. Association for Computing Machinery, New York, NY, USA, 390–397. https://doi.org/10.1145/3487552.3487817

[23] Andrei M Sukhov and Artem V Onoprienko. 2014. Evaluating the effectiveness of geographic routing based on RIPE Atlas data. In *2014 22nd Telecommunications Forum Telfor (TELFOR)*. IEEE, 107–110.

[24] Gerry Wan, Liz Izhikevich, David Adrian, Katsunari Yoshioka, Ralph Holz, Christian Rossow, and Zakir Durumeric. 2020. On the Origin of Scanning: The Impact of Location on Internet-Wide Scans. In *ACM Internet Measurement Conference*.

[25] Bernard Wong, Ivan Stoyanov, and Emin Gün Sirer. 2007. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts.. In *NSDI*, Vol. 7. 23–23.