Location Accuracy of Commercial IP Address Geolocation Databases

Location Accuracy of Commercial IP Address Geolocation Databases

Dan Komosny

Brno University of Technology, Department of Telecommunications, Technicka 12, 616 00 Brno, Czech Republic
e-mail: komosny@feec.vutbr.cz

Miroslav Voznak

VSB-TU Ostrava, Department of Telecommunications, 17 listopadu 15/2172, 708 33 Ostrava, Czech Republic
e-mail: miroslav.voznak@vsb.cz

Saeed Ur Rehman

Auckland University of Technology, School of Engineering, Computer and Mathematical Sciences Private Bag 92006, Auckland 1142, New Zealand; e-mail: saeed.rehman@aut.ac.nz

Corresponding author: komosny@feec.vutbr.cz

This paper deals with finding the geographical location of Internet nodes remotely with no need to communicate with the nodes located (client-independently). IP geolocation is used in a number of areas, such as content personalisation, on-line fraud prevention and detection, and digital media law enforcement. One of the main concerns when studying the accuracy of client-independent geolocation is the groundtruth dataset. As we show in the related work, the used groundtruth influences the results a lot. We construct an error-free groundtruth dataset consisting of nodes with GPS-precise locations. We also record the country, region, city, and ISP for each groundtruth node. Using the created groundtruth, we study the accuracy of eight IP location databases in a number of scenarios, such as effect of city area and population, effect of ISP assignment, and number of not-returned locations.

KEYWORDS: location, geolocation, IP address, groundtruth, accuracy, city, database, MaxMind, DB-IP, IP-2Location, ipinfo, Skyhook, Neustar, Eurek, GeoBytes.

1. Introduction

Geographical location of Internet devices can be obtained with an assistance of the device located (client-dependent) or without its assistance (client-independent). Client-dependent methods use

technologies such as GPS, accelerometers, and triangulation in WiFi or cellular mobile networks. These methods rely on specific properties or features of the devices located. Typically, location accuracy is within tens of meters. On the other hand, client-independent methods are used to locate any Internet device and they do not require any additional properties or features in the devices located. Several methods are used, such as IP geolocation databases. Accuracy of client-independent geolocation is lower, typically at city level.

In this paper we deal with client-independent geolocation that is used for a broad variety of Internet services and applications, including social networks (detection of virtual identity misuse or username/password sharing), web-based services (detection of suspicious logins), e-shops (detection of on-line credit card frauds), banking (prevention of phishing attacks), and electronic content distribution (enforcing territory restriction given by digital media laws).

The contribution of this paper is the following:

1 Based on an inconsistency of the location accuracy results presented in the related work (described in Section 4), we construct an error-free groundtruth dataset. It consists of Internet nodes with known GPS-precise locations. The dataset guarantees the avoidance of wrong location accuracy results.

2 The related work typically depends on the groundtruth nodes coming from large cities. Our groundtruth dataset covers all types of cities, from very small to very large. This allows us to study more properties that influence the location accuracy, such as the effect of city area and population.

3 When locating nodes that belong to the same Internet service provider (ISP), we observe that the results show the same or similar locations in spite of very different real (correct) positions. We particularly study the change of location performance from this ISP node assignment point of view.

The paper is structured as follows: the next section defines the problem that we address. We describe why low accuracy is reached with client-independent geolocation. We give an example of locations provided by several location databases and demonstrate the location error. Section 3 presents the location accuracy as reported by the database vendors. In Section 4 we survey the related work that deals with client-independent location accuracy. Section 5 describes the method for error-free construction of the groundtruth dataset. In Section 6 we present and discuss our location accuracy evaluation. In Section 7 we summarize the results.

2. Current problems of IP geolocation

In this section, we discuss the current problems of finding the geographical location of the Internet devices by their IP addresses.

The use of the IP address space is controlled by IANA (Internet Assignment Numbers Authority). IANA allocates the major segments of IP addresses to five regional registers (RIRs) – AFRINIC (Africa), APNIC (Asia/Pacific), ARIN (North America), LACNIC (Latin America), and RIPE NCC (Europe, the Middle East, and Central Asia). The regional registers further allocate IP address segments ISPs. Such allocation can be direct or through two types of intermediary entities – national internet registry (NIR) and local internet registry (LIR). The records of the allocated IP address segments are stored in a database that is managed by a regional register, as shown in Figure 1. Along with the IP allocation records, the registers maintain contact information of the organizations with the assigned IP addresses. The stored contacts provide a way to locate IP devices in some extent. However, there are no official rules for filling the contact information by the organizations, and thus the provided location information can lead to wrong results. Next major concern is that the IP addresses that fall into one allocation segment can be distributed on a large geographical area depending on the type and size of the organization. A good example are ISPs that operate at the national level or organizations with branches at different locations.

The domain names (DNS) can also indicate the location of IP nodes, however, there are no rules for geographical naming despite some standardization

Figure 1

IP address allocation
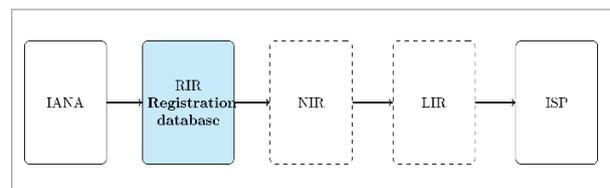
efforts [2]. Some ISPs use internal geographical naming schemes for their networking devices and such information can be used for IP geolocation as described in [24]. The use of domain names for IP geolocation was particularly described in [6, 26, 5]. An enhancement of DNS defines a new LOC record which stores latitude, longitude, and altitude for domain names. However, this enhancement gives a poor location efficiency (high number of not-returned locations) and large location errors [14].

Network measurement is also used for IP geolocation [19, 11, 4]. Measurement-based geolocation works with a positive correlation between communication latency and geographical distance [23]. The latency is measured from a set of servers with known location to the IP address located. The results are converted to maximal geographical distances from the servers to the IP address. These geographical distances delimit the area where the IP address is located. There are several known methods that use this approach, such as Constraint-Based Geolocation [9], Octant [29], Spotter [20], or Topology-based Geolocation [13]. A disadvantage of measurement-based geolocation is the latency instability which leads to false location estimations [18, 28]. Other known problems are long location times [22, 21].

Database-based IP geolocation defines blocks of continuous IP addresses and stores location information for them. These IP blocks may be smaller than the IANA allocation segments and thus provide a better location accuracy. There are two schemes to obtain the location information to be linked with the IP blocks: top-down and bottom-up. These sources of locations are shown in Figure 2. The top-down scheme uses location information available through Internet resources, such as crawling web pages [7, 3, 1, 10] and measuring the network [17]. The bottom-up scheme uses locations that are collected by external resources, such as GPS or WiFi network scanning.

Figure 2

Geolocation database filling schemes
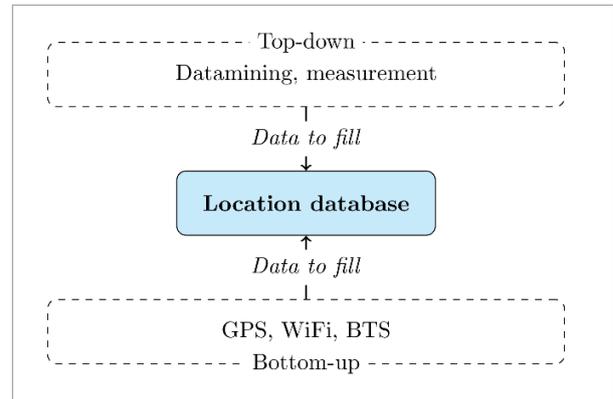
3. Claimed accuracy by location service vendors

In this section, we study the claimed location accuracy by major geolocation database providers. We summarize the found claimed accuracy in Table 1. The databases typically return geographical coordinates (latitude and longitude), country, region, and city. The Skyhook database 'Hyperlocal IP Pro' differs from others by returning region and city only when the estimation reaches a certain level of trustworthiness.

Table 1

Claimed accuracy by vendors

Vendor/database	Country [%]	City [%]	IPv4	IPv6
MaxMind/GeoIP2 Precision	84	40	100 %	YES – N/A
DB-IP/IP address to location + ISP	N/A	N/A	7 mil.	YES – 586,718
IP2Location/DB24	N/A	77	14 mil.	YES – N/A
Neustar/where	99.9	N/A	100 %	YES – 100 %
Eurek/professional edition	N/A	N/A	100 %	NO
Geobytes/Geo IP Location	97	75	98 %	NO

MaxMind publishes the accuracy data for 23 countries (www.maxmind.com/en/geoip2-city-database-accuracy). For the purpose of comparison at the country level, we use the maximum location error of 250 km to evaluate the result as correct. Based on this range, 4 % of the location queries are reported to be resolved incorrectly, and 12 % of location queries are reported to be unresolved (not-returned) at the country level. For the city level, 48 % of the location queries give an incorrect city, and 11 % the location queries are unresolved. DB-IP does not publish any data on location accuracy, only the number of IP address space covered. IP2Location publishes a comprehensive location accuracy data for 250 countries (www.ip2location.com/data-accuracy). The published data cover only the city coverage (the country level is not included). For the purpose of comparison at the city level, we use the maximum error of 50 miles. Neustar does not publish any data on accuracy of their geolocation services. However, we found that the accuracy of the databases was evaluated by Pricewater- house Coopers (www.neustar.biz/resources/product-literature/neustar-ip-intelligence-pwc-audit). The result is 99.9 % accuracy for the country level. Neustar claims to cover all of the IP address space. Eurek does not provide any location accuracy information about their products. The only information provided is that it covers the whole IPv4 address space. Geo-Bytes provides some basic data about their accuracy (www.geobytes.com/faq/). It claims to resolve 98 % of IP addresses with accuracy of 97 % at the country level. Another information published is that 80 % of returned locations are within the maximum error distance of 100 km and 75 % of locations are within the maximum location error of 50 km. ipinfo does not publish any accuracy related information. The same holds for Skyhook.

4. Related work

Comparing data from the related work is difficult due to the use of different evaluation techniques. An example is the use of different distance thresholds for the city-level accuracy. In the related work it varies from 40 to 100 km. Therefore, we use a different measure to compare the related work. We work with the independent cumulative probabilities of locations within a maximal error of 50, 100, 150, and 250 km.

Shavitt and Zilberman [25] use a groundtruth dataset that is based on an algorithm which groups IP addresses into virtual Points of Interest (PoPs). The algorithm discovers the sets of the routers at the same location. For this purpose, they use latency measurements and topology discovery. The accuracy of the results depends on the PoPs identified locations. Six major geolocation databases are evaluated: MaxMind, IP2Location, IPligence, HostIP, Netaculity, and Geobytes. We summarize the results in Table 2. As the source of the accuracy data, we use the cumulative probability function showing the database location deviation from the locations of the identified groundtruth PoPs.

Table 2

Cumulative percentage of estimated locations within maximum location error [km] – created from source [25] [%]

Vendor	<50	<100	<150	<250
MaxMind	68	73	76	78
IP2Location	62	65	66	68
IPligence	73	75	76	78
HostIP	37	39	42	45
Netaculity	45	49	50	54
Geobytes	33	35	40	45

Triukose et al. [27] focus on IP geolocation of mobile devices. They discuss the use of network address translation and how it affects IP geolocation. They study the accuracy of the public and private IP addresses separately. In some cases, they observe large location errors at the scale of inter-continental distances. We particularly study such large errors in section 6.6. We summarize the results of their work in Table 3. Two databases are used: MaxMind and

Table 3

Cumulative percentage of estimated locations within maximum location error [km] – created from source [27] [%]

Vendor	<50	<100	<150	<250
IP2Location	28	34	47	57
MaxMind	15	22	32	45

IP2Location (DB11.LITE). The table shows the values for the public IP addresses studied. The private addresses studied give larger values up to a maximum error of 400 km. After this value, the results are similar with no significant difference. We note that a cellular network is used in this study and such networks give worse results compared to general Internet networks.

Huffaker et al. [12] evaluate eight databases. They point out the absence of a substantial groundtruth dataset. They propose a method to evaluate the accuracy by a centroid-based algorithm working with majority of location votes from the databases. The location accuracy in a form of cumulative probability is only provided for five databases: MaxMind Geo, MaxMind Lite, IPligence, Digital Envoy, and HostIP. For the data shown in Table 4, we use the location accuracy for the PlanetLab groundtruth dataset. The results show a better location accuracy most probably because of the use of the groundtruth dataset, which was created by using the location voting algorithm. Another reason is that the PlanetLab groundtruth dataset is strongly oriented towards the major cities [15, 16].

Table 4

Cumulative percentage of estimated locations within maximum location error [km] – created from source [12] [%]

Vendor	<50	<100	<150	<250
MaxMind	81	85	89	92
Digital Envoy	88	91	93	96
IPligence	78	82	85	89
HostIP	76	82	83	87

Poese et al. [24] focus on the differences between the claimed accuracy by the database vendors and the real-case location accuracy. The accuracy is evaluated by using a groundtruth dataset that was obtained via a large European ISP. They create PoPs based on the network prefixes. The locations of the groundtruth nodes are obtained by using an internal naming scheme of the ISP. The databases studied are HostIP, IP2Location, InfoDB, MaxMind, and Software77. Table 5 shows the results for IP address blocks that are smaller than the groundtruth ISP prefixes.

Table 5

Cumulative percentage of estimated locations within maximum location error [km] – created from source [24] [%]

Vendor	<50	<100	<150	<250
InfoDB	50	63	70	80
MaxMind	35	42	50	63
IP2Location	10	15	18	20

Other papers that deal with database-based IP geolocation are [30, 8]. These papers do not give any results for the cumulative maximum location error probabilities, and they focus on different related topics, such as the covered IP address space.

By summarizing the related work, we show that there are very large differences in the results. The related work shows that:

1 from 10 to 88 % of the nodes can be located within 50 km range,

2 15-91 % of the nodes can be located within range of 100 km,

3 18-93 % of the nodes can be located within range of 150 km,

4 and 20-96 % of the nodes can be located within range of 250 km.

The worst results were achieved when one ISP-based groundtruth dataset was used. The second worst results were achieved when mobile devices were located. The best results were obtained when using a groundtruth dataset that was obtained from the centroid-based algorithm based on location votes from the databases.

5. Construction of error-free groundtruth dataset

One of the main concerns when evaluating location accuracy is the groundtruth dataset. As we show in the related work, the used groundtruth influences the location accuracy results a lot. Being aware of this problem, we construct an error-free groundtruth dataset. We collected the original groundtruth public IP addresses by using a developed mobile application.

The groundtruth geographical locations were obtained by using in-built GPS in the used mobile devices. We also recorded the country, region, city, and ISP for each groundtruth IP address.

The groundtruth construction is as follows:

1 In order to keep the dataset free of the problems that described in the related work, we strictly followed the rule to use only one node per ISP in a city (i.e. there is only one node which belong to an ISP in a city). This restricted the size of the original dataset a lot. On the other hand, it assured the proper distribution of the nodes and the location accuracy results are not influenced by repeating the same or similar locations.

2 We additionally filtered the dataset to store only the nodes that belong to same ISPs and, at the same time, situated in at least 10 different cities for each ISP involved. The reason is that the location databases give the same or a small set of locations for the same-ISP nodes that are correctly located in many different places [27]. An example distribution of the nodes that belong to such an ISP is shown in Figure 3.

3 We aimed to cover a variation of cities in the groundtruth. The cities covered are very small to very large, i.e. we did not focus only on major cities. The city population and area of the groundtruth nodes is shown in Figures 4 and 5. The median value for the city population is 16 925. The median value for the city area is 50 km^2.

Figure 3

Example of same-ISP nodes used in groundtruth dataset
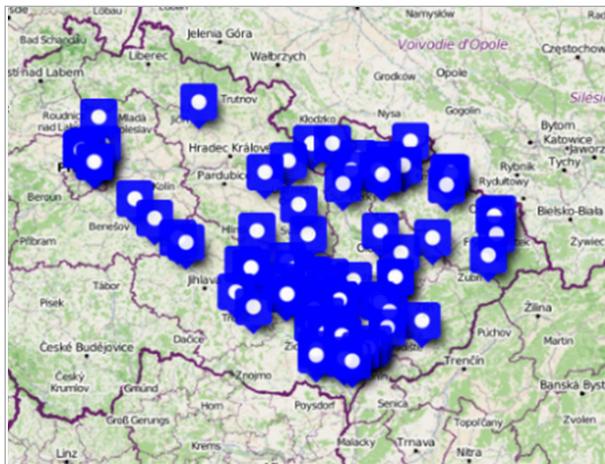

Figure 4

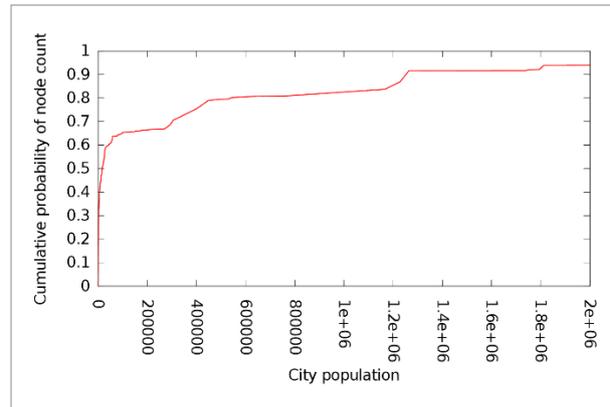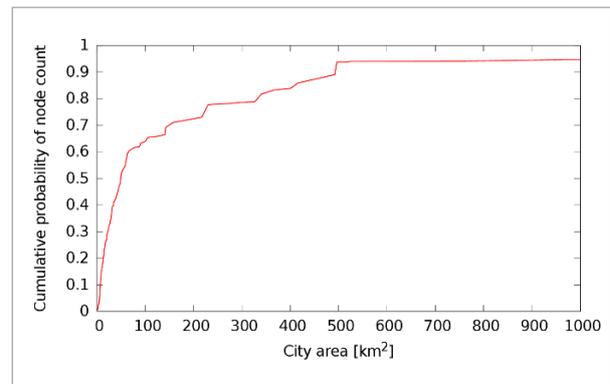Groundtruth nodes in cities with different population

Figure 5

Groundtruth nodes in cities with different area

After applying the proposed method to the original dataset, we obtained about 700 error-free groundtruth nodes. The trusted nodes were situated in 16 countries, 52 regions, and 270 cities. They were assigned to 319 ISPs.

# 6. Evaluation of location accuracy

6.1 Relative location accuracy

By using the groundtruth dataset constructed, we evaluated relative location accuracy in terms of the correctly estimated countries, regions, and cities. The results are shown in Figure 6. During this evaluation, we faced a particular problem of the place names which are sometimes different in English and in the

local language. Some of the databases return the English form, but the others keep the local name. For a proper evaluation, we stored the English forms of place names in the groundtruth dataset. If a database returned a place name in the local language, we found its English form for a proper evaluation.

Figure 6

Relative location accuracy
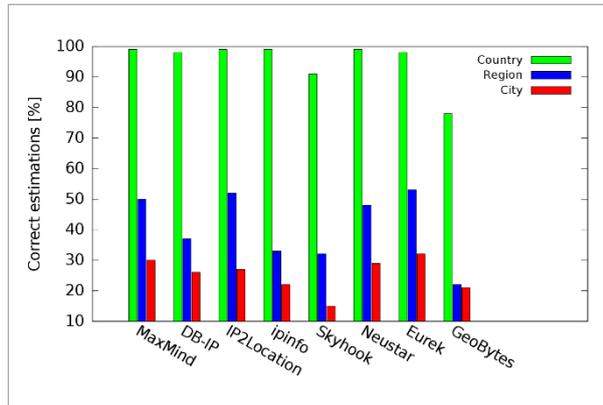
The results show that the majority of the databases achieve nearly 100 % accuracy at the country level. The worst accuracy is achieved by the database Geo-Bytes which is around 80 %. There are much worse results for the region and city estimations. The best databases return around half of the estimations correct at the region level. The worst database is again GeoBytes with about 20 % of the estimated correct regions. At the city level, the best databases give about 30 % of the estimations correct. We notice a significant drop for the Skyhook database which returns only about 15 % of the cities correctly. We, however, note that this database returns a city only when there is a high probability of a correct match.

6.2 Absolute location accuracy

We evaluated absolute location accuracy as the geographical distance between the estimated and correct coordinates. We noticed great differences between absolute and relative location accuracy. A good example is the Skyhook database. Figure 7 shows cumulative probabilities of maximal location errors for each database. The most accurate database is Skyhook. However, regarding relative accuracy (Figure 6), Skyhook gives the worst result at the city level. The reason behind this inconsistency is that Skyhook returns

Figure 7

Cumulative probability of absolute location error
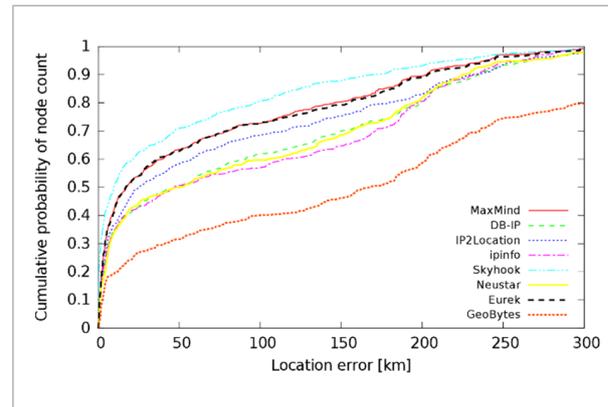
a city only when the estimation is believed to be correct, otherwise the city returned is null.

The most significant differences in absolute location error between all the databases are around a maximal error of 150 km. The best database gives about 90 % of the locations within this error range (Skyhook) while the worst returns only 45 % of the results within this error range (GeoBytes). This difference becomes smaller for large-errors that we separately study in section 6.6. All the databases return almost all locations within maximum error range of 300 km except the database GeoBytes.

Table 6 provides specific numbers on absolute location accuracy. The standard deviation shows large variation of the location errors for the databases DB-IP and GeoBytes.

Table 6

Absolute location error details [km]

Database	Mean	Std Dev	1st q.	Median	3rd q.
MaxMind	64	86	3	18	113
DB-IP	142	775	5	50	181
IP2Location	91	290	4	26	146
ipinfo	95	98	5	47	187
Skyhook	50	76	1	10	70
Neustar	93	100	6	51	181
Eurek	69	112	3	17	119
Geobytes	657	1768	21	168	261

6.3 Effect of city area and population

Figures 8 and 9 show the relative location error change for different city areas and populations. The figures show the regression lines of the median location error values. Both figures indicate better accuracy with increasing city area and population. The reason for these accuracy changes is the source of location information that come from traffic measurement (more populated places generate more Internet traffic). This also holds for crawling and datamining the web servers for location data.

Figure 8

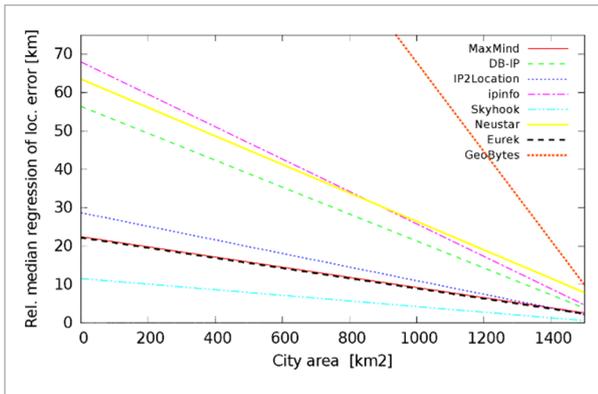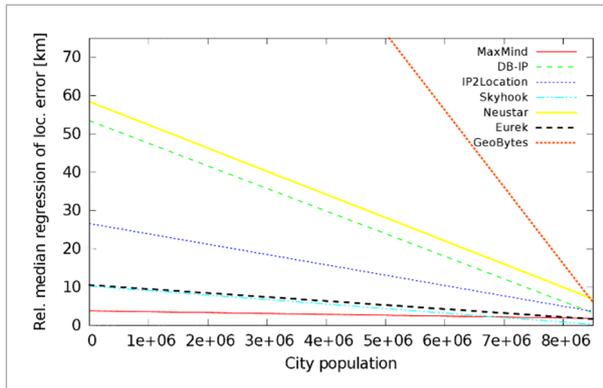Location error change with different city area

Figure 9

Location error change with different city population


6.4 Effect of ISP assignment

The databases typically return the same or a small set of locations for the nodes assigned to the same ISP. We particularly study this phenomenon in Figure 10. The figure shows the cumulative probability for the same-ISP nodes located within a maximum location error. The gap between the location error probabilities for the databases is smaller compared to the general case shown in Figure 7. The graph also shows generally worse cumulative probabilities. The databases again locate all the nodes within a maximum error distance of 300 km, except the database GeoBytes.

Figure 10

Absolute location error of nodes belonging to same-ISP
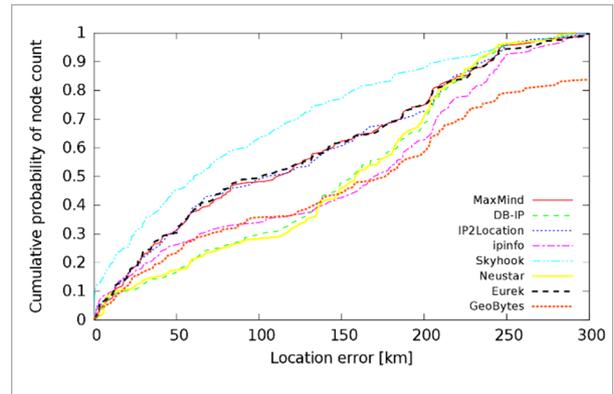
The specific numbers of this evaluation are shown in Table 7. The database with the lowest median location error is again Skyhook. However, the median value for the database Skyhook is about 50 km worse, compared to the general multi-ISP scenario.

Table 7

Location accuracy details for nodes belonging to same ISP [km]

Database	Mean	Std Dev	1st q.	Median	3rd q.
MaxMind	118	86	35	112	201
DB-IP	144	78	78	158	206
IP2Location	118	88	38	107	205
ipinfo	145	91	44	171	215
Skyhook	87	81	21	61	138
Neustar	145	77	80	159	204
Eurek	117	88	35	102	201
Geobytes	483	1142	54	177	236

We also study relative errors for the same-ISP nodes. Figure 11 shows the correctly estimated cities. We conclude that the databases estimate approximately one quarter (on average) of the cities correctly when the same ISPs are used compared to multiple ISPs usage.

Figure 11

Correct city estimations for nodes belonging to same-ISP
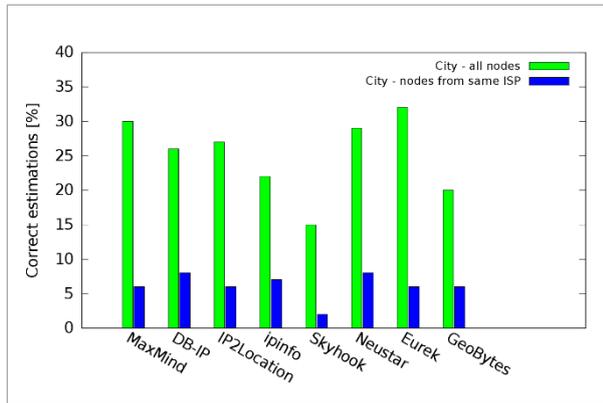
6.5 Unresolved locations

The previous results show that the databases do not resolve (do not return) a location for every node requested. Figure 12 shows the not-returned relative locations (country, region, and city) and the coordinates. The results indicate that the best database in terms of the resolved locations is DB-IP (100 % locations returned), IP2Location IP (100 % locations returned), Neustar, and GeoBytes. We note that Skyhook returns

Figure 12

Unresolved locations
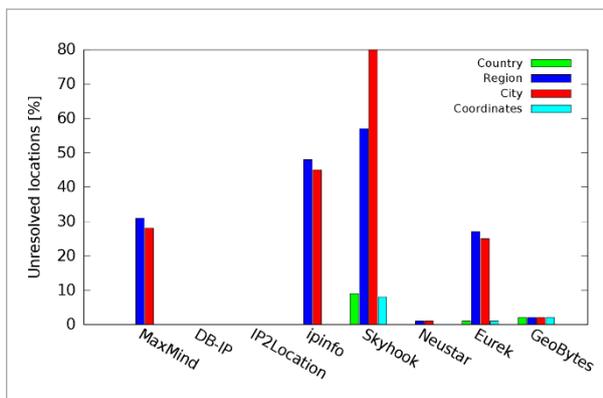
a relative location only when a certain level of trustworthiness is reached. We also note that these numbers do not indicate the location accuracy as some databases return the capital city or the centre of a country when a better estimation is not known. This is better seen when Figures 12 and 6 are compared.
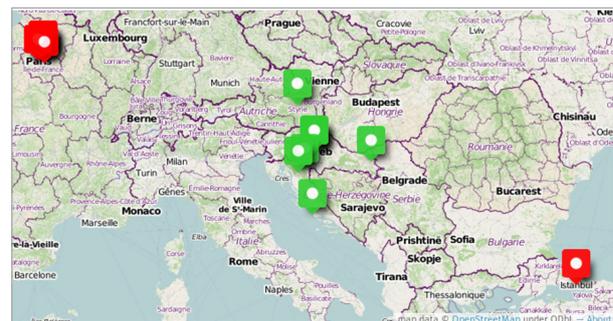
6.6 Large errors

A common problem of IP geolocation is that some locations are returned with a great error. The previous observation shows the percentages of such large-errors in Figure 7. It shows that there is a percentage of location errors over 200 km for each database. The database GeoBytes shows the worst results for location errors over 200 km.

We demonstrate the problem of large errors in Figure 13. The green/bright marks show the locations estimated with some large errors. Additionally, one of the red/dark marks shows an estimated location with an extremely large error. The node is correctly situated in Paris, France but the estimated location points to Istanbul, Turkey.

Figure 13

Example of very large location errors

Differences between the databases for such large errors may be seen better by plotting a matrix of the mutual differences for the last decile of location errors. The heatmap in Figure 14 indicates that some of the database pairs (MaxMind and Eurek, ipinfo and DB-IP) have very small location error differences. On the other hand, GeoBytes seems to be quite unreliable as it has large differences from the other databases reaching a range of hundreds of kilometers. Therefore this database should be excluded from the technique that uses the centre of gravity of multiple location results as the final estimated location.

Figure 14

Databases differences for large location errors (last decile)
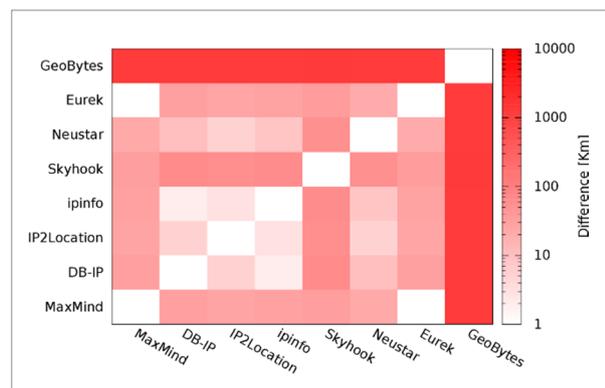
7. Conclusions

The accuracy of client-independent IP location was high at the country level where it reached almost 100 % of correct estimations (except the database GeoBytes). In this case, IP geolocation databases can be trusted. The region-level accuracy was of a worse value, best databases gave around 50 % of correct estimations. The best city-level accuracy was around 30 %. Considering the error distances between the estimated and correct coordinates, the best median error was 10 km for the database Skyhook. We note that all these values were obtained for a high number of small cities involved in the dataset (the median city population was 17 000).

The city population and area have a significant impact on location accuracy. Furthermore, node association to the ISPs also have a significant impact on location accuracy. Multiple nodes that were assigned to the same ISP but in different cities were located less accurately than the nodes that were assigned to different ISPs. There was a 50 km worse median location error for the best performance database Skyhook. Also, the databases estimated approximately one quarter (on average) of the cities correctly when the same ISPs were used compared to multiple ISPs usage.

The percentages of returned locations do not correlate with location accuracy. This was particularly apparent for the correctly estimated and unresolved cities.

Finally, we noticed great differences between absolute and relative location accuracy. Some databases may return a city or region only when there is a high probability of a correct match so as not to confuse the users.

Acknowledgments

References

1. Backstrom, L., Sun, E., Marlow, C. Find Me if You Can: Improving Geographical Prediction with Social and Spatial Proximity. Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, 61-70. https://doi.org/10.1145/1772690.1772698

2. Balakrishnan, H., Lakshminarayanan, K., Ratnasamy, S., Shenker, S., Stoica, I., Walfish, M. A Layered Naming Architecture for the Internet. Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM, 2004, 343-352. https://doi.org/10.1145/1015467.1015505

3. Bo, H., Cook, P., Baldwin, T. Geolocation Prediction in Social Media Data by Finding Location Indicative Words. Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), 2012, 1045-1062.

4. Burget, R., Komosny, D., Kathiravelu, G. Topology Aware Feedback Transmission for Real-Time Control Protocol. Journal of Network and Computer Applications, 2012, 35(2), 723-730. https://doi.org/10.1016/j.jnca.2011.11.005

5. Fioreze, T., Heijenk, G. Extending DNS to Support Geocasting Towards VANETS: A Proposal. IEEE Vehicular Networking Conference (VNC) 2010, 2010, 271-277. https://doi.org/10.1109/VNC.2010.5698258

6. Fioreze, T., Heijenk, G. Extending the Domain Name System (DNS) to Provide Geographical Addressing Towards Vehicular Ad-Hoc Networks (VANETs). IEEE

Vehicular Networking Conference (VNC) 2011, 2011, 70-77. https://doi.org/10.1109/VNC.2011.6117126

7. Graham, M., Hale, S., Gaffney, D. Where in the World Are You? Geolocation and Language Identification in Twitter. The Professional Geographer, 2014, 66(4), 568-578. https://doi.org/10.1080/00330124.2014.907699

8. Gueye, B., Uhlig, S., Fdida, S. Investigating the Imprecision of IP Block-Based Geolocation. In: Passive and Active Network Measurement, Lecture Notes in Computer Science, Springer, 2007, 4427, 237-240. https://doi.org/10.1007/978-3-540-71617-4_26

9. Gueye, B., Ziviani, A., Crovella, M., Fdida, S. Constraint-Based Geolocation of Internet Hosts. IEEE/ACM Transactions on Networking, 2006, 14(6), 1219-1232. https://doi.org/10.1109/TNET.2006.886332

10. Guo, Ch., Liu, Y., Shen, W., Wang, H., Yu, Q., Zhang, Y. Mining the Web and the Internet for Accurate IP Address Geolocations. IEEE INFOCOM 2009, 2009, 2841-2845.

11. Hegr, T., Bohac, L., Kocur, Z., Voznak, M., Chlumsky, P. Methodology of the Direct Measurement of the Switching Latency. Przeglad Elektrotechniczny, 2013, 89(7), 59-63.

12. Huffaker, B., Fomenkov, M., Claffy, K. Geocompare: a Comparison of Public and Commercial Geolocation Databases. CAIDA Tech Report, University of California, 2011.

13. Katz-Bassett, E., John, P., Krishnamurthy, A., Wetherall, D., Anderson, T., Chawathe, Y. Towards IP Geolocation Using Delay and Topology Measurements. Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, ACM, 2006, 71-84. https://doi.org/10.1145/1177080.1177090

14. Komosny, D., Balej, J., Sathu, H., Shukla, R., Dolezel, P., Simek, M. Cable Length Based Geolocalisation. Przeglad Elektrotechniczny, 2012, 88(7a), 26-32.

15. Komosny, D., Pang, S., Pruzinsky, J., Ilko, P., Polasek, J. PlanetLab Europe as Geographically-Distributed Testbed for Software Development and Evaluation. Advances in Electrical and Electronic Engineering, 2015, 13(2), 137-146. https://doi.org/10.15598/aeee.v13i2.1245

16. Komosny, D., Pruzinsky, J., Ilko, P., Polasek, J., Masek, P., Kocatepe, O. On Geographic Coordinates of PlanetLab Europe. IEEE 38th International Conference on Telecommunications and Signal Processing (TSP), 2015, 642-646. https://doi.org/10.1109/TSP.2015.7296342

17. Komosny, D., Voznak, M., Bezzateev, S., Kathiravelu, G. The Use of European Internet Communication Prop-

erties for IP Geolocation. Information Technology and Control, 2016, 45(1), 77-85. https://doi.org/10.5755/j01.itc.45.1.11062

18. Komosny, D., Voznak, M., Kathiravelu, G., Sathu, H. Estimation of Internet Node Location by Latency Measurements – The Underestimation Problem. Information Technology and Control, 2015, 44(3), 279-286. https://doi.org/10.5755/j01.itc.44.3.8353

19. Krajsa, O., Fojtova, L. RTT Measurement and Its Dependence on the Real Geographical Distance. IEEE 34th International Conference on Telecommunications and Signal Processing, 2011, 231-234. https://doi.org/10.1109/TSP.2011.6043737

20. Laki, S., Matray, P., Haga, P., Sebok, T., Csabai, I., Vattay, G. Spotter: A Model Based Active Geolocation Service. IEEE Conference on Computer Communications, 2011, 3173-3181. https://doi.org/10.1109/INFCOM.2011.5935165

21. Moravek, P., Komosny, D., Burget, R., Sveda, J., Handl, T., Jarosova L. Study and Performance of Localization Methods in IP Based Networks: Vivaldi Algorithm. Journal of Network and Computer Applications, 2011, 34(1), 351-367. https://doi.org/10.1016/j.jnca.2010.06.014

22. Moravek, P., Komosny, D., Simek, M., Sveda, J., Handl, T. Vivaldi and Other Localization Methods. TSP 2009: 32nd International Conference on Telecommunications and Signal Processing, Asszisztencia Congress Bureau, 2009, 214-218.

23. Percacci, R., Vespignani, A. Scale-Free Behavior of the Internet Global Performance. The European Physical Journal B – Condensed Matter and Complex Systems, 2003, 32(4), 411-414. https://doi.org/10.1140/epjb/e2003-00123-6

24. Poese, I., Uhlig, S., Kaafar, M., Donnet, B., Gueye, B. IP Geolocation Databases: Unreliable? ACM SIGCOMM Computer Communication Review, 2011, 41, 53-56. https://doi.org/10.1145/1971162.1971171

25. Shavitt, Y., Zilberman, N. A Geolocation Databases Study. IEEE Journal on Selected Areas in Communications, 2011, 29(10), 2044-2056. https://doi.org/10.1109/JSAC.2011.111214

26. Sousa, A., Costa, A., Santos, A., Meneses, F., Nicolau, M. Using DNS to Establish a Localization Service. IEEE International Conference on Indoor Positioning and Indoor Navigation 2014, 2014, 385-392. https://doi.org/10.1109/IPIN.2014.7275506

27. Triukose, S., Ardon, S., Mahanti, A., Seth, A. Geolocating IP Addresses in Cellular Data Networks. Passive

and Active Measurement, 13th International Conference, Springer, 2012, 7192, 115-167. https://doi.org/10.1007/978-3-642-28537-0_16

28. Voznak, M., Rozhon, J., Ilgin, H. Network Delay Variation Model Consisting of Sources with Poisson's Probability Distribution. Proceedings of the 15th WSEAS International Conference on Computers, WSEAS, 2011, 240-244.

29. Wong, B., Stoyanov, I., Sirer, E. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. Proceedings of the 4th USENIX Conference on Networked Systems Design & Implementation, USENIX Association, 2007, 313-326.

30. Zander, S. How Accurate is IP Geolocation Based on IP Allocation Data? Technical Report 120524A, Swinburne University of Technology, 2012.

Summary / Santrauka

This paper deals with finding the geographical location of Internet nodes remotely with no need to communicate with the nodes located (client-independently). IP geolocation is used in a number of areas, such as content personalisation, on-line fraud prevention and detection, and digital media law enforcement. One of the main concerns when studying the accuracy of client-independent geolocation is the groundtruth dataset. As we show in the related work, the used groundtruth influences the results a lot. We construct an error-free groundtruth dataset consisting of nodes with GPS-precise locations. We also record the country, region, city, and ISP for each groundtruth node. Using the created groundtruth, we study the accuracy of eight IP location databases in a number of scenarios, such as effect of city area and population, effect of ISP assignment, and number of not-returned locations.

Straipsnyje aptariamas geografinės interneto mazgų padėties aptikimas nuotoliniu būdu, kai nereikia susisiekti su pačiais mazgais (nepriklausomais nuo kliento). IP adresų geografinė informacija naudojama daugelyje sričių, pavyzdžiui, turinio personalizavimas, sukčiavimo prevencija ir aptikimas internete bei skaitmeninės žiniasklaidos teisėsauga. Vienas iš svarbiausių aspektų, tyrinėjant nuo kliento nepriklausomą geografinę informaciją, yra etaloninių reikšmių duomenų rinkinys. Kaip rodo ankstesnieji tyrimai, naudojamos etaloninės reikšmės turi daug įtakos rezultatams. Autoriai pristato savo sukurtą klaidų neturintį etaloninių reikšmių duomenų rinkinį, sudarytą iš mazgų su tiksliomis, GPS pagalba nustatytomis lokacijomis. Kiekvienam etaloninės reikšmės mazgui taip pat padaromas šalies, regiono, miesto ir ISP įrašas. Naudodamiesi savo sukurtomis etaloninėmis reikšmėmis, autoriai tiria aštuonių IP lokacijų duomenų bazių tikslumą skirtinguose scenarijuose – tokiuose, kaip miesto vietovės ir populiacijos įtaka, interneto ryšio tiekėjo paskyrimo ir negautų atsakymų, įtaka.