Geocompare: a comparison of public and commercial geolocation databases

CAIDA Tech Report, May 16, 2011

Bradley Huffaker, Marina Fomenkov, kc claffy
{bradley,marina,kc}@caida.org
CAIDA, University of California, San Diego*

ABSTRACT

We attempt a systematic quantitative comparison of currently available geolocation service providers. We add depth to previous contributions by analyzing inconsistencies across databases for different geographic (RIR) regions and organization (Autonomous Systems) types. We compare results on a country granularity, using a methodology that compares each database against the majority vote across all databases with answers for a given IP address. On a finer granularity than country, rigorous formal comparison gets trickier. Unlike the discrete country labels, coordinates can have nominally different values yet still represent approximately the same location, We compare the databases at a lat-long granularity using an 80 km threshold for two lat-longs coordinates to be in the same geographic region. We describe our process for selecting this threshold, and our centroid-based algorithm for comparing database lat-long results against a majority of responses from the set of databases we evaluated. While not a foolproof methodology – the databases could all be converging to the same wrong answers over time – it assumes that database providers successfully work toward improving the accuracy of their databases over time. In the absence of substantial ground truth, our method offers a systematic way to study the geolocation databases to reveal insights, summarized at the end of the paper. We intend to re-run the comparison experiment using additional databases later in 2011; we welcome constructive feedback on the methodology so we can further improve our next experiment.

1. INTRODUCTION

Governments, researchers, and commercial entities share an interest in mapping Internet resources to physical locations, a process termed *geolocation*. For example, governments use this data to prepare and plan for adverse events as well as to tax and regulate. Academics use this data to more accurately capture the geographic deployment and utilization of Internet resources. Commercial interests use this data to provide better localized services, target pricing or ads, enforce digital rights management restrictions or data privacy requirements, and assign incoming requests for content to the nearest data center storing it.

One can broadly categorize geolocation techniques based on the main source of knowledge driving the geolocation: delay, database, or topology. Delay-based methods typically gather delay data from a collection of known geographic landmarks and use that knowledge to triangulate a target IP address [29, 34, 31, 37, 40, 30]. Database-driven methods collect and aggregate static mapping information from public and private databases [36, 32]. Topology-driven methods infer geolocations by assuming topologically close addresses are physically nearby each other [33, 34].

These techniques involve non-trivial hurdles that make building one's own service prohibitively difficult for the average user. Private databases, such as a web site's listing of its users' IP addresses and contact information, are generally inaccessible. Public databases, such as the WHOIS [27] and DNS [35], are difficult to parse, hard to keep current, and also may have access controls. Topology and delay data require extensive measurement infrastructure, and are challenging to collect, process, and interpret. Therefore, the majority of users rely upon third-party geolocation services.

Most geolocation databases, both publicly [14, 16, 19, 25] and commercially available [6, 1, 11, 7, 8, 17, 18, 9, 10], map blocks of consecutive IP addresses to geographic locations, usually at the country level. Some providers support city mappings and/or resolve to latitude and longitude coordinates. Most services offer little if any documentation on which techniques they employ in the creation of their geolocation databases, thus complicating systematic attempts at evaluation and comparison. Previous comparison attempts [38, 39] have also noted that the lack of a large and diverse set of ground truth further challenge rigorous comparison.

In this study we attempt a systematic quantitative comparison of currently available geolocation service providers. We compare only the geographic components of the participating databases; most providers offer additional features that we did not evaluate. We add depth to previous contributions by analyzing inconsistencies across databases for different geographic regions and organization (Autonomous Systems)

*Support for this work is provided by DHS N66001-08-C-2029 and NSF 05-51542.

types. We use Regional Internet Registries (RIR) delegation data to classify addresses by the RIR to which the address block was first delegated. We infer the organization type corresponding to an IP address based on the characteristics of its origin AS in the AS graph, building on the approach in [28]. Section 2 describes previous comparative geolocation studies. Section 3 discusses the datasets we used in our work. We present data analysis and results in Section 4 and conclude in Section 5.

2. RELATED WORK

Despite the unavoidable obstacles to systemic comparison, several research groups have made efforts to evaluate geolocation services. Siwpersad *et al.* [39] examined the geographic resolution of MaxMind GeoLite [19] and Hexasoft [8]. They calculated the distance between locations as determined by each service and found that for 50% of addresses the difference was smaller than 100 km. They also compared the database locations with inferences from active measurement data collected by PlanetLab nodes probing 39 landmarks. The authors inferred geographic locations from the collected data using a constraint-based approach with a series of confidence regions. They found that for 90% of probed IP addresses, their location fell outside the confidence regions estimated by active measurements.

Shavitt *et al.* [38] examined HostIP [14], IP2Location [8], IPInfoDB [16], MaxMind GeoIP [18], and Software77 [25]. They evaluated these databases through the lens of the DIMES Project's [26] Points of Presence (PoP) level map. The authors attributed IP addresses to PoPs using their own interface-graph-based inference algorithm [38], and assumed that IP addresses in the same PoP should map to the same geographical location. They found that MaxMind GeoIP, GeoBytes, and Digital Envoy placed between 74% and 82% of a PoP's IP addresses within 1 km of each other while for HostIP the percentage was slightly less (57%). To compare across databases, Shavitt *et al.* defined PoP coordinates as the median latitude and longitude of all the coordinate values found in all databases for all IP addresses at the given PoP. They considered two levels of proximity: a "city" range (40 km), and a "region" range (500 km). For IPligence, MaxMind GeoIP, and IP2Location, the probability of identifying the location of an IP address within the "city" range of its PoP ranged between 62% and 73%, while for GeoIP, HostIP, and Digital Envoy it was between 33% and 47%. MaxMind GeoIP placed over 80% of IP addresses, while Geobytes, HostIP, and Digital Envoy placed 48% to almost 60% of IP addresses into the PoP "region" range.

3. DATA SETS

3.1 Geolocation Databases

Our goal was to include as many geolocation databases as possible into this study. However, despite our best efforts, we could not obtain data from all the databases we know about. GeoBytes [7] and IP2location [8] did not respond to our inquiries. Quova [9] and Akamai [6] were prohibitively expensive, requesting more than $10,000 a year for their services. We used free services from Software77 [25], MaxMind GeoLiteCity [19], HostIP [14], and IPInfoDB [16] and purchased access to Cyscape [1]. CAIDA has an ongoing research agreement with Digital Envoy [11] as our primary geolocation data provider.

In this paper we refer to the database with the name of the organization providing it, e.g., NetAcuity is labeled Digital Envoy. Table 1 lists free and commercial databases evaluated in this study and their basic statistics: the date of the database snapshot; the cost of service; the percentage of the IPv4 address space covered (relative to the range delegated in the RIR delegation files, described in Section 3.1.1); the number of unique address blocks; and the numbers of countries, cities, and latitude/longitude values found in each database.

RIR delegation files, though never intended as a geolocation service, are included in the Table 1 as a baseline. Software77's database is essentially a processed version of the RIR delegation files. Of the databases examined, HostIP covers the smallest fraction of the IPv4 address space because this free, open database is populated by volunteers submitting their geographic locations. IPligence is the least expensive commercial database among those of our study.

We did not receive a full dump of Cyscape's geolocation database, so we sampled its database to infer the full table using the blocks from the largest two other databases: IPligence [17], and MaxMind GeoIP [18]. If geographic answers for two contiguous blocks were identical, we merged these blocks into a single larger block; if the answers for the first and the last address of a particular block differed, then we subdivided the block further until every address in a block had the same geolocation mapping.

IPligence's database had a larger number of unique cities than unique lat-long coordinate pairs, due to some typographical variance in city names. For example in addition to Uplands Vasby we also found Upplandsvasby and Upplands-Vasby. In a few cases many suburb names shared lat-long coordinates, e.g., Kungsholmen, Stockhom, Bandhagen, Johannehov, Johanneshov, Stochholm, and Stockholm are all part of Stockholm City and share 59.33, 18.05.

MaxMind GeoLite is a publicly available less accurate version of MaxMind GeoIP. As described on their web page, IPInfoDB derives its results largely from the MaxMind GeoLite dataset [15], and the two databases performed indistinguishably for most of our metrics. Therefore, we have excluded IPInfoDB data from our plots; the reader can safely assume its answers are the same as those of MaxMind GeoLite.

Digital Envoy's Netacuity[1] database contains the largest

[1]Digital Envoy also offers NetAcuity Edge, which relies on anonymous user supplied data from some of their partners, combined with infrastructure targeting. We hope to test this database in our

Table 1: Geolocation service provider database statistics.

Database	cost[1]	date	addr[2]	blocks	countries	cities	lat,long
RIR$_f$	-	2010.10.31	100.0%	105,380	229	-	-
Software77$_f$	-	2010.12.01	99.5%	105,334	229	-	-
HostIP$_f$	-	2010.10.04	15.9%	780,287	216	-	23,906
IPligence	$	2010.10.26	95.7%	3,155,326	232	59,194	56,353
Cyscape	$$	2010.08.31	96.8%	54,788	234	-	-
MaxMind GeoIP	$$$[3]	2010.12.01	100.0%	5,774,006	239	128,368	130,707
MaxMind GeoLite$_f$	-[4]	-	-	-	-	-	-
IPInfoDB[5]$_f$	-	2010.12.01	100.0%	3,533,709	228	113,209	115,950
Digital Envoy	$$$$	2010.12.02	100.0%	6,082,327	241	33,247	33,195

Indented databases are derived from the database in the row above.
[1] cost of unlimited geolocation for one year: $ = $1-$300 $$ = $300-$900 $$$ = $900-$1800 $$$$ = $1800+
[2] out of RIR delegated addresses
[3] $370 site license one month updates, $90 per month of updates thereafter
[4] IPInfoDB is almost indistinguishable from MaxMind GeoLite and is not individually displayed in the rest of the paper.
$_f$ marks the free datasets

number of address blocks, mapping IP addresses into locations at the finest granularity. MaxMind's GeoIP database has only 4.2% fewer blocks than the Digital Envoy one, but maps addresses to four times as many unique cities. However, as we will discuss in Section 4, this additional granularity at the city level does not result in a performance benefit that was captured by our testing.

3.1.1 Regional Internet Registries Breakdown

The Regional Internet Registry (RIR) delegation files [20, 21, 22, 23, 24] provide a regional baseline for our comparative study since each RIR is responsible for a certain geographic region: **AfriNIC** for Africa; **ARIN** for the United States, Canada, and several parts of the Caribbean region; **APNIC** for Asia, Australia, and neighboring countries; **LACNIC** for Latin America and parts of the Caribbean region; and **RIPE** for Europe, the Middle East, and Central Asia. We limited our study to the addresses covered by address blocks in RIR delegation files. To determine which RIR delegated a given address, we downloaded and parsed the delegation files from October 31, 2010.

Figure 1 displays the breakdown of the whole IPv4 address space explored in our study and of each database address blocks in RIR delegations. The leftmost column is a baseline showing the *percentages of addresses* in the evaluated address space delegated by each registry. All remaining columns show the *percentages of blocks* in each geolocation database by RIR geographic region. Comparing these percentages to the baseline, we notice that databases subdivide the address space into blocks unevenly, sampling different geographic regions with blocks of different average sizes. The majority of databases have a disproportionately large number of blocks from **RIPE** (Europe), reflecting a higher

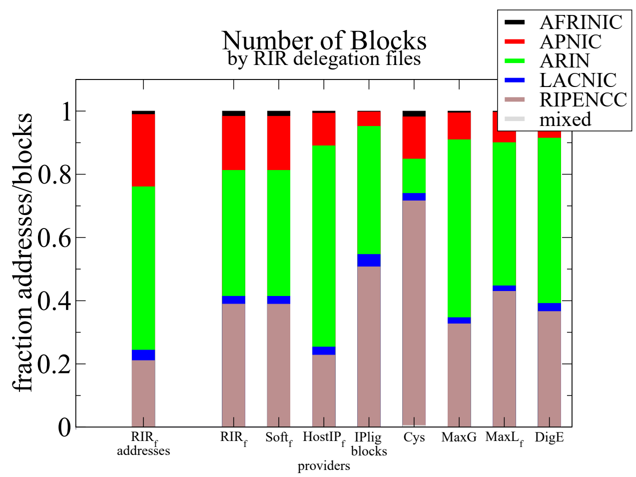


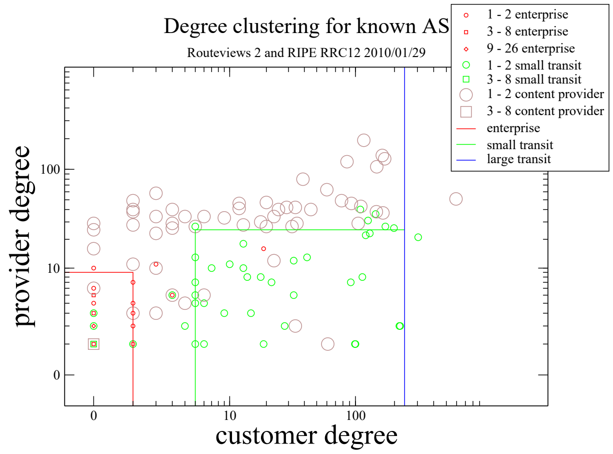
Figure 1: *Database breakdown by RIR. Most databases have more blocks from* RIPE NCC *than other regions, reflecting the higher number of connected countries in the European region.*

number of countries in that region and a necessity to use blocks of smaller average size for these countries. In contrast, the fraction of APNIC region blocks is smaller then the fraction of addresses in this region for almost all databases – meaning that these geolocation databases have relatively larger address blocks for Asia, Australia, and neighboring countries. Finally, block sizes used by Cyscape to cover the ARIN region are larger then those by other geolocation providers, therefore the fraction of ARIN blocks in the Cyscape column is disproportionally small.

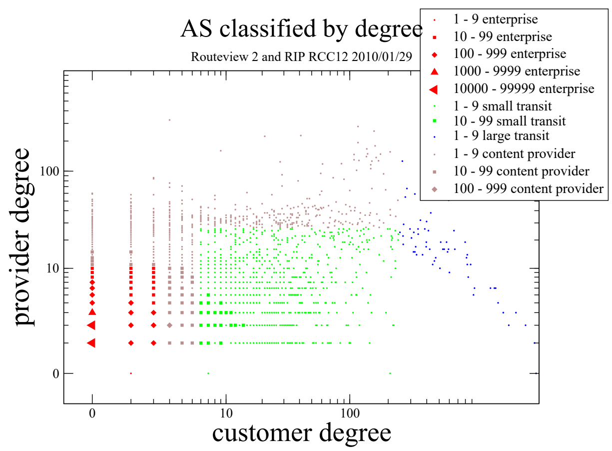
3.1.2 Organizational Breakdown

In order to analyze whether some geolocation databases perform better for certain organization types than others, we classified BGP-announced address blocks by organization type. First, we determined origin ASes for address space

next experiment.



(a) Density plot for the training set of 50 ASes with manually drawn bounding boxes that separate most ASes of a given type from other types.



(b) Density plot for the full AS dataset mapped on the boundaries defined from the training dataset.

Figure 2: ***Classification of ASes into four types: Enterprise Customers, Content/Access/Hosting Providers, Small Transit Providers, and Large Transit Providers.***

blocks probed in our study using tables from the Route Views database [4] and RIPE NCC [3] Border Gateway Protocol (BGP) collectors from the January 29, 2010.

Next, following the approach in [28], we considered four classes of ASes: **Enterprise Customers (EC)**, typically organizations, universities and companies at the edge comprised of mostly users; **Content/Access/Hosting Providers (CAHPs)**, also at the edge, but typically provide content and/or Internet access (e.g. DSL, cable modems, etc); **Small Transit Providers (STP)**, which provide transit to smaller ASes, in addition to any content and access services, but purchase transit for some routes from a large Transit Provider; and **Large Transit Providers (LTP)**, which provide the same services as the **STP**, but have sufficient topological coverage (reach) that they do not need to pay any provider for transit.

In [28], Dhamdhere and Dovrolis manually classified 50 ASes creating a training set and classification rules that can be applied to a larger population. For a sample of ECs, they picked well-known universities and corporations. For STPs, they selected transit providers that are mostly regional in terms of their coverage and customer size. For CPs and AHPs, they picked well-known content providers, hosting sites, and large broadband/dial-up residential and/or business access ISPs.

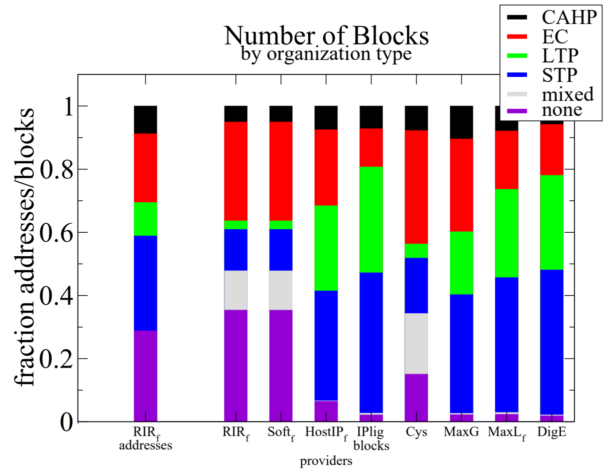


Figure 3: ***Database breakdown by organization type. Almost 28% of addresses and 34% of RIR blocks are not yet announced and so classified as* None*. HostIP, IPligence, Maxmind, and Digital Envoy have fewer then 2% of their blocks in this category, reflecting understandably greater attention on the part of geolocation providers to the actually routed address space.***

For each AS, both in the training set and in the full set, we determined the number of its customers and providers as inferred in the January 29, 2010 version of CAIDA's AS Rank [2] AS Relationship dataset. Using the training set, we selected boundaries separating AS types in the two-dimensional space with customer degree on the X-axis and provider degree on the Y-axis (Figure 2(a)). We then mapped the full set of 35577 ASes onto those boundaries (Figure 2(b)). Our categories are essentially regions of this two-dimensional (customer degree, provider degree) space, established from a ground truth sample of 50 well-known ASes.

Figure 3 presents the breakdown of the whole IPv4 address space explored in our study and of each database address blocks by organization type. Similar to Figure 1, the leftmost column is again a baseline showing the *percentages of RIR addresses* in the tested address space announced by ASes of different types. The remaining columns show the *percentages of blocks* announced by different organization types for the compared geolocation databases. Note that our classification of ASes is based on the BGP tables and therefore available only for routable addresses. However, almost 28% of addresses and 34% of blocks in RIR files are delegated, but not yet announced in the BGP tables. We labeled

Table 2: **Ground Truth set statistics.**

Database	date	addr[1]	countries	cities	lat,long
PlanetLab	2010.12.03	1,067 (0.0%)	1	-	397
French networks	2010.12.24	6,010,880 (0.2%)	1	2,694	2,680
US Tier 1	2011.01.27	23,644 (0.0%)	28	133	133

[1] out of RIR delegated addresses

these addresses as **none**. Also, some delegated blocks become split and announced by organizations of different type according to our classification. These blocks are labeled as **mixed**.

As noticed in Section 3.1, Software77's database is an immediate progeny of the RIR delegation files. Therefore, the breakdowns of their blocks by organization types are very similar. In all other databases in our study, the percentage of the unclassified blocks (**none**) is much lower because understandably, geolocation providers focus their attention on mapping the actually routed address space.

HostIP, IPligence, Maxmind GeoIP and Digital Envoy have 50% or more of their address blocks classified as **LTP** or **STP**, in contrast to less than 16% of such blocks in RIR data. Transit-providing organizations can typically justify sizable allocations from the RIRs, thus their fractions are much larger in the leftmost "RIR addresses" column than in the next "RIR blocks" column. Large transit providers also usually cover a wide geographic region, the largest of them being present on at least three continents. They suballocate parts of their address space to their customers in different geographic locations. This expansive geographic reach means that geolocation providers have to cut their address space into smaller and smaller blocks in order to capture the resulting geographic diversity.

We use this classification of address blocks to compare geolocation databases as a function of inferred organization (AS) type.

3.2 Calibration Datasets

3.2.1 Ground Truth Data

Table 2 lists the three separate ground truth datasets we used as a baseline for our comparisons. PlanetLab is a globally distributed set of computers available as a testbed for computer networking and distributed systems research. PlanetLab publishes latitude and longitude coordinates of its 398 participating sites [13]. Freebox provides a list of French ADSL networks by geographic region [12]. Finally, a large U.S. transit provider gave us a full dump of their DNS hostname to IP address mappings and the heuristic to map the names to geographic locations.

3.2.2 Archipelago Round Trip Time Datasets

As additional calibration data, we used Round Trip Time (RTT) data collected in the course of CAIDA macroscopic IPv4 topology measurements on the Archipelago infrastructure [5]. Ark monitors continuously probe the routed IPv4 address space recording RTT values from each responding hop along the forward path.

RTT is the amount of time it takes for a probe packet to be sent from a monitor to a target address, the target to generate a reply, and that reply to return to the monitor. Many factors that a monitor cannot observe affect the RTT – from congestion delay to indirect paths – thus adding noise to the recorded RTT value. Nevertheless, in Section 4.4 we discuss how we estimate the expected value of RTT, given the known location of our Ark monitor and the latitude/longitude location of the target address provided by each database. We then compare the estimated and actually observed RTT values.

4. ANALYSIS

To compare participating providers, we tried to geolocate every IPv4 address in an RIR delegation and performed four evaluations: consistency of country-level resolution; consistency of lat/long coordinate resolution; comparison with a limited ground truth sample; and calibration against measured RTT. All evaluations included only resolved addresses, i.e., if a database did not have an answer for a given address, then we excluded this address from all counts for this database.

4.1 Consistency of IP-to-country Resolution

Since we lack ground truth for the full dataset, we evaluated the consistency of the country-level resolution by comparing the location provided by a given database with the location provided by the majority of the databases, and then counting incidents of each database agreeing or disagreeing with the majority answer.

To have a successful election of the majority answer, there must be at least two databases resolving an address, and one country must receive more votes than others. We determined the majority answers from the databases of the six providers listed in Table 1. We excluded MaxMind GeoLite and IPInfoDB from the election process since they both derive from MaxMind GeoIP and counting all three would skew the majority answers toward the MaxMind GeoIP version.

We found that for 94.5% of addresses, a majority of databases were able to agree on a single country, and for 1.2% there was a tie. As shown in Figure 4, all but one (IPligence) geolocation databases in our study agreed with the majority

Table 3: Pairwise country comparison between databases. The number in each cell is the percentage of addresses for which the country determined by the row database matched the country provided by the column database. For example, the RIR data matched 99.9% of Software77's country-level answers, but Software77 only matched 99.4% of RIR's country-level answers (since Software77's database is a little smaller). HostIP's low agreement with other databases is due to its small size, while IPligence answers are indeed different from other databases.

	RIR	Soft	HostIP	IPlig	Cys	MaxG	MaxL	DigE	avg[1]
RIR$_f$	-	99.9	88.9	89.3	93.6	94.1	94.2	91.8	93.8
Software77^v_f	99.4	-	88.8	88.6	93.0	93.5	93.6	91.2	91.6
HostIP$_f$	14.1	14.2	-	13.9	15.4	14.4	14.4	14.9	14.6
IPligence^v	85.4	85.3	83.8	-	89.3	89.5	89.6	86.2	87.6
Cyscape^v	90.7	90.6	94.2	90.4	-	93.2	93.3	95.7	92.5
MaxMind GeoIP^v	94.1	94.0	90.9	93.5	96.2	-	99.8	94.9	94.7
MaxMind GeoLite$_f$	94.2	94.1	91.0	93.6	96.3	99.8	-	94.9	95.8
Digital Envoy^v	91.8	91.7	93.9	90.0	98.8	94.9	94.9	-	93.9
average[1]	92.3	90.4	90.3	90.6	94.3	92.8	94.3	92.0	-

color key 0-59 60-69 70-79 80-89 90-100

[1] average is only calculated across ^{v} voting databases
^{v} databases used in the calculation of the average
$_f$ free databases

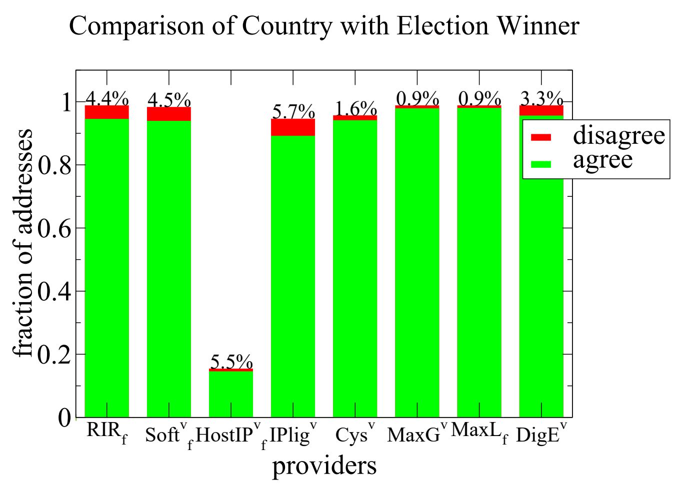


Figure 4: ***Agreement of each database with the majority answers for IP-address-to-country mappings. The value at the top of each column is the percent of addresses for which this database disagreed with the majority. A superscript*** ^{v} ***marks providers voting in the country election process.***

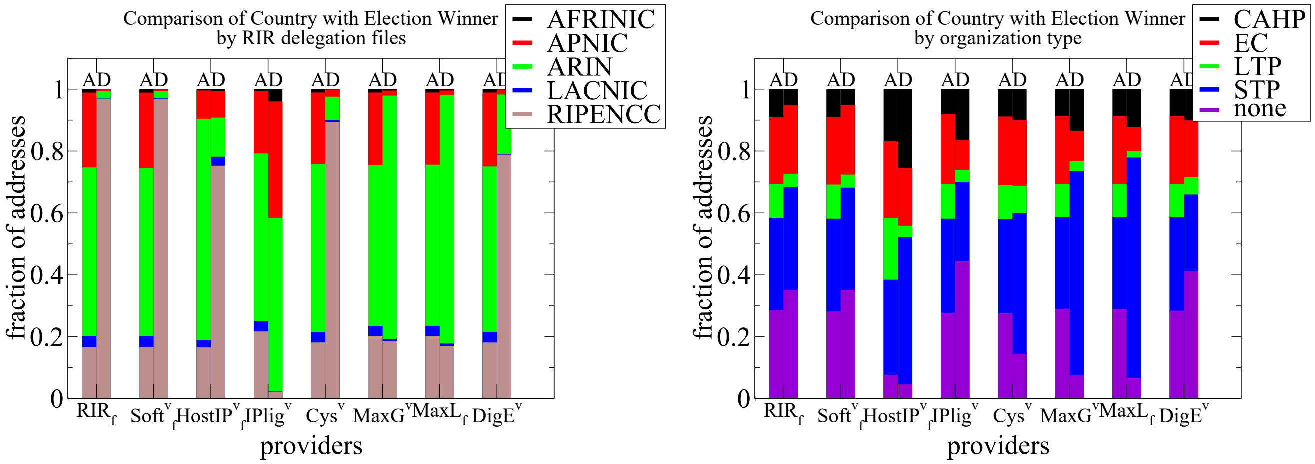
answer in more than 95% of cases. MaxMind GeoLite had the lowest disagreement with the majority (0.9%), followed by Cyscape with 1.6%. IPligence had the highest disagreement (5.7%) with the majority on country determinations. Country identification from RIR delegation files disagreed with the majority answer for 4.4% of addresses, suggesting the prevailing majority of addresses are typically being used in the country to which they were delegated.

Pairwise disagreements between databases are shown in Table 3. Each number is a percentage of the row database's addresses that had a matching country in the column's database. The corresponding average percentages of pairwise agreements are in the last column and the last row. We excluded the RIR data, HostIP, and MaxMind GeoLite from calculation of the averages: HostIP because it is too small, and RIR and MaxMind GeoLite since they would bias upward the influence of Software77 and MaxMind GeoIP, correspondingly.

HostIP typically matched 14-15% of other databases' answers, while other databases matched 84-94% of HostIP's entries. This disparity is the result of HostIP's small size. Next, IPligence had the lowest agreement with others, matching on average only 87.6% of other databases' answers. MaxMind GeoLite matched the largest percentage of entries, 95.8% on average.

Figure 5 explores the consistency of IP-to-country mapping across geographic regions (5(a)) and organization types (5(b)). For each database, two columns are depicted, the first one (marked with **A**) shows the breakdown for the answers agreeing with the majority answer, while the second one (marked with **D**) represents the breakdown of disagreeing answers. To enable comparison of breakdowns for agreement and disagreement columns, we normalize each column to 100% (otherwise the A column, containing > 90% of addresses for each provider, would dwarf the D column). Generally, for each database, the A column's breakdown matches the overall distribution of addresses included in this database.

Considering the breakdown by geographic regions (Figure 5(a)) we notice that the RIR database and its progeny Software77 strongly correlate in how they disagree with the majority, having most inconsistencies for addresses in the RIPE region. HostIP, Cyscape and Digital Envoy also have the largest fractions of their disagreements for RIPE addresses, but with increased fractions for ARIN addresses. In contrast, both MaxMind databases disagree with the majority primar-



(a) breakdown by geographic regions

(b) breakdown by organization types

Figure 5: ***Consistency of country-level resolution. These graphs show the breakdown of addresses for which each database agreed (A column) or disagreed (D column) with the majority vote. To enable comparison between the breakdowns of agreement and disagreement columns, each column is normalized by the total number of addresses that it represents, rather then by the total number of addresses from the given provider. Figure 4 shows the percentage of data represented in the Disagreement columns.***

ily for addresses in the ARIN region, while IPligence has the largest fraction of disagreements for addresses in the APNIC region.

The breakdown by organization types (Figure 5(b)) shows that both MaxMind databases have disproportionately more inconsistencies with the majority answers for **STP** addresses. For all other providers the breakdowns of Agreement and Disagreement columns approximately match.

4.2 Consistency of IP-to-coordinates resolution

On a finer granularity than country, rigorous formal comparison gets trickier. Unlike the discrete country labels, coordinates can have nominally different values yet still represent approximately the same location, We developed a method that we admit is imperfect, but in the absence of ground truth, gives us a systematic way to explore the databases and analyze the consistency of their latitude/longitude answers.

For each IP address, we looked up its coordinates in the HostIP, IPligence, MaxMind GeoIP, and Digital Envoy databases. (As for our country-level analysis, we did not include answers from IPInfoDB and MaxMind GeoLite, since they are essentially subsets of MaxMind GeoIP.) By our definition, the resulting coordinate points obtained for a given IP address from different databases form *a cluster* if they are all within a certain threshold T distance of each other. Points can belong to more than one cluster. A *winning cluster* contains the largest number of points. For each winning cluster, if there is one, we find its centroid, that is the point whose coordinates are the average coordinates of all members of this cluster. We consider this centroid position of an IP address as the yardstick against which we compare coordinate locations given for this address by individual databases. While not a foolproof methodology – the databases could all be converg-

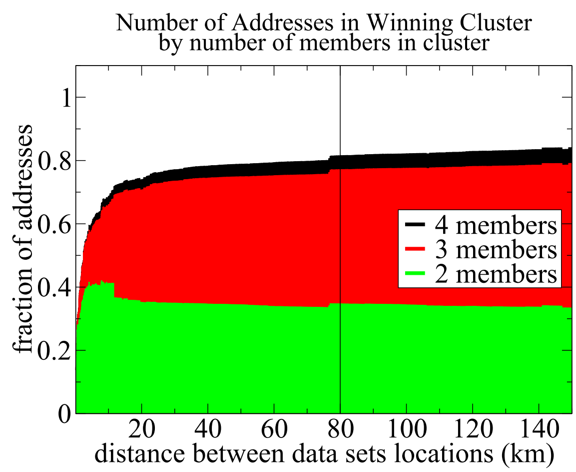


Figure 6: ***Up to 11.5 km, the majority of winning clusters are formed with only two members. After that point, the majority of clusters have three or more members and the rate of adding new clusters slows. The last jump in two member clusters is at 77.7 km. We selected our threshold value T = 80 km to include these addresses.***

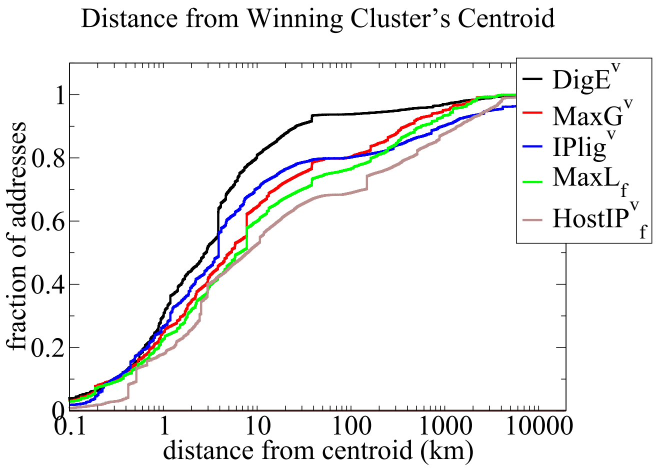


Figure 7: ***CDFs of distances between database-provided geolocations and the winning cluster's centroid. For Digital Envoy, 93% of IP addresses with lat-long coordinates were within 40 km of the winning cluster centroid. The corresponding percentages for MaxMind GeoIP, IPligence, and MaxMind GeoLite were 78%, 78%, and 73%, respectively. HostIP only had a coordinate in the winning cluster for 8% of addresses, 67% of which were within 40 km of the winning cluster's centroid.***

ing to the same wrong answers over time – it assumes that database providers successfully work toward improving the accuracy of their databases over time.

First, we examined how different values of T affect the clusterization. Figure 6 shows that for small thresholds $T \lesssim$ 11.5 km, the majority of clusters contain only two members. Three-member clusters begin to emerge at $T \gtrsim 1$ km and become the majority for $T \gtrsim 11.5$ km where the rate at which new addresses join clusters slows as a function of T.

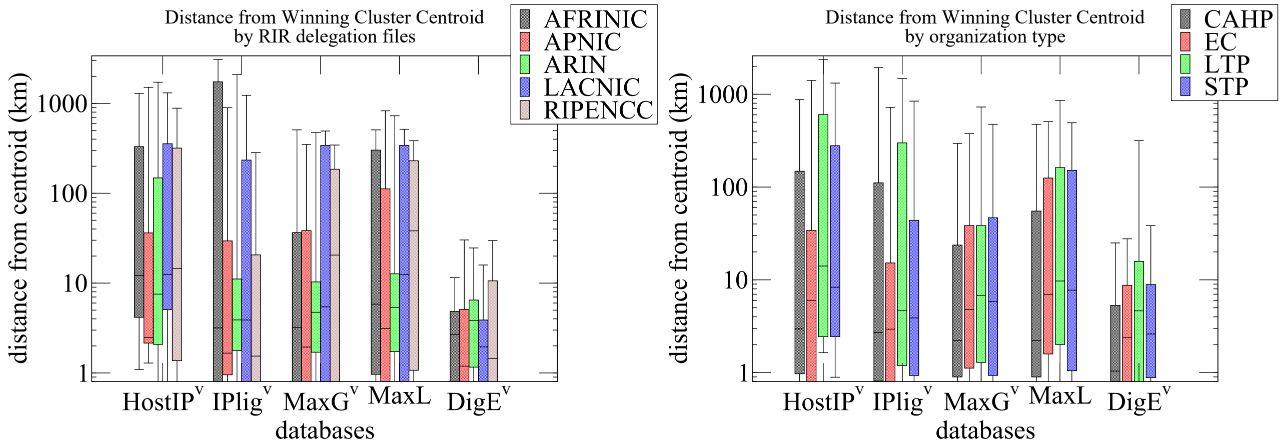
We selected $T = 80$ km for our comparative analysis based on the following considerations. For a given address, a two-member coordinate cluster means that only two (out of possibly up to four) coordinate locations are sufficiently close to each other. In this case, the centroid location is the arithmetic mean of their coordinates. The decline in the number of the two-member clusters (starting at $T \gtrsim$ 11.5 km) indicates that the three-member clusters are being predominantly formed by adding a third point to what was previously a two-member cluster. Correspondingly, the centroid location shifts away from the previous arithmetic mean of the two close locations toward the third, more distant point that merges into this cluster. However, we considered 11.5 km too small to delineate a meaningful geographic region, e.g., a medium size city. The largest threshold which noticeably increased the number of two-member clusters was $T \simeq 77.7$ km. Therefore, we chose $T = 80$ km to include these clusterized addresses. This range is consistent with a size of a large city-metropolis. For comparison, Shavitt *et al.* [38] used ranges of 40 km as "city range" and 500 km as "region range".

Using the selected threshold of 80 km, we were able to find winning clusters for ∼80% of probed addresses and determined the positions of their centroids. Next, for each database, and for each address in a given database, we calculated the distance between the latitude-longitude coordinates provided by this database for this address, and the position of this address' winning centroid (if there was one). Although the lat-long coordinate answers provided by MaxMind GeoLite did not participate in determining the winning clusters, we included its results for comparison with other "voting" databases.

Figure 7 plots the cumulative distribution functions (CDFs) of the calculated distances from centroids. Each CDF is normalized by the number of addresses with winning clusters for this particular database, not as the percentage of all examined addresses. For 93% of the addresses for which Digital Envoy had lat-long coordinates and a winning cluster existed, the coordinates were within 40 km of the winning cluster centroid. The corresponding percentages for MaxMind GeoIP, IPligence, and MaxMind GeoLite were 78%, 78%, and 73%, respectively. Of the 8% of all addresses for which HostIP had a coordinate in the winning cluster, 67% were within 40 km of the winning cluster's centroid.

Figure 8 further elucidates the distributions of distance between a given database's coordinate location and the centroid of the winning cluster for an examined answer. In these plots the thin vertical lines spread between the 10th and 90th percentiles, the ends of color bars mark the 25th and the 75th percentiles, and the black horizontal mark inside each color bar indicates the 50th percentile.

Figure 8(a) presents the breakdown by geographic regions as determined by RIR delegations. All databases located 50% of **APNIC** addresses within 3 km of the centroid, but the variance for those addresses is noticeably larger than the variance for **ARIN** addresses – red bars are longer than green bars for all databases except for Digital Envoy. (Note that Digital Envoy had the tightest variances for all geographic regions, placing 75% of addresses within 10 km and 90% within 20 km of their winning centroids in all geographic regions.) IPligence and both MaxMind databases geolocated 75% of the **ARIN** addresses within 10 km of their winning cluster centroids. In contrast, HostIP located 25% of **ARIN** addresses further than 100 km from the centroid. We found the greatest variance for **LACNIC** addresses (blue bars) – HostIP, IPligence and both MaxMind databases provided coordinates that were further than 300 km from the centroid for 25% of LACNIC addresses. Digital Envoy had 90% of its **LACNIC** addresses within 15 km of the centroid. Although IPligence had 50% of **AFRINIC** addresses' coordinates within 3 km of the centroid, 25% were further than 2600 km.



(a) There greatest consistency across responses from the four databases was for **ARIN** addresses, where there are relatively fewer countries than in the other regions. The greatest inconsistency across databases was found for **LACNIC** addresses.

(b) **CAHP** addresses had the smallest coordinate variance across databases, while **LTP** addresses had the greatest variance.

Figure 8: ***This graphs show the 10th, 25th, 50th, 75th, and 90th percentiles for distance from the centroid for each database, broken down by RIR or AS type.***

We also examined the consistency of IP-to-coordinate mapping by organization type (Figure 8(b)). Looking across all five databases, we found that **CAHP** (Content/Access/Hosting Providers, black bars) addresses had coordinates nearest the centroid, with median distances from the centroid about half that of the next organization type, **EC** (Enterprise Customers, red bars). **LTP** (Large Transit Providers, green bars) had the greatest variance, and median distances greater then any other group type. Only for MaxMind GeoIP was the 75th percentile of its **LTP** distances lower than the 75th percentile of its **STP** (Small Transit Providers, blue bars) coordinates. We surmise that the observed differences in variances reflect more time and effort by the database providers on accurately geolocating edge hosts, which are more likely connected to humans and commerce, and thus more likely to be of type **CAHP** or **EC**.

Table 4 presents pairwise comparisons of coordinate locations between individual databases. Coordinates provided by MaxMind GeoIP and MaxMind GeoLite are the same for at least 75% of their answers. Both these databases and IPligence mostly agree with Digital Envoy, the median differences between provided locations being less than 9 km. HostIP locations are the furthest from other databases, with median distance differences exceeding 100 km between HostIP and IPligence, and exceeding 500 km between HostIP and Digital Envoy. These results are surprising since HostIP data are voluntarily and manually contributed; presumably, these contributors are well aware of their actual locations.

4.3 Comparison with ground truth data

To calibrate the results provided by the participating geolocation databases against the available ground truth datasets described in 3.2.1, we calculated the distances between the identified locations and the known ground truth locations. Figure 9 presents the resulting distance distributions, for the five geolocation services that provide latitude/longitude coordinates of IP addresses and for each ground truth source.

For the PlanetLab IP addresses (Figure 9(a)), all databases placed more than 50% of addresses for which they had coordinates within 10 km of the actual ground truth locations. Digital Envoy had the highest accuracy for PlanetLab IP addresses, with 62% of addresses within 10 km. PlanetLab hosts are primarily located in U.S. academic institutions, typically with a single well-defined geographic location that corresponds to the contact information in the WHOIS database.

We found the largest inconsistencies between geolocation providers when comparing their answers to the French network's ground truth data (Figure 9(b)). MaxMind GeoIP had 58% of addresses within 10 km of the ground truth location, while IPligence had 6%. Only HostIP failed to place all addresses within 700 km (the distance between all points in France and Paris). Interestingly, this database, built on volunteered data from users, located all but 17% of these truly-in-France addresses in Germany. Are those French users intentionally falsifying their location information or are those really German customers using French networks?

The last ground truth dataset, Tier 1's ISP addresses list, includes primarily routers rather than end hosts. Most geolocation providers do not claim to accurately capture geolocations of routers. Digital Envoy's results were the closest to the ground truth for this dataset (Figure 9(c)), placing over 70% of these Tier 1 addresses within 10 km of their actual known locations. For MaxMind GeoIP this percentage is 27% and for IPligence it is 10%. IPligence reaches the 50% mark at 950 km, lagging behind IPHost for this dataset.

Table 4: Pairwise comparison of address coordinate locations. The three numbers represent the 25th, 50th, and 75th percentiles of the distribution of distances between the row and the column database's coordinates for those addresses that had lat-long coordinates in both databases.

	HostIP	IPlig	MaxG	DigE
HostIP$_f$	-	3 / 107 / 1120	9 / 216 / 1140	20 / 511 / 2360
IPligence	3 / 107 / 1120	-	4 / 66 / 685	1 / 9 / 751
MaxMind GeoIP	9 / 216 / 1140	4 / 66 / 685	-	2 / 15 / 318
Digital Envoy	20 / 511 / 2360	1 / 9 / 751	2 / 15 / 318	-

color key 0-49 50-149 150-449 450-1049 1050-

4.4 Calibration vs. RTT

To approximate the expected RTT values, we assumed that the path from the monitor to a target address is roughly a geodesic line, most of the path is made up of fiber, and that cross traffic has little effect on the resulting RTT. Using an index of refraction of 1.538, we estimated the speed of light in fiber as $\sim$195 km/ms.

Figure 10 plots a CDF of the percentage difference between the expected RTT and the observed RTT. We see two distinct groups, the first one including HostIP, IPligence, and Digital Envoy, and the second one consisting of MaxMind GeoIP and MaxMind GeoLite. The first group has 50% of the addresses only 1.3 times the expected RTT vs 1.5 times the "expected" RTT for the second group.

5. SUMMARY

Software77 provided a free downloadable database, with more functionality available for purchase. Although it had .5% fewer addresses than the RIR delegation files, the results were almost indistinguishable. In determining countries for tested IP addresses, it agreed with the majority for 95.5% of addresses. Similar to Cyscape, Software77 has the most inconsistencies with other databases for addresses in European, Middle East and Central Asian regions.

HostIP runs a free, open database populated by volunteer submissions. It has many blocks, which only covered 15.9% of the delegated addresses. Most (63%) of its blocks were in the North American region. It agreed on country with the majority for 94.5% of the addresses. It was within 40 km of the winning centroid for 67% of its addresses. Among all databases in our study, the latitude/longitude locations given by HostIP were farthest from the French Network's addresses, with a median distance of 727 km.

IPligence offered the least expensive commercial database (for geolocation only) that we examined. It had the least agreement on country with the majority, at 94.3%. Most disagreements were in the North American region. It had 78% of its addresses within 40 km of the centroid of distances provided by all databases. Its locations were farthest from the centroid for addresses in the African region. Of the databases we examined it was furthest from the Tier 1 address with an average distance of 1,401 km.

Cyscape divided the address space into the fewest blocks, but still agreed with the majority on country for 98.4% of its addresses. Relative to the other databases examined, it showed the largest percentage of blocks in the RIPE-NCC region. Most of Cyscape's disagreements came from addresses delegated in Europe, Middle East, and Central Asia.

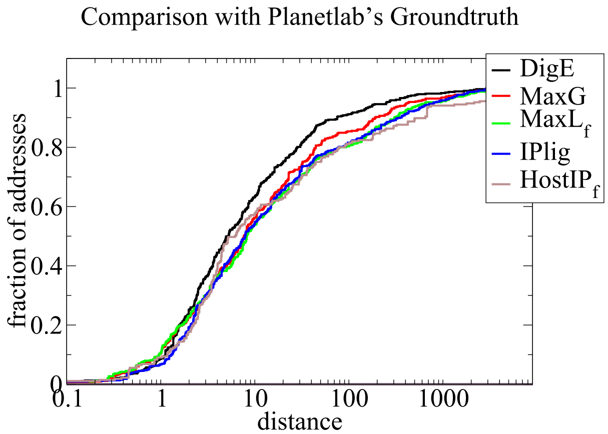
MaxMind GeoIP is the commercial database provided by MaxMind. It agreed with the majority on country for 99.1% of the addresses queried. The majority of disagreements came from the North American region and, disproportionately, among CAHP (Content, Access, Hosting, Providers) address space. Over 78% of its responses were within 40 km of the winning centroid. Its responses were closer to the centriod for Asian addresses than addresses from other regions. The responses were closest to ground truth for the French networks, within 88 km on average.

MaxMind GeoLite, a free database provided by MaxMind, agrees with MaxMind GeoIP's country responses 99.8% of the time and had 75% of its coordinates within 40 km of the centroid. It placed addresses further from all ground truth datasets than MaxMind GeoIP. **IPInfoDB** is a free database with additional functionality on top of **MaxMind GeoLite**.

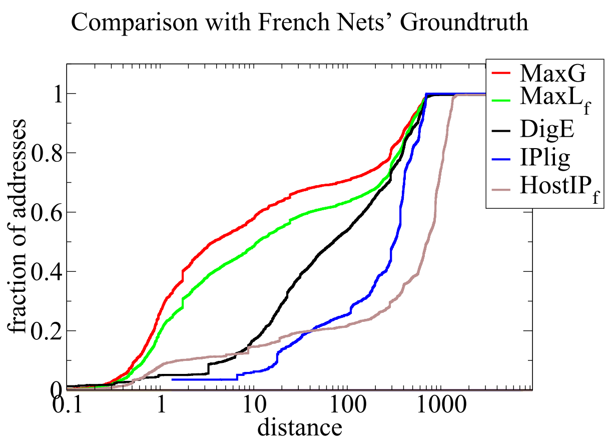
Digital Envoy is a commercial database provider. It agreed on country with the majority for 96.7% of addresses. Its disagreements were disproportionately in Small Transit Providers and from the European, Middle East, and Central Asian address space. 93% of its responses came within 40 km of the centroid. It was nearest on average to two of our ground truth datasets, PlanetLab and the Tier 1 addresses. Digital Envoy's relatively solid performance on router infrastructure IPs is good news for researchers. (Disclosure: For the last seven years, CAIDA has had a Research Agreement to use Digital Envoy's NetAcuity database for many of its public-facing core infrastructure analysis projects.)

6. CONCLUSIONS

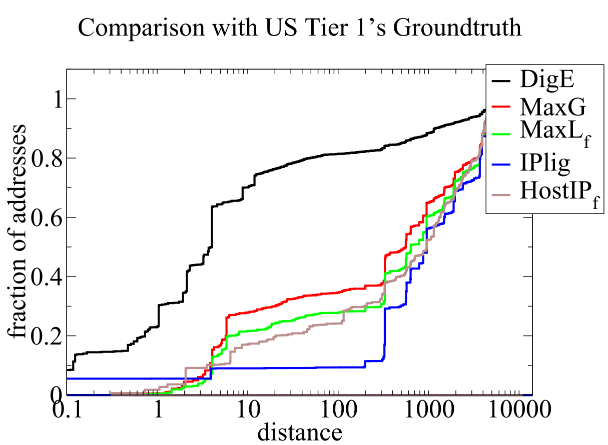
We developed and tested a methodology to support a systematic quantitative comparison of geolocation databases, in terms of how they map IP addresses to countries as well as to actual lat-long coordinates. We analyzed inconsistencies across databases as a function of geographic (RIR) region and organization (Autonomous System) type. To compare country-level granularity, we compared each database



(a) PlanetLab addresses geolocated closest to ground truth in all databases, with the largest spread across databases at 80 km, i.e., over 90% of Digital Envoy's IP addresses but only 79% of HostIP's addresses were within 80 km of ground truth.



(b) The database-provided answers were furthest from ground truth for the French DSL network's IP addresses. MaxMind GeoIP had 70% of addresses within 10 km of ground truth for this data set; IPligence had only 4% within 10 km. HostIP often mapped these French addresses to Germany.



(c) Digital Envoy's mappings had the lowest disparity of all databases against the U.S. Tier 1 ground truth data, capturing the 10% that located at the headquarters and over 50% within 10 km of their known location. IPligence was only correct for IP addresses near the Tier 1's headquarters.

Figure 9: ***Comparison against three ground truth data sets. All databases accurately placed PlanetLab addresses, but Digital Envoy was closest. MaxMind was closer for the French DSL network addresses. Digital Envoy was closest for the Tier 1 router addresses.***

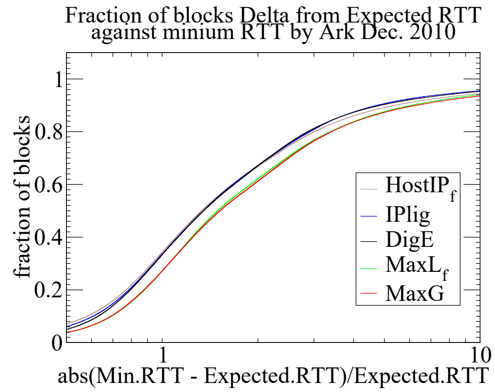


Figure 10: ***Difference between the estimated RTT, a function of the coordinates provided by the database and the speed of light, and the measured RTT. The CDFs of these percentage difference distributions are in two groups: HostIP, IPligence, and Digital Envoy, which have 50% of their addresses with a percentage difference less than 1.3, and the two MaxMind databases, with a median percentage difference less than 1.5.***

against the majority answer across all databases. For lat-long coordinate comparison, we used a centroid-based algorithm to derive a rough consensus across databases, and compared each database's answer to this approximate consensus.

The databases fell into three IP address block granularities: small (Cyscape, Software77, and HostIP) containing 55K-786K blocks; medium (IPligence, IPInfoDB, and MaxMind GeoLite) with 3.1-3.5M blocks; and large (MaxMind GeoIP and Digital Envoy) with 5.7M and 6.1M blocks, respectively. IPInfoDB derives from and is essentially indistinguishable from MaxMind GeoLite in our results.

We found that providers generally agreed on IP-address-to-country mappings. MaxMind GeoLite and MaxMind GeoIP had the highest level of agreement (99.1%) and IPligence had the lowest, with 94.3% of IP addresses agreeing with the majority on country location. The least agreement between any two databases for country mappings was between IPligence and HostIP at 83.8%. Cyscape, Software77, and Digital Envoy showed the greatest disagreement with the majority for RIPE-NCC-assigned addresses.

Digital Envoy and MaxMind GeoIP lat-long coordinates were closest to the winning cluster centroid, with 93% and 78% of addresses within 40 km, respectively. Database response coordinates were closer to the centroid than to other individual database answers, unsurprising since addresses with the largest disagreement with the majority were not part of the winning centroid. Digital Envoy had the shortest median distance to the Tier 1's ground truth router dataset and to PlanetLab's addresses, although all databases fairly accurately geolocated PlanetLab ndoes. MaxMind had the shortest median distance to the French DSL network addresses.

7. REFERENCES

[1] Cyscape's CountryHawk. http://www.cyscape.com/products/chawk/.

[2] data AS:AS Rank. http://as-rank.caida.org/.

[3] data BGP:RIPE NCC's Routing Information Service. http://www.ripe.net/projects/ris/rawdata.html.

[4] data BGP:University of Oregon RouteViews Project. http://www.routeviews.org/.

[5] data RTT:Archipelago Measurement Infrastructure. http://www.caida.org/projects/ark/.

[6] database: Akamai's Edgescape. http://www.ip2location.com/.

[7] database: GeoBytes's GeoServer. http://www.geobytes.com/.

[8] database: Hexasoft's IP2location. http://www.ip2location.com/.

[9] database: Quova. http://www.quova.com/what/products/.

[10] database: Tomas CountryWhoisDB. http://www.tamos.com/products/countrywhois/.

[11] Digital Envoy's Netacuity. http://www.digitalelement.com/.

[12] ground truth:Freebox ADSL networks. http://francois04.free.fr/liste_dslam.php.

[13] ground truth:Planetlab sites. https://www.planet-lab.org/db/pub/sites.php.

[14] HostIP^{free}. http://www.hostip.info/dl/index.html.

[15] IPInfoDB: IP address geolocation SQL database. http://ipinfodb.com/ip_database.php.

[16] IPInfoDB^{free}. http://ipinfodb.com/.

[17] IPligence. http://www.ipligence.com/products/.

[18] MaxMind's GeoIP. http://www.maxmind.com.

[19] MaxMind's GeoLite City^{free}. http://www.maxmind.com/app/geolitecity.

[20] RIR AFRINIC delegations. ftp://ftp.afrinic.net/pub/stats/afrinic/delegated-afrinic-latest.

[21] RIR APNIC delegations. ftp://ftp.apnic.net/pub/stats/apnic/delegated-apnic-latest.

[22] RIR ARIN delegations. ftp://ftp.arin.net/pub/stats/arin/delegated-arin-latest.

[23] RIR LACNIC delegations. ftp://ftp.lacnic.net/pub/stats/lacnic/delegated-lacnic-latest.

[24] RIR RIPE delegations. ftp://ftp.ripe.net/pub/stats/ripnccc/delegated-ripnccc-latest.

[25] Software77^{free}. http://software77.net/geo-ip/.

[26] The DIMES project. http://www.netdimes.org/new/.

[27] AfriNIC and APNIC and ARIN and LACNIC and RIPE. Whois service. http://www.ripe.net/ripencc/pub-services/db/whois/whois.html.

[28] Amogh Dhamdhere and Constantine Dovrolis. Ten Years in the Evolution of the Internet Ecosystem. *ACM Internet Measurement Conference*, Oct. 2008.

[29] M. J. Arif, S. Karunasekera, S. Kulkarni, A. Gunatilaka, and B. Ristic. Internet Host Geolocation Using Maximum Likelihood Estimation Technique. In *AINA '10: Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications*, pages 422–429, Washington, DC, USA, 2010. IEEE Computer Society.

[30] B. Gueye, S. Uhlig, A. Ziviani, and S. Fdida. Leveraging Buffering Delay Estimation for Geolocation of Internet Hosts, 2006.

[31] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint-based Geolocation of Internet Hosts. *IEEE/ACM Trans. Netw.*, 14(6):288–293, 2004.

[32] C. Guo, Y. Liu, W. Shen, H. Wang, Q. Yu, and Y. Zhang. Mining the Web and the Internet for Accurate IP Address Geolocations. In *INFOCOM, IEEE*, pages 2841–2845, Rio de Janeiro, Brazil, 2009. IEEE Computer Society.

[33] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP geolocation using delay and topology measurements. In *IMC '06: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, pages 71–84, New York, NY, USA, 2006. ACM.

[34] S. Laki, P. Mátray, P. Hága, I. Csabai, and G. Vattay. A Model Based Approach for Improving Router Geolocation. *Comput. Netw.*, 54(9):1490–1501, 2010.

[35] P. Mockapetris. Domain Names - Concepts and Facilities, November 1987. Internet Standard 0013 (RFCs 1034, 1035).

[36] D. Moore, R. Periakaruppan, J. Donohoe, and K. Claffy. Where in the World is netgeo.caida.org? In *INET'00: Proceedings of the 10th Annual Internet Society Conference*, 2000.

[37] J. Muir and P. C. V. Oorschot. Internet Geolocation: Evasion and Counterevasion. *ACM Computing Surveys*, 42(1), 2009.

[38] Y. Shavitt and N. Zilberman. A Study of Geolocation Databases. *ArXiv e-prints*, May 2010.

[39] S. S. Siwpersad, B. Gueye, and S. Uhlig. Assessing the geographic resolution of exhaustive tabulation for geolocating internet hosts. In *PAM Proc.*, 2008.

[40] B. Wong, I. Stoyanov, and E. Gun Sirer. "Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts". In *Proceedings of the Symposium on Networked System Design and Implementation*, Cambridge, MA, USA, April 2007.