A Look at Router Geolocation in Public and Commercial Databases

Manaf Gharaibeh
Colorado State University

Anant Shah
Colorado State University

Bradley Huffaker
CAIDA / UC San Diego

Han Zhang
Colorado State University

Roya Ensafi
University of Michigan

Christos Papadopoulos
Colorado State University

ABSTRACT

Internet measurement research frequently needs to map infrastructure components, such as routers, to their physical locations. Although public and commercial geolocation services are often used for this purpose, their accuracy when applied to network infrastructure has not been sufficiently assessed. Prior work focused on evaluating the overall accuracy of geolocation databases, which is dominated by their performance on end-user IP addresses. In this work, we evaluate the reliability of *router* geolocation in databases. We use a dataset of about 1.64M router interface IP addresses extracted from the CAIDA Ark dataset to examine the country- and city-level coverage and consistency of popular public and commercial geolocation databases. We also create and provide a ground-truth dataset of 16,586 router interface IP addresses and their city-level locations, and use it to evaluate the databases' accuracy with a regional breakdown analysis. Our results show that the databases are not reliable for geolocating routers and that there is room to improve their country- and city-level accuracy. Based on our results, we present a set of recommendations to researchers concerning the use of geolocation databases to geolocate routers.

CCS CONCEPTS

• **Networks** → **Network measurement**; *Routers*;

KEYWORDS

IP Geolocation; Geolocation Databases; Router Geolocation

ACM Reference Format:
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 IMC '17, London, United Kingdom, November 1–3, 2017,* 7 pages.
https://doi.org/10.1145/3131365.3131380

1 INTRODUCTION

IP geolocation services map IP addresses to physical locations such as a country, city, or geographic coordinates. Many commercial entities use these services to customize content delivery and advertisements for their users. Networking research uses geolocation services to study the geographic deployment of Internet resources and their utilization. Studying and visualizing routing phenomena to detect BGP threats [31], estimating the geographic presence of Autonomous Systems and detecting routing paths that experience a detour-paths that start and end in the same country but visit other countries in between [28], and studying censorship and monitoring that happens in different countries [25] are just a few examples.

All these studies rely heavily on the accuracy of geolocation services especially for IP addresses that are used for Internet infrastructure (e.g., routers, switches). Quantifying the error margins and identifying regions where geolocation services fail can substantially improve the quality of such studies. Geolocation services are typically available as third-party databases, publicly available [6, 19, 20] or paid [8, 17, 21]. Delay-based geolocation, where delay measurements are mapped to location constraints [14, 22, 24, 32, 33], is another viable option, especially with more public measurement platforms becoming available [7, 11, 23]. However, many users might still prefer the available ready to use geolocation databases.

Previous work on evaluating databases focused on their overall accuracy [13, 15, 26, 29, 30]. However, such work is biased towards evaluating endpoints geolocation since there are far more endpoints than infrastructure in the Internet. Given the importance of router geolocation in understanding geographic aspects of the Internet infrastructure, our work focuses on router geolocation in both public and commercial databases.

Researchers who use geolocation databases to learn the locations of routers need to know how reliable they are in terms of their country- and city-level coverage (i.e., the fraction of addresses a database has country- and city-level resolutions for, respectively), and their accuracy throughout the world. In this work, we study four popular geolocation databases, two of which are free: MaxMind GeoLite2 [19], and IP2Location DB11.Lite [20], and two are commercial: MaxMind GeoIP2 [17], and Digital Envoy NetAcuity [8]. We explain why these databases are selected in §2.2.

Our main contributions in this paper are: (1) we show that the studied databases have many inconsistencies, especially at city-level. We use a set of 1.64M router interface addresses extracted from CAIDA's Ark dataset (§2.1) to study all 4 databases inconsistencies and coverage; (2) we create a ground truth dataset[1] of 16,586 router interface addresses and their locations with city-level accuracy. We create our ground truth using two approaches, a DNS-based approach proposed by Huffaker *et al.* [16] and a delay-based

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IMC '17, November 1–3, 2017, London, United Kingdom
© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
ACM ISBN 978-1-4503-5118-8/17/11…$15.00
https://doi.org/10.1145/3131365.3131380

[1] Our ground truth data is available via IMPACT: https://www.impactcybertrust.org/dataset_view?idDataset=792

approach that utilizes the RIPE Atlas built-in measurements [23]; (3) we use the ground truth dataset to evaluate the databases' country- and city-level accuracy regionally. The results show that all the databases have room to improve their accuracy, even at country-level; (4) our final contribution is a set of recommendations for using the geolocation databases to geolocate routers.

2 DATASETS

2.1 CAIDA Topology Dataset

We use the CAIDA topology dataset [2] collected using CAIDA's Ark measurement infrastructure. Ark monitors around the world collect traceroute data for randomly selected IP addresses from all routed /24 IPv4 prefixes. Using one week of the topology dataset starting from March 9, 2016, we extract a dataset of 1,638K router interface IP addresses, which map to an estimated number of 485K distinct routers according to CAIDA's ITDK alias mapping results [3]. We treat this dataset at IP-level since the geolocation services are supposed to geolocate all IP addresses regardless of their alias resolution. We refer to this dataset as the *Ark-topo-router* dataset. We use this dataset to evaluate the country- and city-level coverage and consistency across the geolocation databases.

2.2 Geolocation Databases

We compare and assess router geolocation reliability in four popular geolocation databases: MaxMind GeoIP2 (referred to as *MaxMind-Paid* in this paper), MaxMind GeoLite2 (referred to as *MaxMind-GeoLite*), IP2Location DB11-Lite (referred to as *IP2Location-Lite*), and Digital Element NetAcuity (referred to as *NetAcuity*). We chose the NetAcuity and MaxMind commercial databases as they are widely considered among the leaders in the geolocation business [15, 29]. On the other hand, comparing the free and commercial versions of MaxMind's databases provides a measure of the improvement between the two. Finally, the IP2Location database is known for providing city-level resolution for most of the IP address space and it appears often in geolocation comparative studies.

2.3 Ground Truth Data

Our ground truth data is basically a set of router interface addresses and their locations at city-level accuracy. We extract our ground truth using two methods as explained next. We discuss the correctness of the ground truth data in §3.

2.3.1 DNS-Based Ground Truth Data. Huffaker et al. [16] geolocate routers by decoding location hints in their hostnames. They inferred an extensive dictionary that maps location strings such as airport codes to physical coordinates, then using domain-specific rules they search for and decode location hints in hostnames to infer their locations. They generated domain-specific rules for 1,398 domains but we only use 7 domains for which they have ground truth rules from the domains' operators. Performing reverse DNS (rDNS) lookups to the Ark-topo-router addresses on May 15, 2016, results in 905K addresses with hostnames, about 13.5K of which belong to the ground truth domains in [16], from which we are able to geolocate 11,857 addresses from the 7 domains: *belwue.de* (23 addresses), *cogentco.com* (6,462), *digitalwest.net* (29), *ntt.net* (2,331), *peak10.net* (170), *seabone.net* (1,405), and *pnap.net* (1,437).

2.3.2 RTT-Proximity Ground Truth Data. We use RIPE Atlas *built-in measurements* [23] to create our second part of the ground truth. These measurements are issued by most of the probes toward well known targets like DNS root servers. We use traceroute measurements collected on May 25th, 2016. The measurements are provided in *JSON* format that specify the measurement origin, target, intermediate hops and their observed RTTs. Since a 0.5ms RTT between two locations maps to a distance of at most 50 km—likely much less due to inflation in RTT measurement—, we use 0.5ms threshold to find all the hops guaranteed to be within 50 km of their probes. We associate such hops with their probes' locations. We find 4960 router interface IP addresses that satisfy our 0.5ms threshold but we only keep 4838 addresses due to the reasons we explain in §3.2. We refer to the set of 4838 IP addresses and their locations as the *RTT-proximity* ground truth dataset. Note that while some of the gathered IP addresses could belong to home routers, more than 80% are at least 2 hops away from their probes indicating otherwise.

2.3.3 Ground Truth Data Regional And Topological Distribution. Table 1 shows statistics for our two ground truth datasets including the total number of addresses (column 2), number of unique countries where the addresses are located (column 3), number of unique coordinates (column 4), and the number of addresses found in each regional Internet registry (RIR) (columns 5 to 9). The RIR for each address is learned from querying Team Cymru whois database [5]. According to CAIDA'S AS rank [1], transit ASes announce 74.5% of addresses in our RTT-inferred ground truth set and 99.9% of addresses in our DNS-based ground truth set.

3 GROUND TRUTH DATA CORRECTNESS

3.1 DNS-Based Data Correctness

We validate part of the DNS-based dataset using two latency measurement datasets including our RTT-proximity dataset, and another similar dataset provided to us by Giotsas *et al.* [12]. Despite the small intersection between the datasets, we see very positive signs for correctness as we explain next.

We identify 109 common addresses between the DNS-based and the RTT-proximity ground truth datasets. The datasets agree within 10 km on the locations of 105 of the addresses and within 43 km on the remaining 4 addresses.

Giotsas *et al.* router geolocation dataset [12] was gathered about 10 months after our DNS-based dataset. Giotsas *et al.* looked for RIPE Atlas probes within 1ms from a set of routers of interest, hereby referred to as 1ms-RTT-proximity dataset. The 1ms-RTT-proximity dataset has about 20.5K router interfaces, but only 384 are common with the DNS-based dataset. Pairwise comparisons show that for 355 addresses (92.45%), the locations from the two datasets are less than 100 km apart. Given the 1ms threshold used to create the 1ms-RTT-proximity dataset, these locations are fairly compatible; in fact, the locations of 337 addresses (87.8%) are less than 40 km apart.

Interestingly, out of the remaining 29 IP addresses with incompatible locations, we find that 19 addresses are likely reassigned to hosts at different locations. We observe this change in their rDNS records. For example, the rDNS lookup result for one IP

Table 1: Location statistics and regional distribution of the DNS-based and RTT-proximity router interface addresses.

Ground Truth	Total	Countries	lat/lon	ARIN	APNIC	AFRINIC	LACNIC	RIPENCC
DNS-based	11,857	53	238	9,588	560	0	0	1,709
RTT-proximity	4,838	118	1,347	1,123	372	131	52	3,160

address was *ae-5.r23.dllstx09.us.bb.gin.ntt.net* on May 2016 and *ae-3.a01.miamfl02.us.bb.gin.ntt.net* on September 2017. The location hint in the prior one indicates the location *Dallas, TX*, while the later indicates *Miami, FL*. All 19 addresses would have similar geolocation to that in the 1ms-RTT-proximity dataset given their updated hostnames, that said, we do not know when exactly the hostnames have changed and if that happened before creating the 1ms-RTT-proximity set. The location disagreement for some of the remaining 10 addresses might be a result of reassigning addresses to hosts at a different location without updating their hostnames leading to misleading location hints. Few RIPE Atlas probes may also have incorrect geolocation.

Overall, between May 2016 and September 2017, 8,197 (69.1%) of the 11,857 DNS-based addresses kept the same hostnames, 2,848 (24%) have different hostnames, and 6.9% no longer have rDNS records. Not all hostnames changes indicate location changes. Geolocating the 2,848 addresses with different hostnames using DRoP's domain-specific ground truth rules shows that 1,927 (67.7%) still have the same location, 877 (30.8%) have different location—i.e., 7.4% of all DNS-based addresses in about 16 months—, and 44 (1.5%) no longer have location hints that match any of the rules.

3.2 RTT-Proximity Data Correctness

The correctness of the RTT-proximity data is dependent on the accuracy of the RIPE Atlas probes locations, which are mostly crowdsourcing-based. While the probes' hosts can easily provide correct city-level locations, it is not guaranteed that they always do. Additionally, a probe might be moved without updating its public location. RIPE Atlas operators informed us that they do some manual checking but nothing structural to validate probes' locations. To increase the confidence in the RTT-proximity data we use two methods to disqualify probes that appear to have inaccurate geolocation.

First, we identify and remove all probes assigned *default* country coordinates. These coordinates are typically near the geographic center of a country [4, 9, 18] and are often located in unpopulated areas (e.g., *N*51°00′00″ *E*09°00′00″ in Germany). Such coordinates are often assigned to IP addresses due to the lack of specific location information. From the set of 1,387 probes associated with our 0.5ms threshold data, we find 19 probes within 5 km of their known country coordinates. Using traceroute measurements we are able to prove that many of these probes indeed have bad geolocation. We find and remove 109 IP addresses associated with these probes.

Our second method is based on the insight that multiple probes near the same router should also be near each other. We call such a group of probes RTT-*nearby* probes. Given our 0.5ms threshold, any two RTT-*nearby* probes should be within a distance of 100 km. We find 495 addresses in the remaining RTT-proximity data with RTT-*nearby* groups of 2 or more probes, out of which, only 12 addresses (2.4%) have RTT-*nearby* probes with inconsistent locations. For example, two probes in Mozambique are RTT-*nearby* to an IP address but their locations are 867 km apart, which means at least one of them has incorrect geolocation. We find 3 other RTT-*nearby* groups that have prominent location inconsistencies. The 8 remaining addresses have relatively small disagreements of less than 128 km between any two probes in one RTT-*nearby* group. One probe in Italy is responsible for 7 of those location disagreements. Overall, we have 223 different probes that are part of one or more RTT-*nearby* groups, out of which, we only disqualify 5 probes (2.2%) and remove 13 interface addresses associated with them. As a result, the final RTT-proximity dataset has 4,838 addresses.

We match the RTT-proximity and the 1ms-RTT-proximity datasets and find an intersection of 1,661 addresses. Comparing the locations from the two datasets for each common IP address shows that 96.8% and 97.4% of the addresses agree within 40 km and 100 km respectively. The small fraction with location disagreements might be a result of IP addresses reassignment to hosts at different locations during the time separating the two datasets.

4 METHODOLOGY

In this paper we seek to answer these questions: (a) what is the probability to find an answer for a router address geolocation query and what would be the resolution of the answer? (b) how consistent are the answers across different databases at both country- and city-level resolution? (c) what is the probability that the database answer is correct? We next explain how we answer these questions.

To evaluate the coverage and consistency of the participating databases, we use the Ark-topo-router dataset (§2.1). To evaluate a database coverage we find the percentage of addresses with location information in each database for both country- and city-level. We also evaluate the pairwise consistency at both resolutions.

While country-level consistency evaluation is as simple as comparing standard ISO alpha-2 or alpha-3 country codes in databases, the city-level consistency evaluation can be tricky, in part because different databases may use different city names. Rather than comparing city names, we compute the distance between one IP address locations (coordinates) in any two databases and check if it is within city range. Comparing coordinates invokes two questions: (a) does the database provide correct city-level coordinates for a given city in a location record? (2) what radius is acceptable as a city range?.

We compare each database coordinates for a given city with the city coordinates from a third party geographical database called *GeoNames* [9]. Since multiple cities can have the same name, we also include the region and country in the matching process. We observe that the distance between city coordinates from any of the geolocation databases and GeoNames is within 40 km more than 99% of the time, indicating that the databases are indeed assigning city-level coordinates when a city name exists in the location record.

Answering the question about city range is tricky, mainly because different cities can have drastically different areas. Previous work [29] used 40 km as their city range, while [15, 16] used the same distance as their threshold to identify if two locations are co-located. However, we note that different databases may assign different coordinates to the same city. We examine the distance

between coordinates assigned to the same city across the databases, and find that one city coordinates from any two databases are more than 99% of the time within 40 km. We conclude it is reasonable to consider any two databases' coordinates within 40 km to be within the same city circumference.

Finally, we evaluate the overall databases router geolocation accuracy using our ground truth of 16,586 interface addresses with city-level accuracy. We also evaluate the accuracy by region where we breakdown the ground truth addresses by their RIRs and report the results for each region at country- and city-level (§5.2.2).

5 RESULTS

5.1 Databases Coverage and Consistency

Using the Ark-topo-router dataset, we analyze databases router geolocation coverage and consistency at country- and city-level. All the databases are accessed shortly after creating the Ark-topo-router set to geolocate its addresses. We find that IP2Location-Lite and NetAcuity both provide near perfect coverage for all interface addresses in the Ark-topo-router dataset at both country- and city-level. The MaxMind-GeoLite and MaxMind-Paid databases both cover about 99.3% of the addresses at country-level, but only 43% and 61.6% of the addresses at the city-level respectively.

Pairwise country-level comparison shows that the MaxMind databases agree on the location of 99.6% of the 1.64M interface IP addresses, while all other pairwise comparisons' agreements range between 97.0% and 97.6%. The overall country-level agreement between all databases is about 95.8% (1.57M addresses). The agreement between the databases might suggest more confidence in the geolocation results, it might also indicate a common incorrect source of the geolocation information (e.g., *registry data*).

We now turn to city-level resolution comparisons. Figure 1 shows pairwise comparison of databases locations (i.e., coordinates) for the Ark-topo-router addresses. For each pair of databases, we compute the distance between the locations from the two databases for each IP address. We then plot the distance distribution for all the addresses. Only the addresses with city-level and *(latitude, longitude)* coordinates in all databases are included (i.e., around 692K IP addresses). The pairwise comparison of the two MaxMind databases shows mostly small differences. 470K addresses (68%) have identical coordinates in the two databases and are truncated from their pairwise distance CDF. But for 11.4% of the addresses, the distance is more than 40 km, indicating that the IP address is likely geolocated to different cities. Other pairwise comparisons show more discrepancies where more than 29% of the addresses are geolocated by different databases to locations more than 40 km apart. The CDFs for IP2Location-Lite and NetAcuity vs. MaxMind-GeoLite are omitted since they are similar to those vs. MaxMind-Paid.

5.2 Evaluation Using Ground Truth Data

Using our ground truth of 16,586 interface addresses (see §2.3), we evaluate the coverage and accuracy of all databases at country- and city-level. The databases are accessed again on early July 2016, to geolocate the ground truth (i.e., about 50 days after creating the DNS-based set). We observe that 7.4% of our DNS-based set addresses likely moved during a 16 months period (see §3.1). The

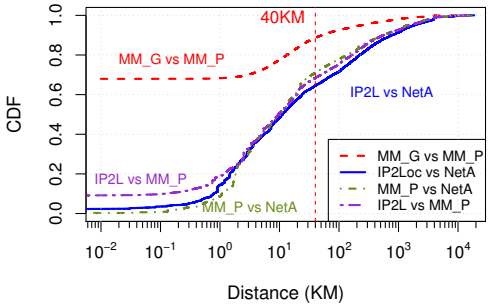


Figure 1: Databases pairwise distance distributions show at least 29% city-level disagreements for different vendors.

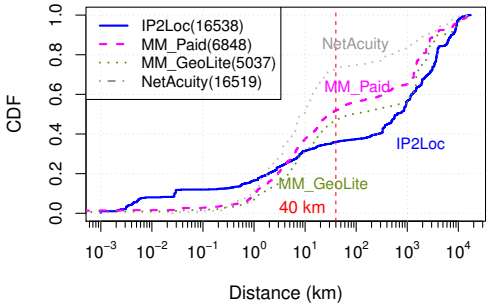


Figure 2: Databases vs. ground truth geolocation error. The number of addresses in each CDF is enclosed in parenthesis.

movement is likely much less in 50 days (i.e., one-tenth of the 16 months) and is unlikely to affect our conclusions.

5.2.1 Databases Coverage and Accuracy Over the Ground Truth. IP2Location-Lite and NetAcuity show near perfect country- and city-level coverage for the addresses in the ground truth. MaxMind-GeoLite and MaxMind-Paid have around 95.4% country-level coverage, and only 30.4% and 41.3% city-level coverage respectively.

The country-level geolocation accuracy is usually reported at higher than 97% by the geolocation service providers [29] (e.g., MaxMind GeoIP2 reports 99.8% accuracy [18]). However, our results over ground truth data show less accuracy for router geolocation. NetAcuity outperforms the other databases at only 89.4% accuracy while IP2Location-Lite and MaxMind databases are comparable with 77.5% to 78.6% accuracy. We discuss country-level accuracy in more depth when we break down results by RIR next in §5.2.2.

Figure 2 shows the distribution of the geolocation error for each database vs. the ground truth for the addresses with city-level geolocation. The vertical red line (at x = 40 km) is our city range threshold. NetAcuity has clearly better accuracy compared to other databases but still incorrectly geolocates some interfaces hundreds of kilometers away from their actual locations. IP2Location-Lite is the least accurate but has much better city-level coverage compared to both MaxMind databases.

5.2.2 Regional Evaluation. To study the accuracy of the databases regionally, we break down the ground truth addresses by their RIRs. Figure 3 shows the country-level accuracy by region. Each column in the graph shows the number of correctly and incorrectly geolocated addresses for each database. The percentage over each column shows the fraction of incorrectly geolocated addresses. From the

graph we see that NetAcuity is the most accurate in all regions but there is still room to improve. We also observe that IP2Location-Lite and the two MaxMind databases' country-level accuracy results are comparable in all regions except for APNIC.

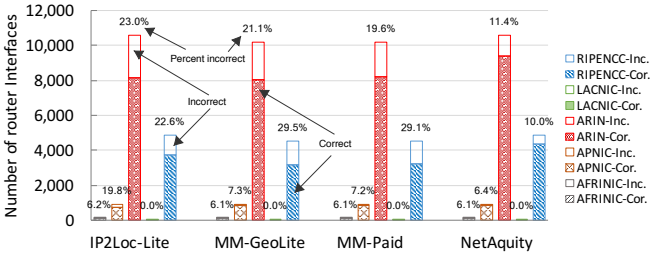


Figure 3: Country-level accuracy breakdown by RIR for ground truth. NetAcuity is the most accurate in all regions.

We go one step further and compare country-level accuracy for individual countries. Figure 4 shows the fraction of addresses correctly geolocated for the 20 countries with most addresses in the ground truth (country code and the number of addresses are depicted on the *x*-axis). While all databases show better than 94% accuracy for addresses in the United States (US) and Russia, their accuracy in most other countries is relatively low, especially the IP2Location-Lite and the two MaxMind databases, which show surprisingly low accuracy in western countries like France and the Netherlands. IP2Location-Light, MaxMind-GeoLite, and MaxMind-Paid agree on the (incorrect) location of 2,277 addresses, which corresponds to around 61%, 64%, and 67% of their incorrectly geolocated addresses respectively. NetAcuity shows the most reliable results with at least 74% country-level accuracy in all 20 countries.

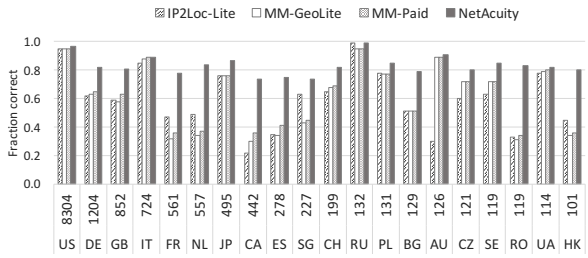


Figure 4: Databases' country-level accuracy is unreliable in most countries but NetAcuity is relatively consistent.

Finally, we evaluate city-level accuracy by region. Figures (5a, 5b) respectively show the distribution of geolocation error with breakdown by RIR for the MaxMind-Paid and NetAcuity against the ground truth data.[2] IP2Location-Lite has almost perfect city-level coverage but the accuracy is lacking, especially for ARIN addresses. Apart from ARIN, MaxMind seems to provide city-level geolocation only when it has some confidence in it, which could explain their low city-level coverage and relatively good city-level accuracy. For example, MaxMind-Paid city-level accuracy for the RIPENCC addresses is 78.9% with only 31.3% coverage compared to only 70.9% country-level accuracy and 93.3% coverage. NetAcuity, again, shows consistent coverage and accuracy results, but like the other databases, it is less reliable for the ARIN addresses.

[2]IP2Location-Lite and MaxMind-GeoLite graphs are omitted for space.

5.2.3 Poor City-level Accuracy at ARIN. The worst city-level accuracy for all the databases is observed for ARIN addresses. We use MaxMind-Paid as a case study to understand reasons for such poor accuracy. ARIN has 10,608 addresses (64%) of the ground truth. 2,793 of those addresses are not located in the US—according to the ground truth data. However, MaxMind-Paid, possibly relying on registry data, geolocates 1,955 of them (70%) to the US. We find that 519 (26.6% of the 1,955 addresses) have city-level geolocation in MaxMind, most of them (504 addresses) have disagreements greater than 1,000 km with the ground truth locations.

Total ground truth addresses located in the US is 8,304 (7,815 from ARIN and 489 from other RIRs). Total ARIN addresses located in the US with city-level information is 3,897, of which 2,267 (58.2%) have geolocation error > 40 km—our city-range. About 91% of them have block-level—/24 block or larger—locations compared to about 78% of the correctly geolocated addresses at city-level. Block-level location assignments can be responsible for large geolocation errors for interface addresses not co-located with the other addresses in their block. We do not investigate blocks co-locality in this work.

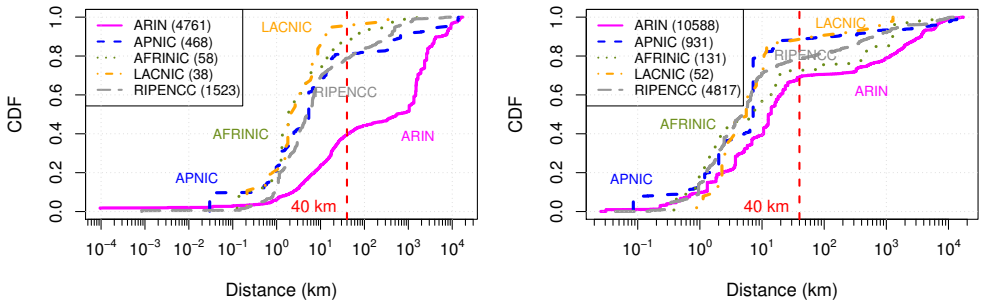
5.2.4 Databases vs. Separate Ground Truth Datasets. We evaluate the databases city-level accuracy against the DNS-based and the RTT-proximity datasets separately to find if they take advantage of the location hints in the hostnames of all the DNS-based dataset addresses. The RTT-proximity dataset has 1,335 addresses (27.6% of RTT-proximity data) that do not have DNS names. We do not know how many of the remaining RTT-proximity addresses have useful city-level location hints in their hostnames. Note that we include the 109 common addresses between the two datasets only as part of the DNS-based dataset. Overall, a database that uses DNS-based techniques to decode location hints in hostnames is expected to perform better on the DNS-based ground truth dataset.

NetAcuity is the only database that shows better city-level accuracy results over the DNS-based data compared to results over the RTT-proximity data. Considering our city-range threshold of 40 km, NetAcuity has 70.1% overall city-level accuracy over the RTT-proximity data and a better 74.2% accuracy over the DNS-based data. All other databases do worse over the DNS-based compared to the RTT-proximity data. MaxMind-Paid, for example, has only 43.9% overall accuracy over the DNS-based data and 66.5% over the RTT-proximity data. Regionally, NetAcuity shows better accuracy in all regions over the DNS-based data. For example, NetAcuity has 55.1% accuracy for ARIN addresses in the RTT-proximity data, and about 70.6% for ARIN addresses in the DNS-based data. According to these results, NetAcuity is the only database that might be using some DNS-based techniques to infer location hints from hostnames.

While the databases results over the RTT-proximity data look more competitive, NetAcuity still outperforms other databases over this dataset considering both accuracy and coverage. NetAcuity has a 70.1% and 99.6% city-level accuracy and coverage respectively. The closest rival, MaxMind-Paid, has a comparable 66.5% accuracy but only 50.3% city-level coverage.

6 RECOMMENDATIONS

Based on our analysis of the geolocation databases using our datasets described in §2, we present our recommendations for using the databases with two thoughts in mind. First, our recommendations



(a) MaxMind-Paid (41.29% of ground truth data). **(b) NetAcuity (99.6% of ground truth data).**

Figure 5: Databases vs. ground truth geolocation error breakdown by RIR. Only routers with city information are included.

are mostly meaningful in ARIN and RIPENCC—where most of our ground truth IP addresses are located—and to a less degree in APNIC regions. Second, NetAcuity might have benefited from the nature of the DNS-based ground truth data (see §5.2.4), but we still argue that it has the best accuracy and city-level coverage as the results over both ground truth datasets show. With that in mind, here are the recommendations:

- If using a geolocation database is the only available option, we recommend NetAcuity to geolocate routers. Note that we think of NetAcuity city-level accuracy of 74.2% over the DNS-based data as an upper bound for its overall accuracy. NetAcuity appears to benefit from the location hints encoded in the hostnames of the DNS-based dataset IP addresses.
- We do not recommend MaxMind databases if high city-level accuracy and coverage are required. The city-level accuracy is especially bad in the ARIN region. But we do see relatively good city-level results for MaxMind-Paid in RIPENCC and APNIC regions. However, the city-level coverage is very low.
- The commercial version of MaxMind is recommended over the public version if city-level accuracy and better coverage are required.
- We do not recommend IP2Location-Lite, the overall accuracy is too low especially at city-level.
- If price is a problem and an overall 78% country-level accuracy is acceptable, the IP2Location-Lite and both versions of MaxMind were comparable. That said, the accuracy can be very low for some countries (see Figure 4).
- We recommend users not to trust city-level accuracy in ARIN regardless of the database used. NetAcuity was the most accurate there, but only 66% of the ground truth interface addresses there are geolocated to within 40 km of their actual locations.

7 RELATED WORK

Several studies have shown that public and commercial databases have coarse-grained granularity and are not reliable at city-level resolution [10, 15, 26, 29, 30]. Poese et al. [26] studied the relationship between prefixes in several databases and those advertised by a large European ISP. They found that databases split large ISP blocks into smaller ones for more accuracy. However, they reported that this did not improve accuracy. Huffaker *et al.* [15] used majority vote across all participating databases to pick the location of a given block of IP addresses and then evaluated the databases according to the resulting location. Shavitt *et al.* [29] examined the coherency of databases using a ground truth dataset of IP addresses with known PoP (Points of Presence). They also used a majority vote across all databases to infer the location of a PoP and then compared all databases to the inferred location. While these studies evaluate the overall accuracy of the geolocation databases, our work focuses on evaluating router geolocation in databases. Our ground truth is not specific to an ISP or region or PoPs as in [26] and [29]. We also do not use delay measurements as in [13] and [10] to study the geographic span and co-locality of IP blocks.

Another category of related work is the DNS-based geolocation that infers location hints from DNS names. Huffaker *et al.* work [16], discussed is §2.3.1, is one example. Scheitle *et al.* work[27] is similar to [16] where location hints are extracted from DNS names and then verified or disqualified using latency measurement. While DNS-based methods can provide good accuracy results, their scope is limited since not all router addresses have DNS names, and not all names have useful geolocation hints. In our work we used [16] to create part of our ground truth dataset.

8 CONCLUSIONS

In this paper, we evaluate router geolocation in four widely used geolocation databases. We examine their consistency and coverage using a dataset of 1.64M router interface addresses. We show that the databases generally agree on the country-level (95.8% of the time), but the databases—from different vendors— show more discrepancy at city-level with more than 29% pairwise disagreements. However, we show that agreement among the databases does not imply correctness. We evaluate the accuracy of the databases with a ground truth dataset we created using DNS-based and latency-based methods. We show that the databases are not accurate in geolocating routers at neither country- nor city-level, even if they agree significantly among each other. A breakdown by RIR of the ground truth shows that the databases are less reliable at the city-level resolution in ARIN compared to other regions. NetAcuity shows the best combination of coverage and accuracy. MaxMind shows relatively good city-level accuracy in regions other than ARIN but it lacks extensive city-level coverage. Overall, comparing our router geolocation accuracy results with previous work on databases evaluation suggests databases geolocate routers with less accuracy compared to end hosts. Researchers need to pay extra caution when using geolocation databases and understand the impact of the databases accuracy on their results.

ACKNOWLEDGMENTS

The authors would like to thank Digital Envoy NetAcuity for making their database available to this work, as well as Vasileios Giotsas for sharing his router geolocation dataset with us. We thank kc claffy, Matt Calder, anonymous IMC reviewers, and our shepherd, Ethan Katz-Bassett, for their valuable comments on earlier drafts. This material is based on research sponsored by the Department of Homeland Security Science and Technology Directorate, Homeland Security Advanced Research Projects Agency, Cyber Security Division, BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement numbers FA8750-12-2-0326, FA8750-12-2-0344, and FA8750-15-2-0224. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Department of Homeland Security, Air Force Research Laboratory or the U.S. Government.

REFERENCES

[1] CAIDA. 2016. AS Rank. http://as-rank.caida.org. (2016).
[2] CAIDA. 2016. CAIDA Topology Dataset. (March 2016). Retrieved March, 2016 from https://topo-data.caida.org/
[3] CAIDA. 2016. Internet Mapping and Annotation. http://www.caida.org/research/topology/internet_mapping/. (2016).
[4] CIA. 2017. The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/2011.html. (2017).
[5] Team Cymru. 2016. IP to ASN mapping. http://www.team-cymru.org/IP-ASN-mapping.html. (August 2016).
[6] DB-IP. 2017. The DB-IP Database. https://db-ip.com. (August 2017).
[7] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. USENIX, Washington, D.C., 605–620.
[8] Digital Envoy. 2016. Digital Element NetAcuity databases. https://www.digitalelement.com/netacuity/. (July 2016).
[9] GeoNames. 2017. The GeoNames Geographical Database. (January 2017). Retrieved January, 2017 from http://www.geonames.org/
[10] Manaf Gharaibeh, Han Zhang, Christos Papadopoulos, and John Heidemann. 2016. Assessing Co-locality of IP Blocks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2016 IEEE Conference on*. IEEE, 503–508.
[11] Vasileios Giotsas, Amogh Dhamdhere, and Kimberly C. Claffy. 2016. Periscope: Unifying Looking Glass Querying. In *Passive and Active Measurement - 17th International Conference, PAM 2016, Heraklion, Greece, March 31 - April 1, 2016. Proceedings*. 177–189.
[12] Vasileios Giotsas, Petros Gigis, Alexandros Milolidakis, Eric Nguyen Duy, Marios Isaakidis, and Edwards. Mukasa. 2016. The Remote Peering Jedi a Portal in the Remote Peering Ecosystem *(RIPE 73)*.
[13] Bamba Gueye, Steve Uhlig, and Serge Fdida. 2007. Investigating the Imprecision of IP Block-based Geolocation. In *Proceedings of the 8th International Conference on Passive and Active Network Measurement (PAM'07)*. Springer-Verlag, Berlin, Heidelberg, 237–240.
[14] Bamba Gueye, Artur Ziviani, Mark Crovella, and Serge Fdida. 2006. Constraint-based Geolocation of Internet Hosts. *IEEE/ACM Trans. Netw.* 14, 6 (Dec. 2006), 1219–1232.
[15] Bradley Huffaker, Marina Fomenkov, and kc claffy. 2011. *Geocompare: a Comparison of Public and Commercial Geolocation Databases*. Technical Report. Cooperative Association for Internet Data Analysis (CAIDA).
[16] Bradley Huffaker, Marina Fomenkov, and kc claffy. 2014. DRoP: DNS-based Router Positioning. *Computer Communication Review* 44, 3 (2014), 5–13.
[17] MaxMind Inc. 2016. MaxMind GeoIP2 City. https://www.maxmind.com/en/geoip2-databases. (July 2016).
[18] MaxMind Inc. 2016. MaxMind GeoIP2 Country Database. https://www.maxmind.com/en/geoip2-country-database. (July 2016).
[19] MaxMind Inc. 2016. Maxmind GeoLite databases. (July 2016). Retrieved July, 2016 from http://dev.maxmind.com/geoip/legacy/geolite/
[20] IP2Location. 2016. IP2Location LITE Databases. (July 2016). Retrieved July, 2016 from http://lite.ip2location.com
[21] IP2Location. 2017. IP2Location Databases. http://www.ip2location.com. (August 2017).
[22] 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*.
[23] RIPE NCC. 2016. RIPE NCC Measurements. (May 2016). Retrieved May 25, 2017 from https://atlas.ripe.net/measurements/
[24] Venkata N. Padmanabhan and Lakshminarayanan Subramanian. 2001. An Investigation of Geographic Mapping Techniques for Internet Hosts. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*.
[25] Paul Pearce, Roya Ensafi, Frank Li, Nick Feamster, and Vern Paxson. 2017. Augur: Internet-Wide Detection of Connectivity Disruptions. In *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*. 427–443.
[26] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP Geolocation Databases: Unreliable? *SIGCOMM Comput. Commun. Rev.* 41, 2 (April 2011), 53–56.
[27] Quirin Scheitle, Oliver Gasser, Patrick Sattler, and Georg Carle. 2017. HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks. In *Network Traffic Measurement and Analysis Conference (TMA)*. Dublin, Ireland.
[28] Anant Shah, Romain Fontugne, and Christos Papadopoulos. 2016. Towards Characterizing International Routing Detours. In *Proceedings of the 12th Asian Internet Engineering Conference (AINTEC '16)*. ACM, New York, NY, USA, 17–24.
[29] Yuval Shavitt and Noa Zilberman. 2011. A Geolocation Databases Study. *IEEE Journal on Selected Areas in Communications* 29, 10 (2011), 2044–2056.
[30] S. S. Siwpersad, Bamba Gueye, and Steve Uhlig. 2008. Assessing the Geographic Resolution of Exhaustive Tabulation for Geolocating Internet Hosts. In *Proceedings of the 9th International Conference on Passive and Active Network Measurement (PAM'08)*. Springer-Verlag, Berlin, Heidelberg, 11–20.
[31] Meenakshi Syamkumar, Ramakrishnan Durairajan, and Paul Barford. 2016. Bigfoot: A Geo-based Visualization Methodology for Detecting BGP Threats. In *2016 IEEE Symposium on Visualization for Cyber Security, VizSec 2016, Baltimore, MD, USA, October 24, 2016*. 1–8.
[32] Yong Wang, Daniel Burgener, Marcel Flores, Aleksandar Kuzmanovic, and Cheng Huang. 2011. Towards Street-level Client-independent IP Geolocation. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation*.
[33] Bernard Wong, Ivan Stoyanov, and Emin Gün Sirer. 2007. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. In *The 4th USENIX conference on Networked systems design & implementation*.