



# ONE-Geo: Client-Independent IP Geolocation Based on Owner Name Extraction

Yucheng Wang<sup>1,2</sup>, Xu Wang<sup>1,2</sup>, Hongsong Zhu<sup>1,2(✉)</sup>, Hai Zhao<sup>3</sup>, Hong Li<sup>1,2</sup>,  
and Limin Sun<sup>1,2</sup>

<sup>1</sup> School of Cyber Security, University of Chinese Academy of Sciences,  
Beijing, China

<sup>2</sup> Beijing Key Laboratory of IOT Information Security Technology,  
Institute of Information Engineering, CAS, Beijing, China  
[zhuhongsong@iie.ac.cn](mailto:zhuhongsong@iie.ac.cn)

<sup>3</sup> School of Computer Science and Engineering, Northeastern University,  
Shenyang, China

**Abstract.** Client-independent Internet Protocol address (IP) geolocation is a critical problem in the Internet World, of which the accuracy is based on highly reliable landmarks. However, most existing methods focus heavily on improving the location estimating method rather than improving the quality and quantity of landmarks. Without sufficient landmarks of high quality, they face difficulties when attempting to further improve accuracy. Even though some existing mining based methods dig massive landmarks from online web resources, most landmarks are of low quality because they do not make full use of these open resources. In this paper, we propose ONE-Geo, a methodology to mine highly reliable landmarks as much as possible by extracting the owner name of web servers. For a given target IP, ONE-Geo extracts the real owner name from web page information and registration records. Utilizing this clue, ONE-Geo determines the correct location by searching address information on an organization knowledge graph and conduct inference. Experimental results show that ONE-Geo achieves a median error distance of 463 m on 165 web servers and a median error distance of 7.7 km on 721 nodes that do not host a website. For web servers, ONE-Geo outperforms existing methods and several commercial tools. To be specific, 66.1% nodes are geolocated by ONE-Geo with an error less than 1 km, which is two times as many as Street-level Geolocation(SLG), which is one of the best existing methods on IP geolocating.

**Keywords:** IP geolocation · Network measurement ·  
Landmark mining

## 1 Introduction

The ability to determine the geographical location of a networking device is essential for many location-aware applications, such as content personalization,

investigating crimes, and location-based access limitation. Even though existing client-dependent geolocation methods are able to achieve high accuracy on locating client devices based on GPS, cellular or Wi-Fi, the client support is a must, which makes them irrelevant in many applications, such as location-based targeted advertising and location-based access restrictions.

As for client-independent methods, they mainly fall into two categories: (1) data mining-based approaches and (2) network measurement-based geolocation approaches. However, most existing approaches can only guarantee a coarse-grained accuracy at city-level, which can hardly meet the demands mentioned above. In most cases, the low precision of measurement-based methods results from the scarceness of landmarks with high precision location. Even though data mining-based IP geolocation approaches can provide solutions to get massive landmarks, they can only provide landmarks of low quality. Since using cloud hosts is a trend today, previous methods that generate landmarks by directly mapping an address that is revealed on a web page to the web site's IP is no longer reliable anymore. Plus, as the structure of websites is becoming more complicated, it is increasingly difficult to mine address information from web pages.

In this paper, we propose ONE-Geo to mine as many highly reliable landmarks as possible by extracting the owner names of web servers. ONE-Geo is based on three findings. (1) Owner names are more commonly shown on web pages than address information. Almost all web pages explicitly expose indications of the owner of the website, such as organization name in the title, copyright information and logo tag. (2) If a website is hosted on the cloud, the owner name usually appears in WHOIS registration records. Even though the WHOIS database only returns the address of the head office for the whole IP block, the organization name can be utilized as a clue to the location. (3) Given the owner name of an IP, it is easy to narrow down its location to several potential coordinates by using organization-location knowledge graph (OKG).

The main contributions of this paper are summarized as follows:

- (1) We propose an efficient method to mine landmarks of high quality without any measurement (ONE-Geo Alpha). Because the structures of websites are more complicated today than they have ever been in the past, it is becoming increasingly difficult to mine address information from web pages. Hence, our highly efficient method can be useful for a variety of IP geolocation efforts.
- (2) We propose a universally applicable approach which depends neither on addresses on web pages, nor addresses in IP registration records, but owner names which are common and easy to obtain. So, ONE-Geo is designed to not only mine the locations of web servers that are hosted locally, but also the ones that are hosted on the cloud. As the trend is shifting towards hosting more websites on the cloud, previous methods that generate landmarks by directly mapping an address that revealed on a web page to the web site's IP is increasingly less reliable. ONE-Geo could play a large part in filling this ever-increasing void of location-related information since it can

deduce the location of a data center by using the OKG and the owner name extracted from WHOIS database.

- (3) Based on the landmarks mined by ONE-Geo Alpha, we designed a highly fault-tolerant inference algorithm to mine as many landmarks as possible. By this method, we were able to make full use of web resources available to the public.

## 2 Related Work

**Data Mining-Based Methods:** GeoTrack [1], DRoP [2], rDNS-Geo [3] and HLOC [4] mine location hints in domain names to geolocate an IP. Structon [5] uses regular expressions to extract location information from web pages. By mapping addresses to the corresponding IPs of the web servers, it generate hundreds of thousands of landmarks. GeoCluster [1] uses the address prefixes in BGP routing tables to cluster IP addresses and then deduce the geographical location of the entire cluster by extracting location information from user registration records in the Hotmail service. Checkin-Geo [6] leverages the location data that users share in location-sharing services and logs of user logins from PCs for IP geolocation. Dan et al. [7] constructs an IP geolocation database by collecting location data from a subset of search engine logs that contain real time global positioning information obtained from mobile devices.

**Measurement-Based Methods:** GeoPing [1] maps a node to a probe's location based on the measured delays from probes to the node. CBG [8] utilizes measured delays to draw constraints and narrows down the possible region that covers a target to a continuous area. TBG [9] proves that network topology can be effectively used to achieve high geolocation accuracy. Octant [10] takes both positive and negative measurement constraints into account. SLG [11] utilizes zip codes on web pages to generate landmarks. An important contribution of this work lies in introducing a method that indirectly estimates the delay between a target and a landmark by finding the closest common router. Geo-NN [12] and LBG [13] train prediction model to geolocate IPs. SBG [14] uses smartphones as landmarks relying on crowdsourcing principles.

Data mining based methods are widely used in commercial systems due to fast response time and easy deployment. However, they can usually only provide city-level precision because some of them, such as Structon and various domain name mining methods, use the open resources in an inefficient way. Furthermore, methods that use the raw data, such as user registration records, are intrinsically unreliable. As for measurement-based methods, they rely heavily on landmarks. Most of them fail to improve accuracy further because of the lack of high-quality landmarks.

Considering the problems mentioned above, we propose ONE-Geo which relies on owner names which are common and easy to obtain. ONE-Geo is not only a more efficient, convenient and universally applicable approach, compared to previous data mining-based methods, but also an excellent solution to providing a considerable number of landmarks for measurement-based methods.

### 3 Owner Name Extraction Based Algorithm

Briefly, our geolocation approach consists of three major steps. First, we scan an IP segment and find all IPs that host websites. We crawl homepages and collect their registration records from regional Internet registry databases. For a given IP, we try to extract the owner name from the web pages and the registration records. Second, we use the owner name as a clue to search for potential addresses by an organization knowledge graph (OKG). If there is only one potential address, we directly map it to the IP address and generate a new landmark. Otherwise, by our election based inference algorithm, we infer the correct location from all candidate geographic coordinates. Third, in order to expand the coverage further, we cluster IPs and map every IP in a cluster to the same location.

#### 3.1 Owner Name Extraction

For registration information, we use the application programming interfaces (APIs) provided by WHOIS databases to get the registration records of a given IP and extract the organization name. For homepages, it is more complicated, the details of which are explained in the following section.

Without a context, existing public name entity recognition (NER) tools do not work well on extracting organization names from titles, logo tags, nor copyright information. We tried Stanford NER and Natural Language Toolkit (NLTK). Both of them return false positives in many cases. Take Stanford NER for example, we feed “Palo Alto Research Center Incorporated; © 2018. All Rights Reserved” into the recognition function and it returns “Alto Research Center Incorporated” back, which can lead to huge errors in the next inference process.

Hence, we employ another two strategies to extract organization names: by regular expression (RegEx) and by an organization name dictionary. Since there are some conventions in displaying the copyright information, the RegEx is a good choice for extracting organization names from copyright information. Different from copyright information, website titles and text in logo image tags are usually organized in a variety of styles. In order to extract organization names exactly and not to introduce redundant characters that can lead to negative effects on the inference part, we collect organization names from public semi-structured and structured knowledge databases, e.g. Wikipedia, yellow pages, and recruiting websites. In case we get more than one organization name, we determine the real owner name by scoring according to their position and frequency.

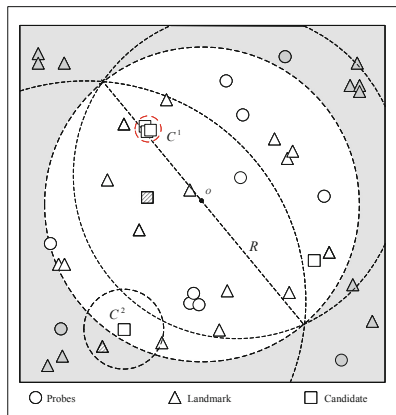
#### 3.2 Election-Based Inference Algorithm

To build our OKG, we first collected a large volume of pointer of interest (POI) data from OpenStreetMap (OSM) [15], Data Center Map [16], and The Real Yellow Pages [17] to generate organization-location links. Then, we crawled Wikipedia [18] for headquarters-subsidiary links. Based on these links, we built an organization knowledge graph for location searching.

Given the owner name of a target IP, we use it as a clue to retrieve all relevant locations of organizations and their subsidiaries from OKG. Before going to the next step, we merge highly clustered locations to a single candidate location (see  $C^1$  on Fig. 1) by calculating their average coordinate. For a given IP, if it has only one owner name and there is only one candidate point for it, we map this IP to the only location and get it into our landmark database. If there is more than one candidate returned, we leave the rest of the job to the next inference process. To our surprise, the number of IPs with only one possible location was substantial (1.2 million of 8 million). In other words, by simply searching for potential locations with the owner name, ONE-Geo got a massive number of landmarks directly without any network measurement. We named the initial ONE-Geo ONE-Geo Alpha. We added the initial landmarks collected by ONE-Geo Alpha to the existing landmark set that were prepared for the next inference part.

As for IPs with multiple candidate locations, we used CBG to filter out points that were far less likely to be the correct location. To determine the possible region of the given IP address, we first measured the network delay time from probes to the target IP by ping. Then, we converted the network delay into a geographical distance. Katz-Bassett et al. [9] and Wang et al. [11] have shown that  $4/9$  is suitable to be adopted as the converting factor between measured delay and geographical distance. Thus, we also adopted this ratio as the converting factor in our calculations. After estimating the geographical distances, by multilateration, we drew an intersection that covers the target IP. Next, we defined a circular area that covers the intersection area and filters out potential locations that are out of this circular area.

For a given target IP, if there is only one possible location left in the possible area, we map the target IP to this location and add this new landmark to the existing landmark set. If there are still more than one candidate locations, we proposed an election-based inference algorithm to determine the final location.



**Fig. 1.** An election example

As shown in Fig. 1, for a target IP, we draw a circle centered at the centroid of the constrained area, with a radius  $R'$  of the maximal error distance. The area bounded by this circle is referred to as the electoral district. In the electoral district, we define an election tuple  $E = (t, C, L, P)$  including the target IP  $t$ , all candidate locations (squares)  $C$ , a set of landmarks (triangles)  $L$ , and a set of probes (circles)  $P$ . The landmark set not only includes landmarks in the district, but also the ones that are out of the district but still in the vicinity of the candidates. For example, the striped triangle near the candidate  $C^2$  in Fig. 1 is also in the set. All the landmarks in the set were selected from the continuous-expanding landmark set which consists of not only the initial landmarks that were collected by ONE-Geo Alpha but also the ones that were generated in every inference loop.

Probes are used to measure network delays and traceroutes from a probe to a landmark or a target IP. Landmarks, in charge of distributing scores to candidate locations, play the role of a bridge that links IPs to their real locations. To be specific, by basing the measurements on probes, we can estimate the distance from any landmark to a given target IP. The closer a landmark is to the target IP, the higher the score the landmark will get. Also, we can calculate the distance from a landmark to any candidate location by computing the great-circle distance [19] between their coordinates. A location closer to a landmark gets a higher score from that landmark. For each election, the candidate location with the highest score wins the election and gets mapped to the target IP.

The final score of a candidate is defined below:

$$score(c_i) = \sum_{l_j \in L} g(l_j, c_i) s(l_j) \quad (1)$$

In this equation,  $s(l_j)$  is the individual score of a landmark  $l_j$  and  $g(l_j, c_i)$  is a redistributing gate which lets through a proportion of the individual score of the landmark  $l_j$ .

As mentioned before, a candidate should get a higher score from a landmark if it is closer to the landmark than the others. Therefore, the redistributing gates are defined below:

$$g(l_j, c_i) = 1 - \frac{e^{dis(c_i, l_j)}}{\sum_{l_k \in L} e^{dis(c_i, l_k)}} \quad (2)$$

Here,  $c_i$  represents a candidate,  $l_i$  represents a landmark, and  $dis(c_i, v_i)$  refers to the great-circle distance between them.

As for the individual score of each landmark, we use the equation below to calculate:

$$s(l_i) = \alpha s_t(l_i) + \beta s_d(l_i) \quad (3)$$

Above,  $s_t(l_i)$  represents the score in the topology perspective and  $s_d(l_i)$  represents the score in the perspective of pure network delay measurement.

The individual score of a landmark depends on the distance to the targeted node. We estimate the distance by network measurement according to two important insights from previous work: (1) Despite the fact that the direct relationship between the real geographic distance and estimated distance is incredibly weak

due to the inevitable circuitous path, queuing and processing delays, the indirect relationship on distance is largely preserved [11]. I.e., the two nearest IP addresses usually have the smallest network delay. (2) For geographically adjacent IP addresses, their network delay measurements resulting from the same probe should be similar [11,20].

In the topology perspective, learning from [11], we manage to calculate the indirect delay distance (IDD) from a landmark to a target IP by doing traceroute measurement and finding the closest common router. Then, we combine IDD into a delay vector and normalize it. Each element can be calculated by the following equation:

$$s_t(l_i) = 1 - \frac{e^{d(l_i,t)}}{\sum_{l_j \in L} e^{d(l_j,t)}} \quad (4)$$

Here,  $d(l_i, t)$  represents the indirect delay time from a landmark to the target IP.

Following this, in the perspective of pure network delay measurement, we measured the pure delay between each probe and every landmark by ping. Then, we embedded all delays into a delay vector. We employ the cosine similarity to estimate the distance. The score of a landmark can be calculated by an equation below:

$$s_d(l_i) = \frac{\text{sim}(l_i, t)}{\sum_{l_j \in L} \text{sim}(l_j, t)} \quad (5)$$

Above,  $\text{sim}(l_i, t)$  represents the cosine similarity of a landmark and the target.

Based on the initial landmarks that ONE-Geo Alpha collected, we first deduced the coordinates of the IPs that had two candidate locations. After completing the inference of the two-candidate IPs, we stepped further into the inference of the ones that had three candidate locations. Every time the inference was done, we mapped the target IPs to the estimated locations to generate a batch of new landmarks. Recursively, we added the new landmarks into an existing landmark set and repeat the inference algorithm until no more IPs that could meet the launch condition of a new loop of inferences existed. A new launch is conditional on the credibility of an election being higher than a threshold  $T$ . The credibility is defined as shown below where  $\text{size}(P)$  and  $\text{size}(L)$  represent the number of the probes and the number of the landmarks separately:

$$c(E_i) = \ln(1 + \text{size}(P)\text{size}(L))^{1/2} \quad (6)$$

### 3.3 Expanding IP Coverage

We refer to the enhanced ONE-Geo as ONE-Geo Beta. After the inference process, almost 4 million landmarks were generated from 8 million IPs. However, all of them were web servers. Therefore, in order to empower ONE-Geo to geolocate nodes that do not host a website, we must go further to enlarge the IP coverage. Freedman et al. [21] claims that 97% of IPs in the same /24 segment are at nearly the same location with a slight distance from each other. Inspired by this finding, we adopted a /24 clustering method to expand the coverage. Instead

of mapping all clustered IPs to the same location, we modified this expanding method a little to avoid introducing errors.

In a /24 IP segment, we cluster existing landmarks and define the largest cluster as the dominant group, which determines the dominant location of this segment. The coordinate of the dominant location is defined as the average coordinate of landmarks in the group. We named the landmarks in the vicinity of the dominant location the followers, and the nearest one the leader. The rest of the dispersed ones are labeled loners.

For the existing landmarks, we did not change their location because the coordinates returned by OKG were very precise. For the rest of the location-unknown IPs, we mapped them to the dominant location of a group, if and only if, they met the requirements to get into the group. We designed 2 rules to evaluate the qualification to join the group. First, we calculated a new node's IDD to the leader, and the IDD between the leader and every single follower. If the IDD of the new IP was less than the IDD of the remotest follower to the leader, it was included in the group. Second, if its IDD to the nearest follower was far less (5 times less in our experiment) than the IDD to any of the loners, it is in the group.

After the addition of the 2 rules mentioned above, a lot of new IPs were selected and clustered into the dominant group. Following this, we mapped the new nodes to the location of the dominant group and got new landmarks. By this method, ONE-Geo expanded the IP coverage even further. We named the fully-fledged ONE-Geo ONE-Geo Epsilon.

## 4 Evaluation

Our research indicates that, SLG [11] is a typical and one of the best existing IP geolocation methods. Therefore, by primarily focusing on a comparison of ONE-Geo with SLG, we can demonstrate the improvement of ONE-Geo on both web server nodes and Internet of things (IOT) nodes. In addition, we compared ONE-Geo with several commercial tools (ipstack [22], MaxMind's GeoIP2 [23], IPIP [24], and ipinfo [25]) to show ONE-Geo achieves much better estimation precision than popular commercial tools on web servers, and achieves similar precision to them on IOT nodes. In our experiments, we limited our dataset to the US. This choice was made for 3 reasons. First, the time needed to deal with cross-language problems when extracting organization names from web pages made it impractical. Second, analysis of ONE-Geo through the lens of multiple language is not focus of this paper. And third, we can get more ground-truth information from the nodes in the United States.

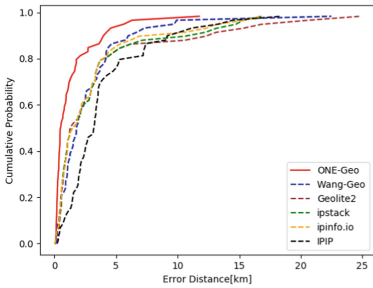
In order to evaluate the accuracy of our method, we collected ground-truth data from PlanetLab [26] and RIPE Atlas [27]. On these two platforms, nodes with reported coordinates are shared with the public. The PlanetLab data set is a commonly used data set in IP geolocation research (e.g. [9–11]), but the quantity of nodes is limited and are all web servers. To complement this data set, we collected data on IOT nodes from RIPE Atlas. In the end, after filtering



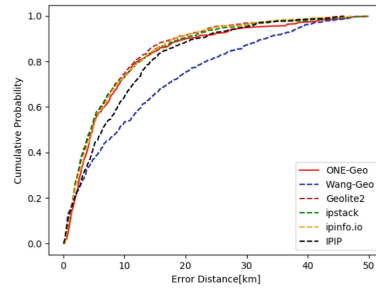
out inaccessible nodes, we had 165 web servers and 980 IOT nodes left. As to evaluating coverage and efficiency, we scanned 64.7 million IPs that host websites and crawled 8,283,809 homepages from nodes in the United States.

#### 4.1 Accuracy

As the average error distance is highly influenced by some abnormal errors from a few nodes, the median error is more commonly used to indicate the accuracy of geolocation systems. In the experiment on PlanetLab nodes, the median error distance of ONE-Geo (Beta), SLG, IPIP, ipstack, ipinfo, and Geolite2 are 463 m, 1768 m, 3161 m, 1463 m, 1463 m, and 1272 m respectively. As we can see from Fig. 2 on web servers, both ONE-Geo and SLG perform well; and ONE-Geo performs much better than SLG. An important difference between SLG and ONE-Geo is that SLG tries hard to find the closest locally-hosted node to the target while ONE-Geo tries to find the web server itself by OKG. Therefore, it is easier for ONE-Geo to control the errors within 1 km when geolocating web servers. To be specific, 66.1% of nodes in this experiment were geolocated by ONE-Geo with an error less than 1 km while only 28.8% were geolocated by SLG within the same margin.



**Fig. 2.** Comparison of error distance on PlanetLab nodes



**Fig. 3.** Comparison of error distance on RIPE nodes

ONE-Geo (Epsilon) covers 721 of the 980 IPs after expanding. Even though ONE-Geo can not cover all of them, this number is high enough to evaluate its ability to geolocate nodes that do not host a website. For these 721 nodes, ONE-Geo also performs well (see Fig. 3). This is not a surprise because of the large number of students, teachers, and local community members that live around schools, universities, and research institutions. Since ONE-Geo can cover most nodes that host websites by these organizations, it can also estimate the location of the nodes around the web servers in a precise way. To be specific, the median error distance of ONE-Geo, SLG, IPIP, ipstack, ipinfo, Geolite2 are 7758 m, 14999 m, 8801 m, 5704 m, 6453 m and 6352 m respectively, in which SLG performed the worst (median error 9295 m higher than the leader) while ONE-Geo

had a more prominent position in the group (median error 2054 m higher than the leader).

As the main idea of SLG is to associate the target’s location with the landmark with the minimum distance and they estimate the distance by the minimum indirect delay, there are 3 reasons that lead to the low precision of SLG on nodes (excluding those with web servers). First, it is not guaranteed to find a common router or the closest common router because of the scarcity of landmarks and probes. Second, the shortest network delay does not always mean the shortest distance [6]. Third, SLG can only utilize local web servers as their landmarks. Since there is a trend to use content delivery network (CDN) to distribute content or use cloud services to store archives, these kinds of local nodes are decreasing, which leads to a lack of high-quality landmarks.

## 4.2 Coverage

Most existing mining methods, like Structon, focus on utilizing the address information or zip codes on the pages to locate the web servers, like Structon. However, our experiment indicates that these methods can hardly cover most potential landmarks. We used 8 million web pages to estimate the proportion of websites that reveal the addresses and zip codes on their homepage. The results on Table 1 show that the coverage of methods mining address or zip codes are much less than the methods that mine owner names, like ONE-Geo.

**Table 1.** Comparison of coverage

Clue	Hit	Pages	Ratio
Address	1,037,271	8,283,809	0.125
Zip Code	203,673	8,283,809	0.024
Owner Name	<b>3,900,443</b>	<b>8,283,809</b>	<b>0.471</b>

**Table 2.** Comparison of efficiency

Methodology	Landmarks	Source	Ratio
Structon	157,407	502,880,364	0.0003
Checkin-Geo	31,634	92,153	0.3432
ONE-Geo	<b>3,900,443</b>	<b>8,283,809</b>	<b>0.4708</b>

Our research suggests that, the best two landmark mining methods are Structon and Checkin-Geo. In their reports, Structon mines 157K landmarks from 502M web pages and Checkin-Geo mines 31K landmarks from 92K login records. Compared to them, ONE-Geo is a more efficient method, which gets 3.9M landmarks from 8.2M pages (Table 2). Structon spends too much time and computing

resources on processing unnecessary pages on the same website. Correspondingly, ONE-Geo can extract the owner names utilizing only the homepage, of which the efficiency is three orders of magnitude higher than Structon. Checkin-Geo has a relatively higher efficiency than Structon. However, Checkin-Geo is not universally applicable because the raw material sources they use are the private data of certain social network companies, which are not open to the public.

## 5 Conclusion

In this paper, we propose ONE-Geo, a methodology which exploits owner names revealed on homepages and registration records to mine landmarks. Experimental results show ONE-Geo achieves fine-grained precision and large coverage. ONE-Geo Alpha is an efficient method to mine landmarks of high quality without any measurement. By constructing inference, ONE-Geo Beta makes full use of web resources and registration records to mine as many landmarks as possible. It is also a universally applicable approach which depends on neither addresses on web pages nor addresses in IP registration records, but owner names which are common and easy to obtain.

**Acknowledgements.** This work was supported in part by the National Key R&D Program of China (Grant No. 2018YFB0803402, No. 2017YFB0802804), the Key Program of National Natural Science Foundation of China (Grant No. U1766215), and the National Natural Science Foundation of China (Grant No. 61702503, No. 61702504).

## References

1. Padmanabhan, V.N., Subramanian, L.: An investigation of geographic mapping techniques for internet hosts. In: ACM SIGCOMM Computer Communication Review, vol. 31, pp. 173–185. ACM (2001)
2. Huffaker, B., Fomenkov, M., et al.: DRoP: DNS-based router positioning. ACM SIGCOMM Comput. Commun. Rev. **44**(3), 5–13 (2014)
3. Dan, O., Parikh, V., Davison, B.D.: IP geolocation through reverse DNS. arXiv preprint [arXiv:1811.04288](https://arxiv.org/abs/1811.04288) (2018)
4. Scheitle, Q., Gasser, O., Sattler, P., Carle, G.: HLOC: hints-based geolocation leveraging multiple measurement frameworks. In: 2017 Network Traffic Measurement and Analysis Conference (TMA), pp. 1–9. IEEE (2017)
5. Guo, C., Liu, Y., Shen, W., Wang, H.J., Yu, Q., Zhang, Y.: Mining the web and the internet for accurate IP address geolocations. In: INFOCOM 2009 IEEE, pp. 2841–2845. IEEE (2009)
6. Liu, H., Zhang, Y., Zhou, Y., Zhang, D., Fu, X., Ramakrishnan, K.: Mining check-ins from location-sharing services for client-independent IP geolocation. In: 2014 Proceedings IEEE INFOCOM, pp. 619–627. IEEE (2014)
7. Dan, O., Parikh, V., Davison, B.D.: Improving IP geolocation using query logs. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 347–356. ACM (2016)
8. Gueye, B., Ziviani, A., Crovella, M., Fdida, S.: Constraint-based geolocation of internet hosts. IEEE/ACM Trans. Netw. (TON) **14**(6), 1219–1232 (2006)

9. Katz-Bassett, E., John, J.P., Krishnamurthy, A., Wetherall, D., Anderson, T., Chawathe, Y.: Towards IP geolocation using delay and topology measurements. In: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, pp. 71–84. ACM (2006)
10. Wong, B., Stoyanov, I., Sirer, E.G.: Octant: a comprehensive framework for the geolocalization of internet hosts. In: NSDI, vol. 7, pp. 23–23 (2007)
11. Wang, Y., Burgener, D., Flores, M., Kuzmanovic, A., Huang, C.: Towards street-level client-independent IP geolocation. In: NSDI, vol. 11, pp. 27–27 (2011)
12. Jiang, H., Liu, Y., Matthews, J.N.: IP geolocation estimation using neural networks with stable landmarks. In: 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 170–175. IEEE (2016)
13. Eriksson, B., Barford, P., Sommers, J., Nowak, R.: A learning-based approach for IP geolocation. In: Krishnamurthy, A., Plattner, B. (eds.) PAM 2010. LNCS, vol. 6032, pp. 171–180. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-12334-4\\_18](https://doi.org/10.1007/978-3-642-12334-4_18)
14. Ciavarrini, G., Luconi, V., Vecchio, A.: Smartphone-based geolocation of internet hosts. *Comput. Netw.* **116**, 22–32 (2017)
15. Openstreetmap. <https://www.openstreetmap.org/copyright>
16. Data center map. <https://www.datacentermap.com>
17. YP.com. The real yellow pages. <https://www.yellowpages.com/>
18. Wikipedia. The free encyclopedia. <https://www.wikipedia.org/>
19. Vincenty, T.: Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Surv. Rev.* **23**(176), 88–93 (1975)
20. Shue, C.A., Paul, N., Taylor, C.R.: From an IP address to a street address: using wireless signals to locate a target. In: WOOT (2013)
21. Freedman, M.J., Vutukuru, M., Feamster, N., Balakrishnan, H.: Geographic locality of IP prefixes. In: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, pp. 13–13. USENIX Association (2005)
22. ipstack - free IP geolocation API. <https://ipstack.com/>
23. Maxmind GeoIP2. <https://www.maxmind.com/en/geoip2-services-and-databases>
24. IPIP.NET. <https://www.ipip.net>
25. IPinfo. <https://ipinfo.io>
26. PlanetLab. <https://www.planet-lab.org/>
27. RIPE Atlas. <https://atlas.ripe.net>