On the Accuracy of Country-Level IP Geolocation

Ioana Livadariu
Simula Metropolitan
ioana@simula.no

Thomas Dreibholz
Simula Metropolitan
dreibh@simula.no

Anas Saeed Al-Selwi
Simula Metropolitan
anasal@simula.no

Haakon Bryhni
Simula Metropolitan
haakonbryhni@simula.no

Olav Lysne
Simula Metropolitan
olav.lysne@simula.no

Steinar Bjørnstad
Simula Metropolitan
steinar@simula.no

Ahmed Elmokashfi
Simula Metropolitan
ahmed@simula.no

ABSTRACT

The proliferation of online services comprised of globally spread microservices has security and performance implications. Understanding the underlying physical paths connecting end points has become important. This paper investigates the accuracy of commonly used IP geolocation approaches in geolocating end-to-end IP paths. To this end, we perform a controlled measurement study to collect IP level paths. We find that existing databases tend to geolocate IPs that belong to networks with global presence and those move between networks erroneously. A small percentage of IP geolocation disagreement between databases results in a significant disagreement when geolocating end-to-end paths. Geolocating one week of RIPE traceroute data validates our observations.

CCS CONCEPTS

• **Networks → Network measurement**;

KEYWORDS

IP Geolocation; Geolocation Databases; Geolocation Approaches

ACM Reference Format:
Ioana Livadariu, Thomas Dreibholz, Anas Saeed Al-Selwi, Haakon Bryhni, Olav Lysne, Steinar Bjørnstad, and Ahmed Elmokashfi. . In *Applied Networking Research Workshop (ANRW '20), July 27–30, 2020, Online (Meetecho), Spain.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3404868.3406664

ANRW '20, July 27–30, 2020, Online (Meetecho), Spain

ACM ISBN 978-1-4503-8039-3/20/07…$15.00
https://doi.org/10.1145/3404868.3406664

1 INTRODUCTION

Online services are increasingly distributed due to recent advances in content distribution, cloud computing and faster connectivity. Today, an online service is a collection of microservices that are hosted at diverse geographic locations that may be under different jurisdictions and are not contained within national borders. Examples include services such as Facebook, Whatsapp and PayPal, as well as infrastructures such as DNS, and authentication microservices that are critical to national services including public services, health and online banking. These intricate digital value chains breed unprecedented vulnerabilities. One such example is the 2017 NotPetya Ransomware attack that hit several major corporations causing $10 billion in damage [7]. Due to intricate transnational digital value chains, NotPetya also impacted several hospitals in the US. The attack, allegedly originated in Russia, targeted a popular Ukrainian accounting software. Furthermore, the attack spread to offices around the world denying service to a large number of doctors. The digital value chains global feature complicates the task of cyber risk assessment in terms of service localization, i.e., traffic for particular services may cross countries with comprehensive censorship or surveillance. We need not only to know where various services are hosted, but also to geolocate end-to-end Internet paths that are used for reaching these services.

IP geolocation is an active research area. Previous research focused on end hosts geolocation and largely ignored routers and peering points. Recent work shows that geolocation databases tend to perform rather poorly, even at the country level, when applied to Internet infrastructures [6]. In this paper, we conduct a controlled measurement study to investigate the accuracy of existing geolocation databases and services in geolocating end-to-end IPv4 and IPv6 paths at the country level. To this end, we conduct measurements between multi-homed end hosts in seven countries. The scale

of our experiment allows us to dissect generated IP geomappings in depth and understand possible causes of geolocation errors. Using this dataset, we evaluate the accuracy of both known geolocation databases and active measurements based approaches. We have also devised a simple approach for narrowing down the location of an IP by probing it from within the autonomous system (AS) that advertises it. We find that existing approaches tend to wrongly geolocate IPs from geographically spread (global) ASes as well as IPs that change ownership due to merger and acquisition. These mis-inferences can result in skipping countries that are actually on the path as well as mistakenly making claims about path tromboning and detours. Evaluating the geolocation approaches against a week long RIPE Atlas traceroute data confirms that our findings extend to the Internet at large.

2 APPROACH AND DATASETS

We use active measurements to collect IP level paths between end sites. We further employ five approaches for geolocating the measured IP hops to respective countries. We describe the measurement setup and the geolocation approaches.

2.1 Active measurement dataset

Measurement setup: Figure 1 shows the country location of the end-sites in our measurement setup. Most sites are multi-homed via an educational network and at least one commercial Internet Service Provider (ISP). A significant fraction of sites has IPv6 connectivity. The figure lists, in brackets, the number of IPv4 and IPv6 sites per country.

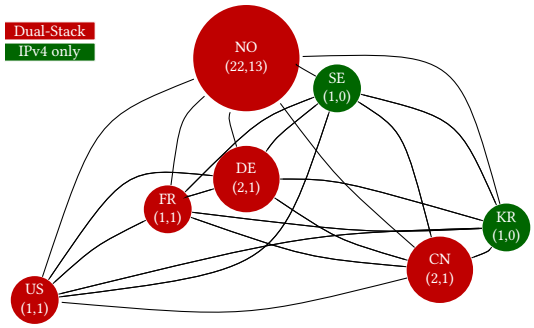

Figure 1: End-site country locations and paths between end-sites.

Data collection and pre-processing: To identify paths between end-sites, we regularly (ca. every 10-15 min) run traceroute between them from March to September 2018. To avoid load balancing effects during these traceroute runs, similar to Paris Traceroute [1], we fix the first 4 bytes of the IPv4/IPv6 payload – which are used by load balancers. We extract and analyze the traceroutes from the first seven days of every month. We collect on average 93,916 and 22,199 unique traces over IPv4 and IPv6, respectively. Since we control both ends when tracerouting, we measure both the forward and reverse paths between a pair of end-sites. Prior to analyzing the collected data, we remove duplicate paths, i.e., we keep only one copy for all paths that appear more than once. Hence, we reduce the number of paths to an overall average of 2752 and 837 over IPv4 and IPv6, respectively. This yields to approximately three IPv4/IPv6 paths per pair on average. In terms of ASes, we find an overall number of 73 ASes. Using CAIDA's AS classification [2] datasets we find that 70 of these ASes are "Transit/Access" networks.

2.2 IP geolocation datasets and methods

We describe the employed geolocation approaches [1].

Registration information: Regional Internet Registries (RIRs) publish files comprising their Internet resources assignments and allocations, i.e., IP addresses and autonomous system numbers (ASNs) [20]. For each IP address we extract the corresponding country and construct IP-to-country mappings which we further refer as *Delegation*.

Geolocation databases: We use the non-commercial versions of two popular databases: MaxMind GeoLite2 [17] and IP2Location BD11.Lite [13], which we refer to in the following as *MaxMind* and *IP2Location*, respectively. Our choice of the non-commercial versions is motivated by observations by Gharaibeh *et al.*, i.e., there are minor differences between the commercial and non-commercial versions in the country-level IP geolocations [6].

Geolocation approaches: IP geolocation methods employ active measurements from vantage points to narrow down the IP address location. We use two geolocation methods: IPmap [22] and HLOC [23], and further refer to the IP geomappings obtained from running these tool as *IPmap* and *HLOC*, respectively. IPmap mainly relies on RTT measurements from the RIPE Atlas probes [22]. It also uses other methods like crowdsourcing and anycast detection from RIPE Atlas anchors to improve its geolocation. HLOC first extracts geo-hints from DNS names, then selects a number of RIPE Atlas probes based on the extracted geo-hints and measures RTTs between them and the IPs it would like to geolocate.

3 COVERAGE AND DIFFERENCES

We compare the coverage of the five geolocation approaches above, and investigate to what extent they agree on associating an IP with a particular country.

3.1 Coverage

Figure 2 shows the percentage of IPs covered by each approach for May and September 2018. Note that our analysis shows no clear difference between IPv4 and IPv6 coverage as well as no differences between the two months. Hence, we use the measurements from May in the remainder of this

[1]We have attempted to use the data published by Gharaibeh *et al.* [6], but ultimately decided against that, as it only covers 1% of the collected IPs.

paper. Delegation and the commercial databases have a high coverage. IP2Location has a full coverage, while MaxMind and Delegation geolocate about 80% of the measured IPs. The partial coverage by the Delegation mappings is surprising, given that all IPs are expected to be tracked by RIRs. A closer look shows that all IPs are actually covered by the delegation files, however, 20% of them are not mapped to a particular country but to the EU. These IPs are mostly legacy IPs that were assigned by IANA prior to the inception of RIPE.

The geolocation methods have limited coverage. IPmap covers more IP addresses than HLOC and geolocates IPs across most of the ASes. Half of the unmapped IPv4 addresses are owned by three organizations – China Unicom Backbone, Korea Telecom and Telia. Similarly, half of the unmapped IPv6 addresses come also from three organizations – China Next Generation Internet, Internet2 and Telia. HLOC functionality is conditioned on the existence of DNS names with meaningful geo-hints for IP addresses. Some of the addresses with DNS names did not respond to ping from RIPE Atlas. HLOC rejects over half of the geolocations – indicated by the empty area in the bars – when the closest probe is over 1000km away for the inferred geolocation or when the RTT between the probe and hint location is greater than the expected delay over fiber optic plus 9ms to account for packet scheduling.

Digging deeper into the rejected IP geolocations we find the main root cause is that their DNS names involves several geo-hints and HLOC ends up choosing the geo-hint that is further away from the location of the IP. Having an understanding of naming conventions followed by different ASes can help avoiding these problems (e.g. the recently proposed approach by Luckie *et al.*[16]).

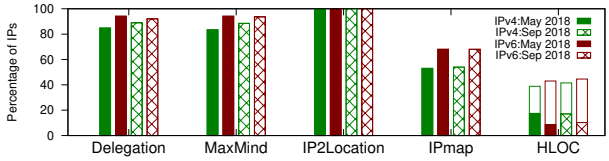

Figure 2: Geolocation approaches coverage.

3.2 Do geolocation approaches agree?

We examine to what extent the country-level IP geolocation approaches agree. We exclude from this analysis the IPs geolocated only by IP2Location. Most of the IPs covered by the three datasets are geolocated to the same country. We also observe both partial and complete IP location disagreements among the three datasets. For 5% and 8% of the IPv4 and IPv6 addresses, respectively, only two of the three datasets agree on the country location. Most of these agreements occur between Delegation and MaxMind. 10 IPv4 addresses are mapped to different countries by all the datasets.

We further investigate the similarity among the IP-to-country mappings in HLOC and IPmap data. We find partial agreement between the two datasets, i.e., 149 and 22 of the 268 IPv4 and 37 IPv6 addresses, respectively are placed in the same country by both approaches. 71 IPv4 and 15 IPv6 of these addresses have also similar mapping in the three geolocation datasets. These numbers indicate a low extended of agreements in the commonly seen IP addresses.

For the set of IP addresses mapped differently by HLOC and IPmap, we further seek to understand the possible causes. To this end, we select 20 IPv4 and 15 IPv6 owned by different organizations and investigate the HLOC country locations in terms of extracted geo-hints. Recall that HLOC's mapping approach relies first on valid geohints in the IPs DNS names and then on ICMP-based measurements from vantage points positioned most likely towards the edge of the networks. Our analysis shows the incorrect geo-hints are the main error cause. For 12 of the 20 IPv4 addresses the DNS name contains the name of the city, but HLOC selects as geo-hint another sequence of three characters. For example, the DNS name for TATA's IP addresses 116.0.82.89 is if-ae-7-2.tcore1.hk2-hong-kong.as6453.net. HLOC placed this IP in Vietnam based on the "hong" hint. The DNS name for the remaining 8 IPv4 addresses contains location code, but HLOC chooses a different sequence of characters and thus misplaces them.

4 IMPROVING GEOLOCATION ACCURACY

Having seen that the geolocation approaches map, at the country level, a sizable fraction of our IPs differently, we devise a simple method for geolocating IP addresses.

4.1 Geolocating infrastructures approach

Our approach is based on RTT measurements and hinges on a simple idea: a location of a router can be greatly narrowed down if we probe it from within its AS. First, we use existing methods (WHOIS services, DNS names and geolocation approaches) to find the owner and the possible physical location of the IP. Second, we use these insights to search for suitable vantage points (VPs) to traceroute to the target IP. A VP is judged suitable if it lies within the IP owner's AS and in close proximity to the initially guessed location. As VPs, we depend on publicly accessible looking glasses (LG). We choose LGs because they are often well provisioned and are not affected by last mile effects that probes from other platforms like RIPE Atlas maybe exposed to. Finally, we consider the IP in the same country as the LG if traceroute confirms a topological proximity (e.g. within a few router level hops and a latency of sub-20 ms) [2]. We proceed to select another LG, if the previous proved far away from the IP under test.

Is our approach feasible? The feasibility of our LG-based approaches is determined by the possibility of extracting

[2]We acknowledge that the proximity may differ across countries, and we plan to incorporate an adaptive threshold in our future work.

initial information on potential country and ownership of the IP(s) in question as well as the availability of LGs in the owner AS or close to it. To gain an initial impression about the feasibility, we evaluate 953 IPv4 and 346 IPv6 addresses on the two points above. These IPs come from 26 organizations. First, we check for availability of geo-hints and information on ownership. For 62.85% (IPv4) and 78.90% (IPv6) of these IP addresses we extract both the city location and owner name from the DNS name, while for the remaining we extract the country and owner from the WHOIS record. Second, we analyze how many organizations provide access to LGs within their network. 21 of the 26 organizations make available LGs [18]. In term of IPs, we find that only 27 IPv4 and 4 IPv6 addresses are mapped to the non-LG organizations. For these organizations, we further investigate whether we can still leverage the LG-based approach, if upstreams or peers of these networks have LGs. We find that three organization have in fact as an upstream Telia and one has a sibling with an LG. Moreover, for four IPv4 addresses we are able to apply our LG-based approach as these are ingress points between different organization. Note that there is no aggregated information on the availability of LGs within organizations. In the future, we plan to collect and make available such information. Note that LGs availability is a necessary condition but not sufficient as they may be located far away from the IP that we want to geolocate resulting in a poor inference.

Two case studies. We use our LG-based approach to geolocate 187 and 65 IPv4 addresses from Cogent (AS174) and NORDUnet (AS2603), respectively. We choose these two ASes because they are heavily represented in the set of IPs where all approaches disagree. Figure 3 shows the distribution of RTTs to the identified LGs for the Cogent and NORDUnet IP addresses. We manage to reach at least 85% of the selected LG within 2 ms. At the same time 98% of these LGs are reachable within 3 router hops. We notice that we reach about 10% of Cogent's IPv4 addresses from the respective LGs in more than 10ms. The number of hops, however, does not increase in a similar manner – we see at most one extra hop. For example, the highest latency (33ms) corresponds to an IPv4 addresses located in Hawaii that is discovered with an LG located in Los Angeles. In term of hops the LG is 3 hops away from the IP address. All the IPv4 addresses located in the tail of the Cogent distribution fall into this category and are in the US. Hence, this does not result in a wrong country level geolocation. IPs geolocated in European countries, we do not record such problems as organizations do not concentrate all their LGs in only one country.

Next, we compare LG-based method's country level geo-mappings to those obtained from the aforementioned approaches. We group the results into four categories depending on whether the geo-mapping places the IP address in the same country, neighbouring country, same region (i.e. continent) or a different region. Figure 4 shows the result of this comparison, which reveals that the geolocation databases mis-locate about half of the IPs. Delegation, MaxMind and IP2Location demonstrate a non-trivial disagreement with the LG-based method. IPmap correctly geolocates at least half of the considered IP addresses. HLOC, however, appears to misplace a significant percentage of the Cogent IPv4 addresses. Active probing based methods are vulnerable to failures in obtaining representative geo-hints from DNS names (HLOC) as well as the last-mile connectivity of the probe. We expect a higher error if the probes are located at the edge (e.g. increased latency because of poor edge performance, bufferbloat, etc). Indeed many RIPE Atlas probes are hosted at the edge. However, relying on LGs helps addressing this caveat.

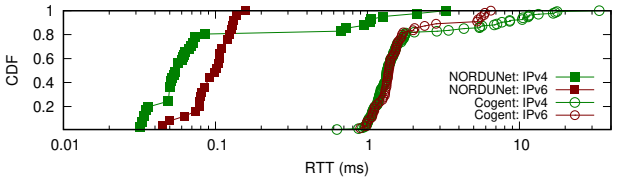

Figure 3: Distribution of RTT values towards the closest LGs for Cogent and NORDUnet IP addresses.

Sources of disagreement. We next dig deeper into understanding potential reasons behind the observed disagreements. Here, we focus on all IPs in our dataset where there is either full or partial disagreement.

Complete IP geolocation disagreement: 10 IPv4 addresses owned by Cogent (6), Level3 (3) and GTT (1) are mapped to a different country by each geolocation database. Delegation and IP2Location map Cogent IPs to the United States and Canada, respectively. MaxMind places four in France and two in Italy. IPmap geolocates only four: two in Sweden and two in Spain. To find the actual locations of these IPs, we pick the starting Cogent LGs [4] based on their DNS names and geo-hints in WHOIS records. For three of the five IPs that have DNS names we retrieve geo-hints that place two IPs in Norway and one in France. Using our LG-based approach, we geolocate these IPs to Norway, Germany and France. Hence, the DNS geo-hint for the German IP is wrong. We geolocate the remaining IPs to the three countries also. Thus, Cogent's IP addresses are inaccurately geolocated by all approaches. To understand the cause of these disagreement, we use the WHOSWAS service from ARIN to trackback the ownership of the six Cogent IPv4 addresses. [3] These IPs come from 149.6.0.0/16 which belonged to PSI Net, and moved to Cogent after part of PSI Net was acquired by Cogent in 2002 [3, 11]. Thus, the database error seems to be caused by inconsistent RIR due to M&A.

[3] https://www.arin.net/reference/research/whowas/

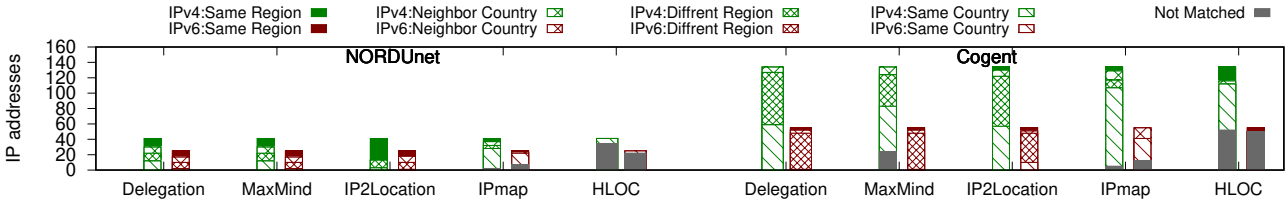

Figure 4: LG-based country location IP agreement among Delegation, MaxMind, IP2Location, IPmap and HLOC.

The GTT IP address is incorrectly geolocated by all three approaches. Using the GTT LG [8], we place this IP address in Great Britain which is also hinted by its DNS name. The three Level3 IP addresses lack DNS names. We use geo-hints in the WHOIS network name to select the initial LGs that suggest that two IPs are in Germany while one is Denmark. The same data places the Level3 European subsidiary in Great Britain. Our LG-based indicates the two IPs are in Germany, however the third turned out to be in Norway. MaxMind correctly maps the IP addresses in Germany and incorrectly maps the IP in Norway to Spain. Delegation places the Level3 IPs in Great Britain, while IP2Location maps the German IPs to United States and the Norwegian IP to Denmark. MaxMind and IP2Location appear to build the IP-to-country mappings based on the network name provided in the WHOIS record, while the Delegation mappings use the organization location. Accordingly, these errors are due to reliance on WHOIS records, which contain imprecise information on IP locations.

Partial IP geolocation disagreement: Apart from the complete IP geolocation disagreements, there are 87 IPv4 and 51 IPv6 addresses mapped to the same country by only two of the three geolocation datasets. These IPs are allocated to both educational network like NORDUnet and GEANT, but also to transit providers like GTT, TATA and Level3. 49 of the 87 IPv4 addresses belong to the NORDUnet core network and map to three IPv4 /24 address blocks (109.105.102.0/24, 109.105.97.0/24 and 109.105.98.0/24). WHOIS records place these IPs in Sweden. These IPs are geolocated in Great Britain by IP2Location and in Sweden by both Delegation and MaxMind. Thus, we hypothesize that MaxMind uses WHOIS to build their IP-to-country mapping for the NORDUnet IPv4 addresses. Our approach maps these IPs across seven countries [4]. Similarly, the NORDUnet IPv6 addresses are inaccurately mapped.

This analysis indicates yields the following conclusions. First, dedicated geolocation databases like MaxMind and IP2Location appear to use information from the WHOIS records (country, network name) to build their IP-to-country mappings. Second, IPs owned by organization with international presence are often geolocated wrongly. Third, merger and acquisition of organizations is a key source of IP geolocation inaccuracies.

[4] 16 in SE, 10 in US, 7 in DK, 6 in NL, 5 in NO, 3 in GB and 2 in DE.

5 ERRONEOUS GEO-MAPPINGS NEGATIVE IMPACT

IP geolocation disagreement have implications for applications that derive conclusions based on which countries are on the path. For example, a number of recent studies have depended on IP geolocation to quantify the extent of path tromboning and routing from a sovereign state point of view [5, 24]. We explore how erroneous mappings impact inference of path tromboning and detours as well as on having an accurate view on which countries are on the path.

5.1 Path tromboning and detours

We check whether our mapped IP paths exhibit tromboning or continental detours. By tromboning, we refer to the case where the source and destination are in the same country, yet there is an intermediate foreign hop(s). By continental detours, we refer to paths connecting points in the same continent, yet traverses another continent.

Path tromboning. We first investigate whether our collected IP paths experience tromboning. Recall that a high percentage of our end-sites are spread across Norway. Hence, approx. 30% IPv4 and 26% IPv6 paths start and end in Norway. Additionally, we also extract two IPv4 paths that start and end in Germany. We do not find any tromboning in IPv4 paths. For the IPv6 path we find false positives caused by inaccurate IP-to-country mappings in the BKK Digitek AS core network resulting in tromboning 33 IPv6 paths. We find the accurate location using the proposed LG-based approach.

Path detours. To investigate detours we consider 519 IPv4 and 254 IPv6 paths that start and end within Europe and check them for IPs geolocated in other continents. 40% and 51% of the considered IPv4 and IPv6 paths are detoured though the United States in at least one geolocation database. IPv4 detour paths appear due to incorrect geo-mappings for 96 IPv4 addresses owned by Level3, Cogent and Telia. Incorrect geo-mappings for 64 IPv6 addresses owned by Level3 and Cogent are the culprit for detours on IPv6 paths. The LG-based method successfully avoids these erroneous mappings.

5.2 How many countries are missed?

To check whether wrong mappings results in leaving out a country or more altogether from the path, we crosscheck for each path the unique countries on the country-level paths.

Surprisingly, a high percentage of the paths appear to miss a country or more when mapped with different geolocation datasets. 77% IPv4 and 65% IPv6 country-level mappings miss at least one country along the path. IP paths within the same continent (short haul paths) and between different continents appear to miss countries (long haul paths).

Long haul paths. Approximately half of the selected IP paths that appear to miss a country are between Europe and China. For example, an IPv4 path from China to Norway is geo-located as follows. Delegation mappings indicate that the path passed through China, United States and ends in Norway. MaxMind maps the path to China, United States, France and Norway. IP2Location locates the path to China, United States, Canada and Norway. In this case we mark Canada and France as missing countries. However, using our LG-based approach we map this path to China, United States, Canada, Netherlands, Germany, Sweden and Norway. Between Europe and China, we find an overall of 12 and 8 countries missing on IPv4 and IPv6 paths, respectively. Moreover, IP paths between Europe and United States account for approx. 10% and 12% of the affected IPv4 and IPv6 paths, respectively. These paths miss an overall of 10 countries.

Short haul paths. Geomapping of IP paths with Europe appear to miss 8 countries for both IPv4 and IPv6 paths. In the case of IPv6 paths, these countries are located only in Europe. For IPv4 paths we find Canada and United State as missing countries. However, these two countries are false positives (see the the continental path detours analysis).

6 LARGE SCALE ANALYSIS

Analyzing our small-scale dataset helped us identifying several weaknesses that plague existing approaches to geolocations. However, it is not clear whether these findings apply to the Internet in general. To explore this question, we leverage successful public RIPE Atlas traceroutes collected during the first seven days of April 2019. [21]. Overall, we have 10,715,357 IPv4 and 7,094,870 IPv6 unique traces. We remove traces that comprise unresponsive hops and obtain 561,199 IPv4 and 101,378 IPv6 unique addresses. We geolocate these IPs in the Delegation, MaxMind and IP2Location databases.

More than 90% of the collected IP addresses are mapped to a country across the three databases spanning across 241 and 176 countries for IPv4 and IPv6 addresses, respectively. When investigating the IP country location agreement we find that more than 85% of the IP addresses are geolocated to the same country. Consequently, we observe a small percentage of IP addresses with partial or complete geo-mapping disagreements. Such IPs, however appear to affect a high percentage of IP-paths. Difference in the IP-to-country mapping generate, for at least 65% of the collected paths, different country-level end-to-end mapping. Our analysis also indicates that, for 42% IPv4 and 32% IPv6 paths, the country-level mappings miss at least one country in one of the databases. These findings are aligned with observations extracted from the small-scale dataset. Moreover, we note that the percentage of geomapping disagreements is comparable among the two datasets.

7 RELATED WORK

Several previous work has focused on devising methods for improving IP geolocation. These include delay-based methods (e.g. [10]). Others have attempted to improve the results of RTT based methods via incorporating additional constraints and data sources like topology and DNS names (e.g. [14], [27], [12, 26] and [23]). We provide a fresh way to geolocating infrastructures that leverages looking glasses as a probing source.

Several studies have also focused on evaluating existing geolocation datasets [6, 9, 19, 25]. Recent work published by Gharaibeh *et al.* [6] focuses on evaluating the consistency and coverage of 1.6M IPv4 addresses assigned to router interface location in four geolocation datasets. We take one step further by exploring different geolocation disagreement sources, including IPv6 and investigating the impact of erroneous geo-mappings on inferred country level paths.

8 CONCLUSIONS

In this work, we investigate the reliability of existing geolocation approaches in providing country level geolocations of infrastructure IPs. We collect IP router-level paths between a number of sites that are spread globally. Our analysis shows a non-trivial disagreement between various approaches and highlights various issues that can reduce the accuracy of DNS and active probing based methods. We find that geolocation databases tend to erroneously geolocate IPs that belong to ASes with global presence and IPs that change ownership due to merger and acquisition. Unfortunately, this is only going to increase given the increasing number of IPv4 transfers as a response to IPv4 depletion [15]. Such disagreements can falsely indicate path tromboning or path detours. Also, these databases appear to miss or add countries to end-to-end paths. This observation has great security implications as it indicates that, depending on popular geolocation databases, end-hosts might be unaware of the countries their Internet traffic is traversing. We devised a simple method for narrowing down the location of IPs that is based on probing these IPs from within the ASes that advertise them. Our method has yielded promising results, which we plan to develop further.

9 ACKNOWLEDGMENTS

This research was supported by Norwegian Research Council grant # 288744 GAIA.

REFERENCES

[1] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Clémence Magnien, and Renata Teixeira. 2006. Avoiding Traceroute Anomalies with Paris Traceroute. In *Proc. of ACM IMC*.

[2] CAIDA. [n. d.]. AS classification dataset. ([n. d.]). https://www.caida.org/data/as-classification/.

[3] Cogent. 2019. Cogent Communications Acquires U.S. Operations of PSINET. (2019). http://www.cogentco.com/en/news/press-releases/279-cogent-communications-acquires-us-operations-of-psinet.

[4] Cogent: Looking Glass. 2019. (2019). http://www.cogentco.com/en/network/looking-glass.

[5] Anne Edmundson, Roya Ensafi, Nick Feamster, and Jennifer Rexford. 2016. Characterizing and Avoiding Routing Detours Through Surveillance States. (2016). arXiv:cs.NI/1605.07685

[6] 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 *Proc. of ACM IMC*.

[7] Andy Greenberg. 2018. The untold story of NotPetya, the most devastating cyberattack in history. *Wired, August* 22 (2018).

[8] GTT: Looking Glass. 2019. (2019). www.as3257.net/lg.

[9] Bamba Gueye, Steve Uhlig, and Serge Fdida. 2007. Investigating the Imprecision of IP Block-based Geolocation. In *Proc of PAM (PAM'07)*.

[10] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. 2004. Constraint-Based Geolocation of Internet Hosts. *IEEE/ACM ToN* 14, 6, 1219–1232.

[11] Company Histories. 2019. Cogent Communications Group, Inc. (2019). http://www.company-histories.com/Cogent-Communications-Group-Inc-Company-History.html.

[12] Bradley Huffaker, Marina Fomenkov, and kc claffy. 2014. DRoP:DNS-based Router Positioning. *ACM SIGCOMM CCR* 44, 3, 6–13.

[13] IP2Location.com. 2018. IP2Location BD11.LITE (accessed September 2018). (2018). https://lite.ip2location.com/.

[14] 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 *Proc. of ACM IMC*.

[15] Ioana Livadariu, Ahmed Elmokashfi, and Amogh Dhamdhere. 2017. On IPv4 Transfers Markets: Analyzing Reported Transfers and Inferring Transfers In The Wild. *Elsevier COMCOM* (Oct 2017).

[16] Matthew Luckie, Bradley Huffaker, and k claffy. 2019. Learning Regexes to Extract Router Names from Hostnames. In *Proc. of ACM IMC*.

[17] MaxMind. 2018. Maxmind GeoLite2 databases. (accessed September 2018). (2018). https://dev.maxmind.com/geoip/geoip2/geolite2/.

[18] Peering DB. [n. d.]. ([n. d.]). https://www.peeringdb.com/.

[19] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP Geolocation Databases: Unreliable?. In *ACM CCR*.

[20] RIPE NCC. [n. d.]. Allocation and assignment resource file. ([n. d.]). ftp://ftp.ripe.net/pub/stats.

[21] RIPE NCC. [n. d.]. Data Sets. ([n. d.]). https://labs.ripe.net/datarepository/data-sets.

[22] RIPE NCC. 2019. RIPE IPmap. (2019). https://ipmap.ripe.net/.

[23] Quirin Scheitle, Oliver Gasser, Patrick Sattler, and Georg Carle. 2017. HLOC: Hints-based geolocation leveraging multiple measurement frameworks. In *Proc. of TMA*.

[24] Anant Shah and Christos Papadopoulos. 2015. Characterizing International BGP Detours. In *Technical Report CS-15-104 (Colorado State University)*.

[25] Y. Shavitt and N. Zilberman. 2011. A Geolocation Databases Study. *IEEE JSAC* 29, 10, 2044–2056.

[26] Neil Spring, Ratul Mahajan, and David Wetherall. 2002. Measuring ISP Topologies with Rocketfuel. In *Proc of ACM SIGCOMM*.

[27] Yong Wang, Daniel Burgener, Marcel Flores, Aleksandar Kuzmanovic, and Cheng Huang. 2011. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. In *Proc. of USENIX NSDI (NSDI'11)*.