

Article

Evaluation Method of IP Geolocation Database Based on City Delay Characteristics

Yuancheng Xie [†] , Zhaoxin Zhang ^{*}, Yang Liu [†], Enhao Chen and Ning Li

Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China; 20b903060@stu.hit.edu.cn (Y.X.); liuyang20200905@163.com (Y.L.); chenenhao741369@163.com (E.C.); li.ning@hit.edu.cn (N.L.)

^{*} Correspondence: zhangzhaoxin@hit.edu.cn

[†] These authors contributed equally to this work.

Abstract: Despite the widespread use of IP geolocation databases, a robust and precise method for evaluating their accuracy remains elusive. This study presents a novel algorithm designed to assess the reliability of IP geolocation databases, leveraging the congruence of delay distributions across network segments and cities. We developed a fusion reference database, termed CDCDB, to facilitate the evaluation of commercial IP geolocation databases. Remarkably, CDCDB achieves an average positioning accuracy at the city level of 94%, coupled with a city coverage of 99.99%. This allows for an effective and comprehensive evaluation of IP geolocation databases. When compared to IPUU, CDCDB demonstrates an increase in the number of network segments by 18.7%, an increase in the number of high-quality network segments by 13.2%, and an enhancement in the coverage of city-level network segments by 20.92%. The evaluation outcomes reveal that the reliability of IP geolocation databases is not uniform across different cities. Moreover, distinct IP geolocation databases display varying preferences for cities. Consequently, we advise online service providers to select suitable IP geolocation databases based on the cities they cater to, as this could significantly enhance service quality.

Keywords: IP geolocation databases; evaluation; network segment; delay distribution; city delay characteristics



Citation: Xie, Y.; Zhang, A.; Liu, Y.; Chen, E.; Li, N. Evaluation Method of IP Geolocation Database Based on City Delay Characteristics. *Electronics* **2024**, *13*, 15. <https://doi.org/10.3390/electronics13010015>

Academic Editor: Tania Cerquittelli

Received: 31 October 2023

Revised: 8 December 2023

Accepted: 15 December 2023

Published: 19 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid proliferation of the Internet and the ubiquitous adoption of mobile applications have precipitated an unprecedented demand for the precise geolocation of IP addresses [1,2]. In the recent past, a multitude of IP geolocation methodologies have been proposed, finding wide-ranging applications across diverse sectors such as online services, targeted advertising, social sharing, network performance enhancement, and bolstering network security [3–5]. The most commonly employed technique for IP geolocation hinges on database queries, which involves soliciting an IP geolocation database to extract the geolocation specifics of IP addresses. A number of Internet corporations have rolled out sophisticated IP geolocation libraries for public utilization, notable among them being MaxMind GeoIP2 [6] and IP2Location [7]. Patricia Callejo and her team underscore that a staggering 50% of online advertisements necessitate location information of IP addresses, which are furnished by these IP geolocation libraries [8].

However, the veracity of IP geolocation databases is often called into question. These databases typically remain reticent about the methodologies they deploy in their construction. Moreover, a significant number of these databases are beleaguered by the absence of location data and high incidences of address migration [9]. As a result, different IP geolocation databases may proffer inconsistent results for the same IP address. Poese et al. inferred that while IP geolocation databases furnish reliable location information at the national level, their city-level data leave much to be desired [10]. Gharaibeh et al., in their

appraisal of four databases using a limited dataset, deduced that these databases exhibit a 95.8% location consistency at the national level; however, this consistency plummets to 71% at the city level [11]. Ben Du et al., in their evaluation of RIPE IPmap's performance, discovered that its city-level positioning accuracy hovered around 80.3% [12]. Gouel et al., through a decade-long longitudinal study, unearthed significant short-term dynamics in the GeoIP2 database [9]. These studies collectively underscore a pervasive issue: the positioning information proffered by IP geolocation databases for IP addresses is not entirely trustworthy. This is particularly pronounced for city-level positioning information and smaller geographic regions. Therefore, effectively evaluating the quality of IP geolocation repositories and providing corresponding recommendations has emerged as a viable approach for individuals selecting IP geolocation repositories.

The current evaluation methodologies for IP geolocation databases primarily bifurcate into two categories. The first involves the research team amassing a ground truth collection of IP addresses and assessing the database's accuracy predicated on this collection [13–16]. While this method can accurately assess verified IP addresses, it is not without its limitations. The range of verifiable IP addresses is circumscribed, precluding a comprehensive assessment of the IP geolocation database. Additionally, procuring the ground truth collection poses a formidable challenge. The second approach evaluates IP geolocation databases predicated on consistency [17–19]. This is achieved by comparing the similarity between multiple databases, such as data consistency rate or delay similarity rate. This method addresses the limited detection range issue of the previous scheme. However, the accuracy of the evaluation results is subpar, and its reliability is heavily contingent on the original IP geolocation repository itself. This underscores the need for more robust and comprehensive evaluation methodologies for IP geolocation databases.

The accurate and comprehensive evaluation of IP geolocation databases presents a formidable challenge. In this study, we unveil a novel method for evaluating IP geolocation repositories predicated on city delay characteristics. This approach is designed to address the conundrum of striking a balance between the scope and accuracy of traditional evaluation methods. Initially, we delve into the feasibility of a strategy that evaluates IP geolocation repositories based on cities, network segments, and delay distributions. We propose a strategy that gauges the reliability of each IP geolocation repository from a city-centric perspective, predicated on the similarity between the delay distribution of network segments and that of a city. Subsequently, we construct a converged reference database, christened CDCDB, which boasts the capability to encompass all IP addresses and furnish highly accurate localization results, which is a requirement that remains unfulfilled by other evaluation methods. Our contributions are primarily encapsulated in the following points:

- We propose a minimum network segment matching mechanism. This mechanism effectively integrates the network segment information from multiple IP geolocation databases, addressing the issues of address overlap between different databases and candidate address problems within segments.
- We introduce a nearest neighbor network segment matching mechanism. This mechanism expands the candidate cities for a network segment, which is crucial for improving the location accuracy of the segment.
- We put forth the concept of city delay feature (CDC). It accurately describes the delay distribution and delay pattern of a target city obtained from a specific detection point, offering a novel solution to improve the city-level geolocation accuracy of an IP address.
- Finally, we propose a network segment verification algorithm based on city delay features. By leveraging the similarity between the delay distribution of network segments and that of cities, we propose network segment delay feature values for candidate cities and select candidate cities for network segments based on these values.

The remainder of this paper is structured as follows: Section 2 provides a review of the relevant evaluation methods for IP geolocation databases. Section 3 discusses three distinct evaluation strategies. Section 4 offers a detailed description of the methodology for

evaluating IP geolocation databases based on city delay features. Section 5 presents the experimental and evaluative procedures conducted. Finally, Section 6 provides a summary of the paper.

2. Related Work

A multitude of IP geolocation methodologies have been proposed, finding wide-ranging applications across diverse sectors [20,21]. For instance, Q Li et al. achieved IP geolocation through fine-grained and stable webcam landmarks [1], while Z Ma et al. proposed an IPv6 geolocation algorithm predicated on neural networks [22]. Despite the rapid strides in IP geolocation technology, numerous scholars have underscored that current IP geolocation databases provide geolocation information for IP addresses that is reliable at the national level but falls short at the city level [10–12]. M Cozar et al. pointed out that a number of online databases and services furnish geolocation data for IP addresses with varying degrees of accuracy and reliability [23]. Livadariu et al. demonstrated that existing databases tend to incorrectly geolocate IPs belonging to global networks and IPs moving between networks [24]. D Ganelin et al. unearthed evidence of bias in IP geolocation: it makes greater geographic errors for users from more economically disadvantaged areas, and it disproportionately places users from more economically disadvantaged areas in prosperous areas [25]. IP geolocation databases constitute the most prevalent conduit for individuals to procure geolocation information of IP addresses. However, the precision of some geolocation services leaves room for enhancement [26]. Consequently, the selection of a suitable IP geolocation database emerges as a topic of significant importance.

At present, methodologies for evaluating IP geolocation repositories can be broadly bifurcated into two categories. One approach entails researchers evaluating IP geolocation repositories by independently constructing ground truth collections of IP addresses. Specifically, researchers assess the accuracy of these databases by utilizing a set of IP addresses with known or estimated locations and juxtaposing them with the locations reported in the databases. Researchers have amassed ground truth collections of IP addresses through a variety of means. For instance, Shavitt et al. constructed ground truth sets using virtual node aggregation [13], while Jiang et al. gathered IP location data from various sources [14], including government websites, universities, well-known companies, and shopping forums. Komosny et al., on the other hand, collected IP address data from other websites [15]. J Sommers harnessed geolocation information embedded in non-standard HTTP response headers and unencrypted HTTP cookies [27]. Saxon et al. utilized validation set data obtained from devices such as mobile terminal GPS [16]. O Dan et al. enhanced IP geolocation by mining search engine click logs [28]. These methods can procure location-accurate IP addresses, but the data size is generally small. Consequently, validation-set-based assessment methods can only assess the reliability for a limited number of IP addresses in the database, and cannot comprehensively assess all IP addresses in the database. Moreover, in constructing a ground truth set of IP addresses, ensuring the accuracy of the dataset poses a formidable challenge [29].

The second approach is the consistency-based assessment method, which evaluates IP geolocation repositories by comparing data consistency or latency consistency between databases. Huffaker et al. utilized data consistency to assess the reliability of IP geolocation repositories [17], premised on the assumption that the majority of IP addresses in these repositories yield consistent localization results. Their findings suggest that IP geolocation repositories tend to prioritize the localization accuracy of routable IPs. H Li et al. extended this work by proposing a weighted voting database fusion model [18], DEROSSO, in conjunction with data consistency rates. Their results indicate that Chinese providers' IP geolocation repositories are more adept at locating IP addresses within China compared to foreign providers' repositories. X Bo et al. introduced the delay consistency rate from a delay similarity perspective and enhanced the DEROSSO model based on this metric [19]. They proposed the D&D model for evaluating IP geolocation repositories. The D&D model supplements the active validation method based on the data consistency

rate, thereby improving the reliability of data consistency rate-based evaluation methods. While consistency-based evaluation methods enable large-scale evaluation of IP geolocation repositories, these methods lack location validation means, and their reliability heavily depends on the original IP geolocation repositories. Notably, they are unable to validate numerous non-routable IP addresses.

3. Evaluation Strategies

In this section, we will delve into the relevant strategies employed for evaluating IP geolocation repositories. These strategies form the theoretical foundation for our proposed method of database evaluation, which is based on urban delay characteristics.

3.1. City-Based Assessment Strategies

The literature [10–12] highlights that while different IP geolocation libraries perform commendably in national-level geolocation efforts for IP addresses, they exhibit significant disparities in city-level geolocation efforts. This observation motivates us to consider the primary differences between IP geolocation repositories for effective evaluation. It is a promising approach to assess the quality of different IP geolocation repositories based on the differences in city-level geolocation of the databases.

In order to substantiate the conclusions drawn earlier, we selected four widely used IP geolocation databases for similarity experiments. These databases encompass IPUU [30], IP2Region [31], IP2LocationLite [7], and GeoLite2 [6]. The IPUU database was procured from AIWEN TECH, while the remaining three databases are freely accessible versions. Initially, we counted the number of network segments contained in all experimental IP geolocation libraries. During this process, we discerned that a significant number of IP addresses within the same database lack city location information, and the number of IP addresses with city location information varies across different databases. We define the percentage of IP addresses with city location information in the IP geolocation database as the city coverage of the database. In a similar vein, we define the country coverage of the database. Consequently, we synchronized the number of network segments and coverage across all experimental IP geolocation databases, as depicted in Table 1.

Table 1. Number of network segments and coverage of IP geolocation repositories.

Database	Network Segments			Coverage	
	ALL	Country	City	Country	City
IPUU	13,704,888	13,701,348	10,838,666	99.97%	79.08%
IP2LocationLite	3,841,655	3,834,796	3,831,202	99.82%	99.73%
GeoLite2	3,463,148	3,463,046	3,463,046	99.99%	99.99%
IP2Region	1,023,551	1,023,533	338,311	99.99%	33.05%

Our analysis reveals that IPUU performs the best, containing over 10 million network segments with country location information and over 10 million network segments with city location information. IP2LocationLite and GeoLite2 perform similarly, containing information for over 3 million network segments. In contrast, IP2Region performs the worst, containing just over 1 million network segments with country location information and over 330,000 network segments with city location information. All four experimental databases have close to 100% country coverage, but there is a significant difference in terms of city coverage. IP2LocationLite and GeoLite2 have city coverages of 99.73% and 99.99%, respectively, and IPUU has a city coverage of 79.08%, while IP2Region has a city coverage of only 33.05%. This suggests that all the experimental databases prioritize national-level geolocation of IP addresses, but their focus on city-level geolocation varies significantly, which aligns with findings reported in other works. From Table 1, we infer that the primary difference between different IP geolocation repositories is manifested in the city-level geolocation work.

We introduce the comparative agreement rate metric to further assess the similarity of city-level geolocation information. The design of the comparative agreement rate is inspired by the data agreement rate proposed by Huffaker [17]. It refers to the ratio of the number of IPs with consistent positioning to the number of all IPs provided by two IP geolocation repositories for the same city. A higher comparative agreement rate indicates greater similarity between the localization results of the two IP geolocation repositories in that city. We conducted pairwise comparisons of the four experimental IP geolocation repositories. To streamline the data volume for the experiments, we selected 42 international cities with common coverage across the four experimental databases as the experimental subjects. These 42 experimental cities are situated in 32 countries worldwide and encompass approximately 560 million IP addresses. Therefore, the experimental results possess considerable objectivity. The experimental results are depicted in Figure 1.

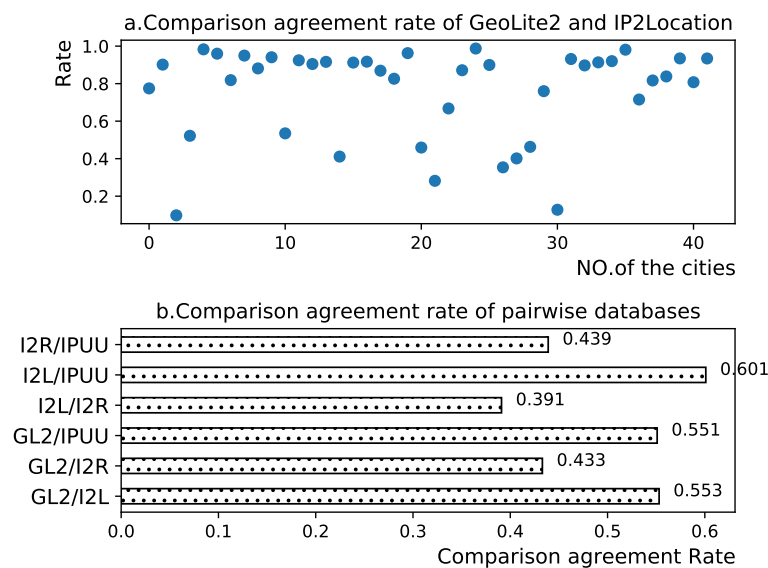


Figure 1. (a) Schematic of the comparison agreement rate results for IP2LocationLite and GeoLite2. (b) Schematic of the comparison agreement rate results of all databases.

Figure 1a presents the schematic results of the comparative agreement rate between IP2LocationLite and GeoLite2 across the 42 experimental cities. We observe that the comparative agreement rate between IP2LocationLite and GeoLite2 exceeds 80% in most cities. This suggests that the majority of cities in IP2LocationLite and GeoLite2 yield consistent localization results, thereby resulting in a high comparative agreement rate. However, there are certain cities where the comparative agreement rate is notably low. For instance, in Denver, USA, the comparative agreement rate is a mere 9%. This indicates a significant discrepancy between the location information provided by the two IP location repositories in that city. Overall, approximately 25% of cities exhibit a substantial conflict between the geolocation information provided by IP2LocationLite and GeoLite2. This underscores that different IP geolocation libraries offer varying levels of reliability for IP addresses across different cities.

Figure 1b provides a schematic of the comparative agreement rates obtained from pairwise comparisons of the four experimental databases. We observed that the comparative agreement rates between different IP geolocation databases vary, with overall rates being low. Among the 42 selected cities, the two databases exhibiting the highest similarity are IPUU and IP2LocationLite, with an average comparative agreement rate of only 60.1%. Conversely, IP2LocationLite and IP2Region, which display the lowest similarity, have an average comparative agreement rate of a mere 39.1%. This suggests that while most cities have location-aligned IP addresses provided by multiple IP geolocation repositories, this high degree of similarity does not extend to all cities. In a few cities,

there is significant conflict between the positioning information provided by different IP geolocation repositories.

Table 1 and Figure 1 elucidate that at the city-level geolocation granularity, the reliability of location information furnished by the same IP geolocation repository exhibits considerable variation from city to city. Furthermore, there exist discernible differences in the location information provided by different IP geolocation databases for the same city. This observation serves as a catalyst for us to evaluate IP geolocation repositories from a city-centric perspective. This approach promises to yield more nuanced insights into the performance and reliability of these repositories, thereby facilitating more informed decision-making for users leveraging these databases.

3.2. Network Segment-Based Evaluation Strategy

We further examine the difference in network segment quality among IP geolocation repositories. IP geolocation repositories typically store data in the format of {net segment, location}, where a net segment represents a set of numerically neighboring clusters of IP addresses. The quantity of network segments can, to a certain extent, reflect the quality of an IP geolocation database: the more network segments the database contains, the more detailed the information it provides, and consequently, the more credible its geolocation information becomes. We posit that the longer the prefix length, the fewer IP addresses contained in the network segment, and consequently, the higher the quality of the segment. Therefore, we categorize the quality of network segments based on the number of IPs they contain. We consider network segments with prefix lengths longer than /24 as high-quality segments, segments equal to /24 as normal-quality network segments, and segments smaller than /24 as low-quality network segments. Based on this categorization, we conducted quality statistics on all the network segments of the experimental IP geolocation repository. The experimental results are depicted in Figure 2.

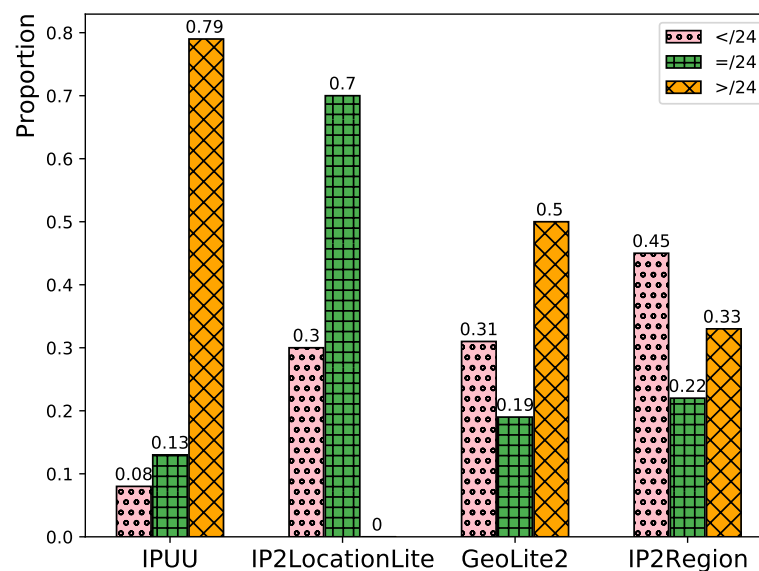


Figure 2. Quality of network segments for IP geolocation repositories.

Our analysis reveals that IPUU exhibits superior performance, comprising 79% high-quality network segments and 13% normal-quality network segments. GeoLite2 follows, with 50% of its network segments classified as high-quality and 19% as normal-quality. The distinction between IP2LocationLite and the preceding two databases primarily lies in the high-quality network segments; IP2LocationLite contains merely 70% normal-quality network segments and 31% low-quality segments. The least effective is the IP2Region database, with a significant 45% of its network segments categorized as low-quality. Diverse IP geolocation databases proffer varying numbers of network segments, engendering sig-

nificant disparities in the quality of network segments in the databases. The more superior the quality of a network segment, the higher the likelihood that its internal IP addresses are situated in the same area. Consequently, the proportion of high-quality network segments can be leveraged to evaluate the overall reliability of the IP geolocation database.

Determining the geographical location of the IP address in the form of network segments also boasts higher credibility. While most current IP geolocation techniques are based on IP addresses, Gharaibeh et al. proposed that geolocation can be achieved more efficiently by utilizing clusters of IP addresses located at the same address [32]. Yong Gan et al. demonstrated that the geolocation of an IP address can be inferred from sequences of neighboring IPs [33], and that the average error for the method based on these neighboring sequences is 20–30 km. These studies suggest that the geolocation of IP addresses can be more accurately determined by using clusters of neighboring IP addresses, which aligns with the practices of network operators in assigning IP addresses. Moreover, when constructing a reference database, the computational effort required to construct a reference database based on network segments is significantly lower than that required for a database based on individual IP addresses. Therefore, we will evaluate the IP geolocation database based on network segments.

3.3. Evaluation Strategies Based on Delay Distribution

A significant number of IP geolocation techniques rely on the round-trip delay information of IP addresses for geolocation. For instance, Q Zhao et al. utilized delay as a feature to elucidate the relationship between distance and delay and found that delay can be effectively employed in geolocation efforts for IP addresses [34]. However, delay-based geolocation methods do not always yield accurate results. As pointed out by Marchetta et al., the round-trip time (RTT) from the last hop of the traceroute and the RTT from the middle hop are not comparable [35]. They originate from different ICMP responses, which are handled differently by the router. Consequently, the middle hop may register a higher latency than the destination.

We corroborate this conclusion as well. We actively probed the IP addresses of the top 100 cities globally from three probe sources in Weihai, London, and Newark, and plotted the average RTT-distance curves as depicted in Figure 3. The distance is the Euclidean distance between the target city and the probe node calculated based on latitude and longitude. In Figure 3, we observe that the round-trip delay information does not consistently exhibit a linear relationship with the distance. Once the distance between source and destination addresses surpasses a certain threshold, the relationship between RTT and distance begins to deteriorate. Simultaneously, the assumption that a shorter distance corresponds to a smaller RTT is not necessarily accurate, as evidenced by numerous “spurs” in the delay curves of cities with similar distances in Figure 3. This suggests that the feedback of network delay on distance is influenced not only by the magnitude of the distance but also by the network environment. Consequently, network delay does not accurately reflect distance, and IP geolocation algorithms based solely on network delay are susceptible to errors.

F Zhao et al. proposed a novel IP geolocation algorithm predicated on the similarity of directional routers and local delay distributions, providing a fresh perspective on potential solutions [36]. F Zhao highlighted that this IP geolocation algorithm, based on the similarity of directional routers and local delay distributions, improves the accuracy by an average of 19% over the learning-based method at city-level geolocation granularity. The average accuracy of the algorithm proposed by F Zhao et al. exceeds 92% at city-level localization granularity.

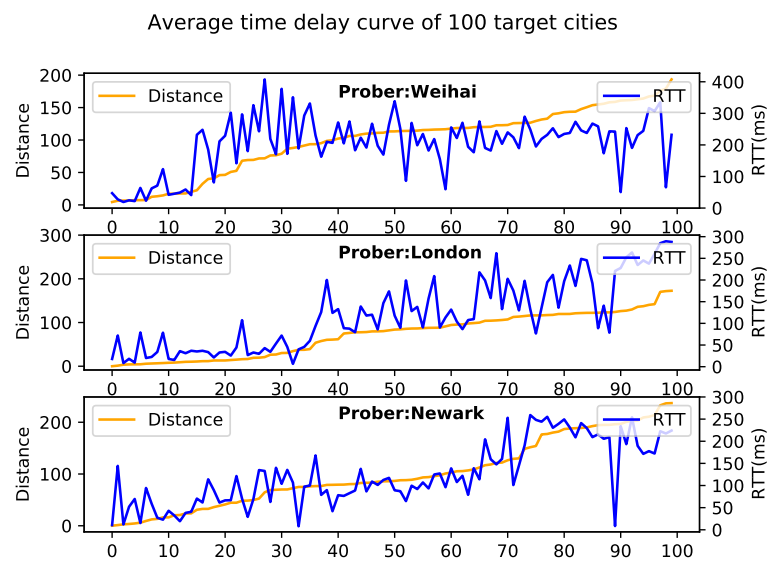


Figure 3. RTT–distance curve graph of 100 cities.

To verify the reliability of the method of the localized delay distributions' similarity proposed by F Zhao et al., we conducted delay probing from the London node for IP addresses localized in four cities, namely London, Paris, Brussels, and Rotterdam. Subsequently, we tallied the delay frequency of different RTT messages located in the same city according to their frequency of occurrence. By considering the delay frequency as a characteristic of RTTs, we plotted the frequency–delay graph as depicted in Figure 4.

From Figure 4, we observe that the pattern of delay distribution across cities is not uniform. There is an overlap in the delay ranges of neighboring cities, which clearly illustrates that the network delay-based approach is susceptible to errors when cities are in close proximity to each other. Although the delay distributions of neighboring cities all resemble a normal distribution, the delay distribution patterns of different cities are entirely distinct. This suggests that the localization results obtained from IP geolocation based on network delay distribution are indeed more reliable than those based solely on network delay. Therefore, we will evaluate the IP geolocation library based on the distribution pattern of network delay.

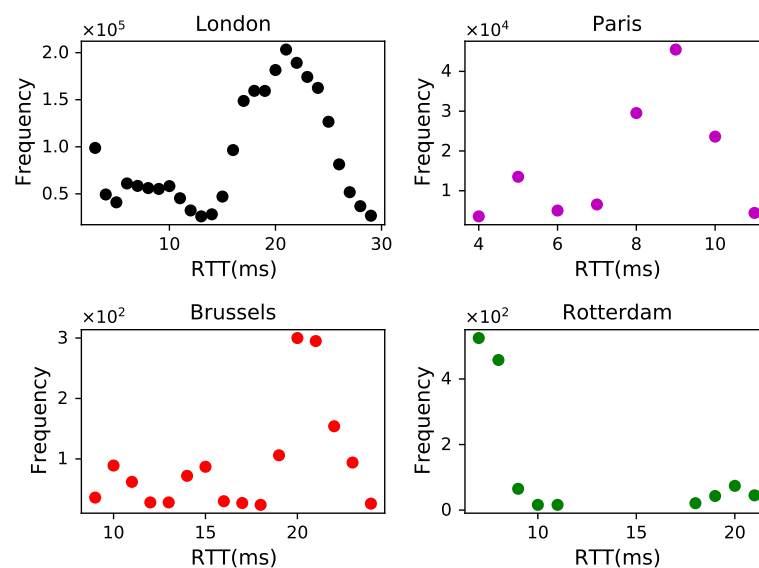


Figure 4. Frequency–delay diagram of London, Paris, Brussels, and Rotterdam (London).

4. Method

4.1. Overview

Having discussed the strategies employed to evaluate IP geolocation libraries, we now delve into the specifics of our approach. In our analysis, we introduce two concepts of IP address clusters:

- **Network Segment:** The first is a network segment, which is a set of numerically neighboring IP address clusters, denoted in the text in the form $[firstIP, endIP]$.
- **City IP Address Group:** The second concept, referred to as a city IP address group, represents a cluster of IP addresses located in the same city.

We aim to construct a reference database based on the similarity between the delay distribution of network segments and the delay distribution of city IP address groups. Subsequently, we will conduct a comprehensive and effective evaluation of the IP geolocation database using this reference database. To address the challenges associated with the swift and efficient construction of the reference database, the absence of city information for certain IP addresses, and the low accuracy of city information for IP addresses, CDCDB introduces two major enhancements. Figure 5 illustrates the architecture of the method for constructing the reference database CDCDB.

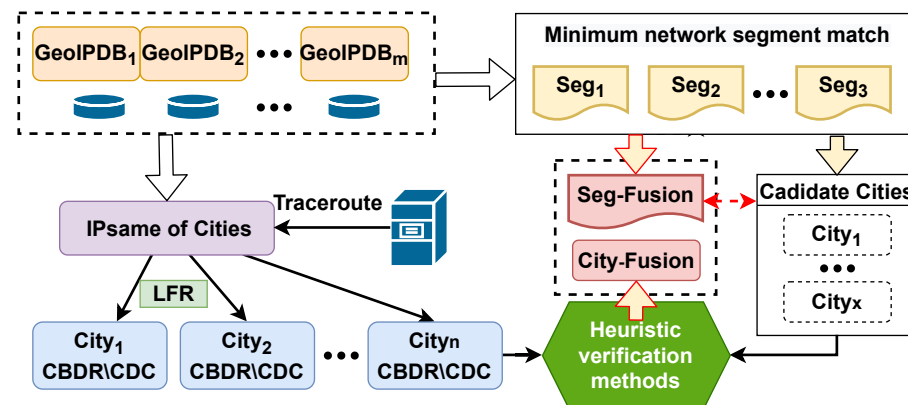


Figure 5. Architectural diagram of the construction methodology of the reference database CDCDB.

To mitigate issues such as the difficulty in rapidly constructing the reference database and missing city information for some IP addresses, we propose an approach based on the minimum network segment matching mechanism. We determine city location information of IP addresses based on network segments, thereby enhancing the efficiency of building reference databases. Then, we enhance the number and quality of segments in the reference database CDCDB by slicing segments with overlapping addresses between each database. We also improve city coverage of IP addresses by retaining candidate city information.

To tackle the issue of low accuracy of city location information for IP addresses, we initially utilize the minimum network segment matching mechanism to increase the likelihood that IP addresses within the same network segment are located in the same region. Then, we leverage the similarity between delay distribution of network segments and delay distribution of city IP address groups to verify locations of network segments where multiple candidate cities exist. Through a heuristic location verification algorithm based on city delay characteristics, we effectively assist network segments with location conflicts in determining city location information, thereby resolving the issue of low accuracy of city location information for IP addresses.

4.2. Minimum Network Segment Matching Mechanism

We propose the amalgamation of location information from various IP geolocation repositories to construct a reference database. By integrating multiple IP geolocation repositories, we can acquire as much location information of IP addresses as possible, thereby

facilitating the construction of a more comprehensive and accurate reference database. We denote this reference database as CDCDB (City Delay Characteristics DataBase). Given that IP geolocation databases contain a vast amount of data, and different databases provide varying numbers of network segments, different qualities of network segments, and different reliability levels of IP address location information, the primary challenge we need to address is how to efficiently merge multiple IP geolocation databases. We aim to ensure that the IP addresses have high city coverage and that the location information of the IP addresses is highly accurate.

In Section 3.2, we address the feasibility of a network segment-based scheme to improve IP geolocation accuracy. In an intuitive sense, the fewer the number of IPs within a segment, the higher the quality of that segment. This also increases the probability that the IP addresses within the segment are located in the same region. Upon analyzing the number of network segments contained in the experimental IP geolocation databases, along with the quality of these segments, we observed a noticeable overlap of assigned network segments among multiple databases. To address this issue of overlapping addresses, we propose a method based on a minimum segment matching mechanism. This method not only expands the number of segments in the reference database but also enhances segment quality by slicing mid-stacked address blocks.

The minimum network segment matching mechanism operates by dissecting network segments with overlapping addresses across multiple IP geolocation databases. This process involves dividing larger network segments into smaller ones and subsequently annotating each segment with potential location information as a candidate location. This approach serves a dual purpose. Firstly, it refines the network segments, thereby enhancing the likelihood that IP addresses within the same network segment are localized in the same region, i.e., it improves the quality of the network segments. Secondly, by remarking all potential location information of the network segment, the city coverage of the segment can be augmented. Concurrently constructing the reference database based on network segments can significantly expedite the construction efficiency of the entire reference database and reduce the volume of data requiring location verification. A schematic representation of the minimum net segment allocation mechanism is depicted in Figure 6.

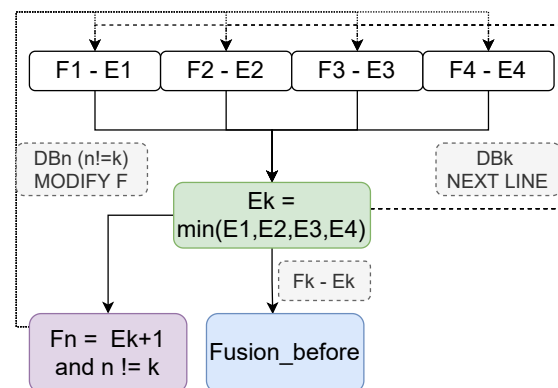


Figure 6. Schematic diagram of the minimum network segment matching mechanism.

Initially, we arranged network segments of all the IP geolocation databases in ascending order and uniformly converted the expression form of network segments into the format $[FirstIP, EndIP]$. Subsequently, we selected one network segment from each of the databases in sequence post-sorting and executed a “smallest network segment” comparison. In the first round of comparison, all databases provided the first sorted network segment, hence all segments involved in the comparison shared the same *FirstIP*. We designated the network segment with the smallest *EndIP* in this round of comparison as the fusion network segment, recorded the city information of this segment across all databases as the candidate location of this fusion network segment, and subsequently stored the fusion network segment in the “Fusion-before” database. Following this, we selected the next

segment from the database that provided the fusion network segment for the subsequent round of comparison. The remaining network segments were sliced from the smallest *EndIP*, and the residual portion was formed into a new segment for the next round of comparison. This process was repeated until all segments had been compared.

Utilizing the minimum segment matching mechanism, we construct the Fusion-before database. The network segments within this database may concurrently contain multiple candidate cities, necessitating the verification of network segment locations with multiple candidate cities. If a fusion network segment possesses only a single candidate city or lacks any candidate city, we directly deposit the segment into the “Fusion” database, thereby completing the final processing of the reference database construction.

4.3. City Delay Characteristics

For a converged network segment with multiple candidate cities, it becomes necessary to discern the true location city of the segment from among the multiple candidates. However, prior to executing the location verification, we must first ascertain the delay characteristics of the network segments of the candidate cities. In Section 3.3, we explore the feasibility of determining IP address geolocation schemes based on delay distribution, and in our analysis in Section 3.1, we also discover that different IP geolocation inventories exhibit certain city preferences. Consequently, we can verify the location of network segments with multiple candidate cities in the Fusion-before database from the perspective of city delay distribution.

Our objective is to discern the delay characteristics of cities, yet accurately obtaining the delay information of cities and determining their delay distribution model pose significant challenges. Figure 7 illustrates our frequency–delay graph from the London node for four cities: London, New York, Beijing, and Singapore. In contrast to the three cities closer to London in Figure 4 (Paris, Brussels, and Rotterdam), the three cities further from the London node in Figure 7 (New York, Beijing, and Singapore) exhibit a large RTT mode and a broad distribution of RTTs, as well as the phenomenon of RTT outlier nodes and RTT clusters. Their delay distributions are so irregular that they cannot be adequately represented by any mathematical model.

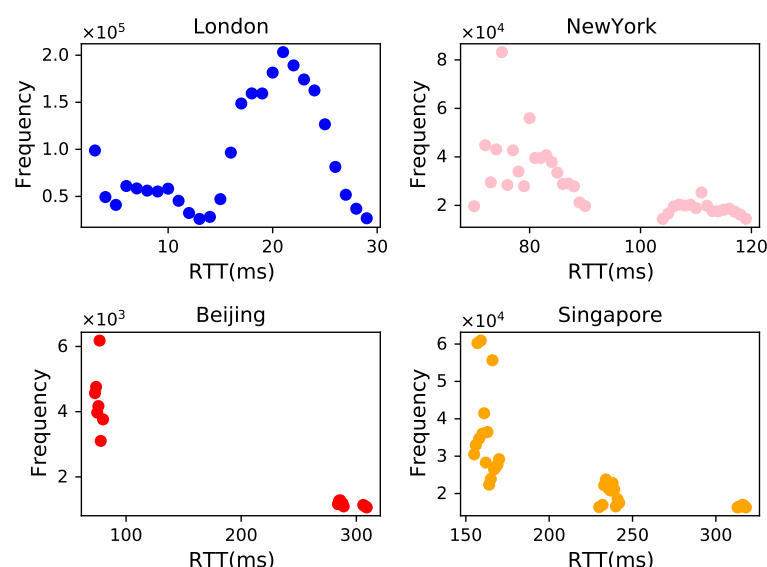


Figure 7. Frequency–delay diagrams for London, New York, Beijing, Singapore (London).

Figures 3, 4, and 7 demonstrate that when the source IP is proximate to the target IP, the size of the RTTs still exhibits a linear pattern with distance, and the city’s delay distribution pattern is also evident. Conversely, this law does not hold when the source IP is significantly distant from the target IP. This suggests that only when the target IP is closer

to the original IP does the obtained delay information prove relatively accurate and the city's delay distribution regular. Therefore, only cities closer to the probe node can yield accurate and centralized delay information, and only the delay distribution model obtained based on this accurate delay information can be utilized for geolocation of IP addresses. Consequently, we propose a distance-based collection point assignment method to gather city delay information.

The principle of the distance-based acquisition point allocation method is depicted in Figure 8. Initially, we select multiple probe nodes globally for delay detection. Subsequently, we calculate the Nearest Haversine Distance [37] from the target city to different probe points based on latitude and longitude, selecting the probe point with the shortest distance as the city's probe point. Finally, the specific probe node conducts delay detection on the IP address of the assigned city to obtain the corresponding city delay information.

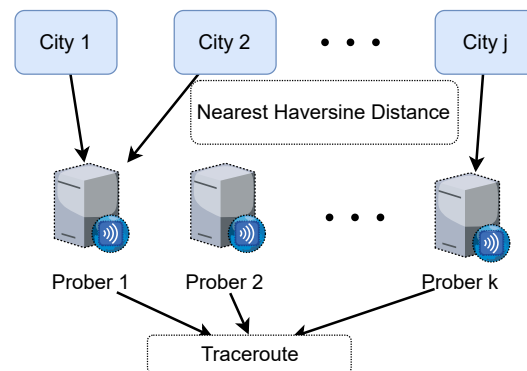


Figure 8. Distance-based collection point assignment method.

In the process of constructing the city's delay characteristics, we encounter two significant challenges. Firstly, how do we identify credible IP addresses that are localized to the target city? Secondly, how do we construct a reliable city delay distribution model from the obtained delay information?

To effectively obtain trusted IP addresses located in the target city, we propose a dynamic comparison mechanism based on multiple databases. As observed in Section 3.1, different IP geolocation databases often contain many identical IP addresses within the same city. These location-consistent IP addresses are more trustworthy than other IP addresses with location conflicts. Consequently, we treat all IP addresses that are locationally identical across multiple IP geolocation databases as trusted IPs. These are then divided by city to obtain a set of trusted IP addresses for a city, referred to as the set of city IPs (IPsame). If an IP address lacks city information in a certain database, then that database is not involved in the comparison. This dynamic comparison mechanism based on multiple databases maximizes the number of trusted IP addresses.

Among the 42 experimental cities we selected, the average total number of city IPsame datasets obtained based on three experimental IP geolocation databases is 7.95 million (excluding IP2Region), while the average number of trusted IP addresses obtained based on four experimental databases exceeds 2.54 million. Therefore, our delay characterization of cities based on the city IPsame dataset is sufficiently objective and reliable. The dynamic comparison mechanism based on multiple databases is depicted in Figure 9.

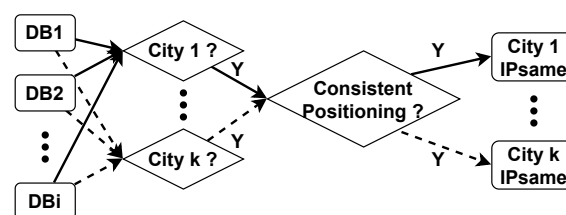


Figure 9. Schematic diagram of dynamic comparison mechanism based on multiple databases.

We utilize the delay information of the city IPsame dataset, obtained from the special detection points, as the city's delay dataset. However, we observe that even the delay information derived from the city IPsame dataset is broadly distributed and contains numerous RTT outlier nodes. We attribute this phenomenon to potential interference from network firewalls or network congestion, resulting in a small number of IP addresses with distorted feedback delay. The extensive distribution range of the city delay dataset is not conducive to discerning the city's delay characteristics and also impacts the computational efficiency of these characteristics. Therefore, it becomes necessary to perform data preprocessing on the collected raw time-delay data.

To address the issue of the extensive distribution range of the city delay dataset, we employ a "Low-frequency rejection" method for data cleansing. Although different cities exhibit distinct delay distributions, the delay distribution range for each city should be concentrated within a certain value range. RTT values that are excessively small or large may indicate distortion in delay information. Consequently, we propose that the frequency of occurrence of an RTT signifies the credibility of that RTT: an RTT with a higher frequency of occurrence is less likely to be distorted, while an RTT with a lower frequency of occurrence is more likely to be distorted. We compute the frequency value of each RTT based on the urban delay dataset. We eliminate extreme data with excessively low frequency, a process referred to as "Low-frequency rejection". By employing "Low-frequency rejection", we can significantly reduce the distribution range of delay information, enhance computational efficiency, and simultaneously retain the primary characteristics of urban delay. Specifically, RTTs with a delay frequency lower than $1/a$ are rejected. Utilizing the "Low-frequency rejection" method, the reliability of urban delay characteristics can be substantially improved. We will discuss the related parameter a in the subsequent section.

Finally, we construct the corresponding city delay characteristics. We arrange the reliable RTT dataset of the city in ascending order to establish the city's boundary delay range, denoted as $CBDR = [t_1, t_2, \dots, t_n]$. Concurrently, we utilize the frequency of occurrence of each RTT in $CBDR$ as a feature of the RTT. For any city delay information t_k , assuming its frequency of occurrence is p , the $RTT = t_k$ is characterized by p , denoted as $\{t_k : p\}$. Then, based on the city's boundary delay range and the frequency of occurrence, we derive the city delay characteristics: $CDC = [\{t_1 : p_1\}, \{t_2 : p_2\}, \dots, \{t_n : p_n\}]$.

4.4. Geolocation Verification Mechanism

As discussed in Section 3.3, we propose an evaluation strategy based on delay distribution. We argue that IP geolocation algorithms solely based on network delay are susceptible to errors, while determining the geographic location of IP addresses based on the network delay distribution of the region proves more reliable. Our objective is to geolocate the network segments in the Fusion-before database where multiple candidate cities exist. Therefore, upon obtaining the city delay characteristics, we propose a location verification method predicated on city delay characteristics.

We obtain the delay information of all IP addresses in a network segment through active probing and derive the network segment delay distribution of the candidate city by analogy with the delay characteristics of the candidate city. We then express the confidence that the candidate city is the real location of the network segment with the eigenvalues of the net segment delay characteristics of the candidate city. Consequently, we select the candidate city with the highest network segment delay feature value as the actual location city of the network segment. The flowchart of the location verification method based on city delay features is depicted in Figure 10. We partition the geolocation verification of network segments into four distinct stages.

Firstly, we measure the delay information of the network segment based on the candidate cities of the segment. Given that different candidate cities may select different probing nodes, we choose an appropriate probing node to perform delay probing on the network segment based on the candidate cities. We then perform traceroute probing on all IP addresses in the network segment from the corresponding probing point and record the

RTT information of the last hop of the routing path fed back by each IP address. The RTT information fed back by all IP addresses is aggregated into the network segment delay: $NET = \{N_{t1}, N_{t2}, \dots, N_{tm}\}$.

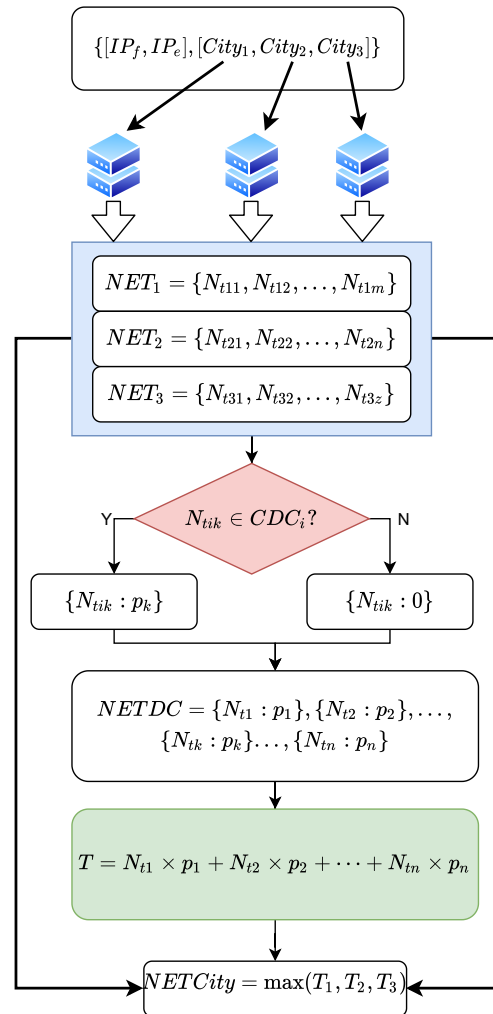


Figure 10. Schematic diagram of the principle of location verification method based on city delay characteristics.

Secondly, we compare the network segment delay with the corresponding candidate city delay characteristic CDC. For a candidate city's network segment delay information $NET = \{N_{t1}, N_{t2}, \dots, N_{tm}\}$, if any $RTT = N_{tk}$ is located in a candidate city with $CDC = [\{t_1 : p_1\}, \{t_2 : p_2\}, \dots, \{t_n : p_n\}]$, we then record the feature p_k of N_{tk} in the CDC into the corresponding segment delay as the segment delay feature of the candidate city $NETDC = [\{N_{t1} : p_1\}, \{N_{t2} : p_2\}, \dots, \{N_{tm} : p_m\}]$. If $RTT = N_{tk}$ does not exist in the CDC of the candidate city, its characteristics are defaulted to $p_k = 0$.

Next, we compute the network segment delay characteristics value T of the candidate city. For the candidate city j to be determined in the net segment, its corresponding net segment delay characteristics are denoted as $NETDC_j = [\{N_{jt1} : p_{j1}\}, \{N_{jt2} : p_{j2}\}, \dots, \{N_{jtm} : p_{jm}\}]$; we define the delay eigenvalue T_j of this candidate city as $T_j = N_{jt1} \times p_{j1} + N_{jt2} \times p_{j2} + \dots + N_{jtm} \times p_{jm}$.

Finally, we determine the real location of the network segment. We complete the location verification by taking the candidate city with the largest value of the network segment's delay feature as the actual location city of the network segment. The network segment after location verification is stored in the Fusion database to complete the construction of the reference database CDCDB.

To enhance the credibility of the candidate cities of a network segment, we propose a method to extend the candidate cities of a network segment. Yong Gan et al. suggested that the geolocation of an IP address can be determined by a sequence of neighboring IPs [33]. Therefore, we propose to extend the list of candidate cities of a network segment based on the nearest neighbor matching mechanism of a network segment. Specifically, we compare the network segment to be verified with the network segments in the Fusion database, identify the network segment from the Fusion database that is the closest neighbor to the network segment to be verified, and add the localized city of that segment into the list of candidate cities for the network segment to be determined. Figure 11 provides a schematic diagram of the candidate city expansion method based on the nearest neighbor network segment matching mechanism.

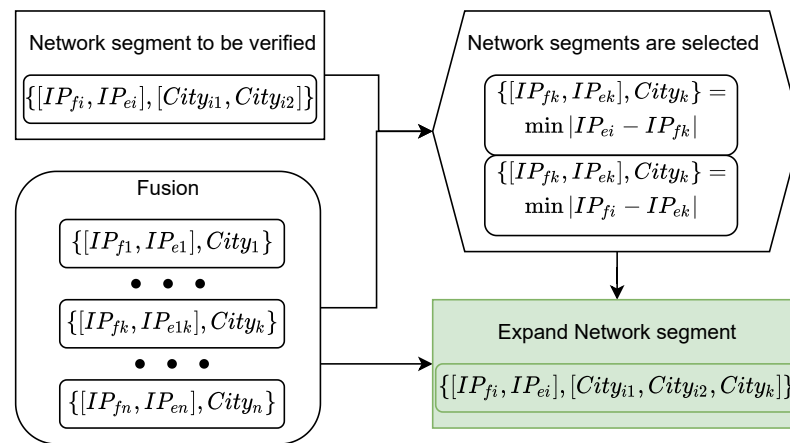


Figure 11. Schematic diagram of the candidate city expansion method based on nearest neighbor network segments.

For any segment $NET_i = \{[IP_{fi}, IP_{ei}], [City_{i1}, City_{i2}]\}$ to be verified, we search for the nearest neighbor NET_k to NET_i from the Fusion database. $NET_k = \{[IP_{fk}, IP_{ek}], City_k\}$ in the Fusion database is the nearest neighbor of NET_i if NET_k satisfies the condition $|IP_{ei} - IP_{fk}| = \min |IP_{ei} - IP_{fj}|$ or $|IP_{fi} - IP_{ek}| = \min |IP_{fi} - IP_{ej}|$, where $NET_j = \{[IP_{fj}, IP_{ej}]\}$ is any network segment in the Fusion database. If NET_k is the closest neighbor of NET_i in the Fusion database, we then add the localized city $City_k$ to the list of candidate cities of NET_i to complete the candidate city expansion, resulting in $NET_i = \{[IP_{fi}, IP_{ei}], [City_{i1}, City_{i2}, City_k]\}$.

5. Experimentation and Evaluation

The experimental and evaluative components of our study are organized into two distinct sections. The first section delves into the relevant parameters involved in computing urban delay features. The second section provides an overview of the fundamental aspects of the reference database CDCDB, accompanied by a discussion on the validity of the CDCDB. Then, we evaluate the reliability of four experimental IP geolocation libraries utilizing the reference database CDCDB as a benchmark.

5.1. Discussion of Parameter a

In order to enhance the reliability of urban delay features, we introduced the “Low-frequency rejection” data cleaning method in Section 4.3. In this section, we explore the parameter a in “Low-frequency rejection” to determine the optimal cleaning scheme.

Reliable city boundary delay ranges are typically characterized by a high frequency of RTTs and a highly concentrated distribution of RTTs. The more concentrated the distribution range of city delay, the more pronounced the characteristics of delay distribution in the city. Our objective is to eliminate as much distorted RTT information as possible, while retaining as many original delay data points as possible, in order to discover the delay

characteristics of the city as comprehensively and accurately as possible. Therefore, we take the data retention rate as the primary measurement index for the selection of parameter a . the smaller the range of delay distribution of the city, the more pronounced the delay characteristics of the city; the higher the retention rate of raw data, the more comprehensive the delay characteristics of the city.

Table 2 presents the results of the experiments conducted in Shanghai following several data cleansing sessions. Prior to data cleaning, the latency data distribution range for Shanghai was 112,672 milliseconds. However, after applying the “Low-frequency rejection” method, the delay range for Shanghai was significantly reduced. When $a = 100$, the delay range is reduced to 49 milliseconds (ms), with a retention rate of 66%; when $a = 200$, the delay range expands to 79 ms, but the retention rate increases to 90%. When we set the parameter a to 1000, the distribution range of delay increases to 101 ms, but the data retention rate reaches 95%. The experimental results demonstrate that the data cleaning method based on “Low-frequency rejection” can effectively reduce the delay distribution range of the city, minimize the impact of distorted RTT information on the delay characteristics of the city, and enhance the reliability of the delay range of the city boundary.

Table 2. Data cleaning results based on “Low-frequency rejection” (Shanghai).

Parameter	Delay Range (ms)	Proportion of Remain IPs
$a = 100$	49	66%
$a = 200$	79	90%
$a = 1000$	101	95%

We further analyzed the experimental results of data cleaning based on “Low-frequency rejection” for 42 large cities, as detailed in Table 3. The average latency ranges for the 42 cities are 26.78 ms, 40.21 ms, and 72.57 ms, while the data retention rates are 74.5%, 83.8%, and 91.4% for the values of parameter a of 100, 200, and 1000, respectively. The overall delay distribution and data retention ratio are similar to that of Shanghai. Therefore, we posit that when the delay distribution of a city ranges around 100 ms and the data retention ratio exceeds 90%, the remaining RTT data can effectively reflect the delay characteristics of the city. Consequently, we ultimately choose to make parameter $a = 1000$.

Table 3. Data cleaning results based on “Low-frequency rejection” (42 cities).

Parameter	Delay Range (ms)	Proportion of Remain IPs
$a = 100$	26.78	74.5%
$a = 200$	40.21	83.8%
$a = 1000$	72.57	91.4%

5.2. Analysis and Evaluation

Leveraging the minimum segment matching mechanism and the city delay characteristics, we construct the reference database CDCDB for evaluating the IP geolocation repository. CDCDB comprises 16.27 million network segments, of which 6.38 million segments with unique candidate cities account for 39.1% of all segments in the database. Consequently, we only need to verify the location of 9.89 million network segments where multiple candidate addresses exist. Compared with the traditional IP address-based method to construct the reference database, our proposed method improves computational efficiency nearly a hundredfold.

In the CDCDB database, there are a mere total of 1.269 million non-high-quality network segments with prefix length no greater than /24, constituting 7.8% of the number of all network segments, and the coverage of city-level network segments reaches 99.99%. Compared with IPUU, the number of network segments in CDCDB increases by 18.7%,

the number of high-quality network segments increases by 13.2%, and the coverage of city-level network segments increases by 20.92%, as shown in Table 4. This demonstrates that our proposed minimum segment matching mechanism indeed expands the number of segments, improves segment quality and city-level coverage, and is crucial for enhancing city-level localization accuracy of IP addresses.

Table 4. Basic information of CDCDB and IPUU.

Database	NO. of Net-Segments	High-Quality Network Segments	City Coverage
CDCDB	16,274,675	15,005,250	99.99%
IPUU	13,704,888	10,826,862	79%

We evaluated the accuracy of CDCDB in 42 cities using the macro Internet Topology Data Kit (ITDK) provided by CAIDA [38]. This kit amalgamates well-known Internet exchange point information, Hoiho hostname mapping, and city granularity GeoLite2 geolocation data freely available from MaxMind. After filtering out the location information provided by MaxMind, we verified the geolocation accuracy of the IP addresses of the 42 cities in the CDCDB. We found that the average accuracy is as high as 94%. The results are depicted in Figure 12.

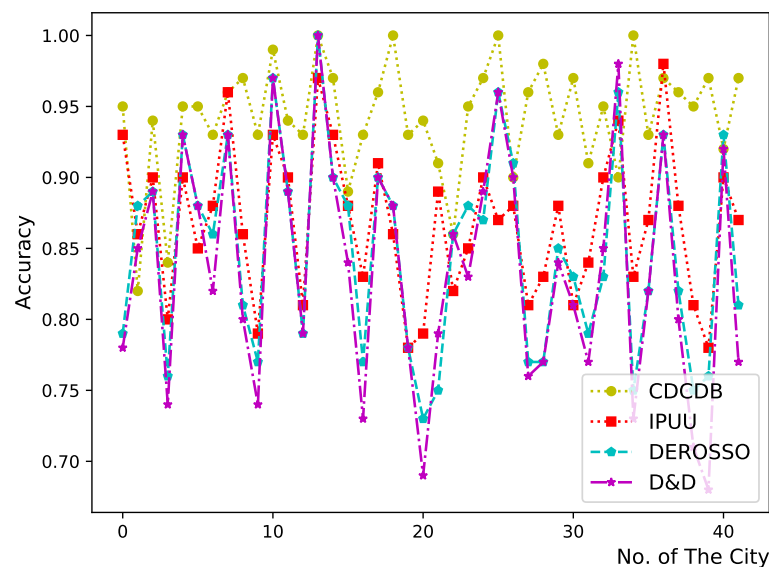


Figure 12. Schematic representation of accuracy validation results for multiple models.

We also evaluated the geolocation accuracy of the DEROSSO model proposed by H Li et al. [18] and the D&D model proposed by X Bo et al. [19], as well as IPUU, in 42 experimental cities. As depicted in Figure 12, we find that the overall accuracy of both DEROSSO and D&D is lower compared to CDCDB, indicating that our proposed evaluation method effectively enhances the accuracy of the reference database. We also observe that the accuracy curves of DEROSSO and D&D models follow almost the same trend as that of IPUU, and do not demonstrate any superiority over the original IP geolocation database IPUU. This suggests that the reference databases constructed by DEROSSO and D&D models are highly dependent on the IP geolocation database IPUU and do not significantly improve geolocation accuracy. Conversely, the overall accuracy of CDCDB is significantly improved compared to that of IPUU, indicating that our constructed reference database effectively improves city-level geolocalization results of IP addresses and is less dependent on the original IP geolocation database.

The results depicted in Figure 12 suggest that we conclude that the reference database CDCDB, constructed through the minimum segment matching mechanism and city delay characteristics, has highly reliability in evaluating IP geolocation databases. Furthermore, this reliability increases as the number of integrated IP geolocation databases increases.

Lastly, we assessed the reliability of the four experimental IP geolocation libraries utilizing the reference database CDCDB. As identified in Section 3.1, IP geolocation libraries exhibit varying preferences for different cities, implying that the reliability of geolocation information provided by IP geolocation libraries differs across cities. Next, we validated this finding through the reference database.

We assess the reliability of the IP geolocation database in a specific city by selecting the city to be evaluated. We then compared the IP geolocation database under evaluation with the reference database, CDCDB, and calculated the data similarity to gauge the reliability of the IP geolocation database in that city. Specifically, we evaluated the reliability of the IP geolocation repository in a given city using three metrics: Database Accuracy (*DA*), Database Recall (*DR*), and Database Reliability Score (*DRS*). The definitions for these three metrics are as follows:

Consider a scenario where the quantity of IP addresses, which share the same localization between the IP geolocation database and the reference database CDCDB within a specific city, is denoted as IP_{sim} . The total number of IP addresses localized in that city by the IP geolocation database is represented as IP_{GeoDB} , and the quantity of IP addresses localized in that city by the CDCDB is symbolized as IP_{CDCDB} .

1. **Database Accuracy (*DA*):** This metric represents the proportion of IP addresses whose geolocation aligns with both the IP geolocation database and the reference database CDCDB, relative to the total number of IPs assigned to the IP geolocation database in that city. Mathematically, it can be expressed as:

$$DA = \frac{IP_{sim}}{IP_{GeoDB}} \quad (1)$$

2. **Database Recall (*DR*):** This measure signifies the ratio of the number of IPs with consistent geolocation between the IP geolocation repository and the reference database CDCDB, to the total number of IPs assigned to the reference repository in that city. It can be formulated as:

$$DR = \frac{IP_{sim}}{IP_{CDCDB}} \quad (2)$$

3. **Database Reliability Score (*DRS*):** This score is a weighted reconciled average of both the database accuracy (*DA*) and database recall (*DR*) of the IP geolocation repository. Mathematically, it can be expressed as:

$$DRS = \frac{1}{1 + \beta^2} \cdot \left(\frac{1}{DA} + \frac{\beta^2}{DR} \right) \quad (3)$$

Database Accuracy (*DA*) serves as a reflection of the reliability of IP addresses provided by IP geolocation databases that are situated in a specific city. A higher *DA* signifies a greater degree of reliability of the IP information furnished by the database for that city. On the other hand, Database Recall (*DR*) mirrors the comprehensiveness of the IP addresses provided by the IP geolocation database in a particular city. A higher *DR* indicates a more extensive coverage of IP addresses by the IP geolocation database in that city. *DA* and *DR* encapsulate the accuracy and comprehensiveness of IP information provided by IP geolocation databases separately. Yet, we lack a comprehensive assessment indicator. Therefore we proposed the Database Reliability Score (*DRS*) as a comprehensive analytical indicator, which amalgamates *DA* and *DR*. This score provides a holistic measure of the performance of the database, thereby facilitating more informed decision-making for users leveraging these databases. In this study, we introduce a parameter, β , which serves as a

measure of the relative importance of database accuracy (*DA*) versus database recall (*DR*). The value of β influences the balance between these two metrics in the following manner:

- When $\beta = 1$, it signifies that *DA* and *DR* hold equal importance.
- When $\beta > 1$, it indicates a greater emphasis on *DR*.
- Conversely, when $\beta < 1$, *DA* is given more weight.

In the context of this paper, we assign a higher priority to database accuracy. This decision is driven by the expectations of IP geolocation database users, who generally anticipate that the positioning information provided by these databases will be highly accurate. Consequently, we set the parameter β to 0.5, thereby giving more influence to *DA* in our analysis.

In this study, we selected 12 cities across China, Europe, and the United States to evaluate the city-specific performance of the IP geolocation repository. Figure 13 presents a schematic representation of the database accuracy and database recall results for four experimental IP geolocation libraries across these 12 cities.

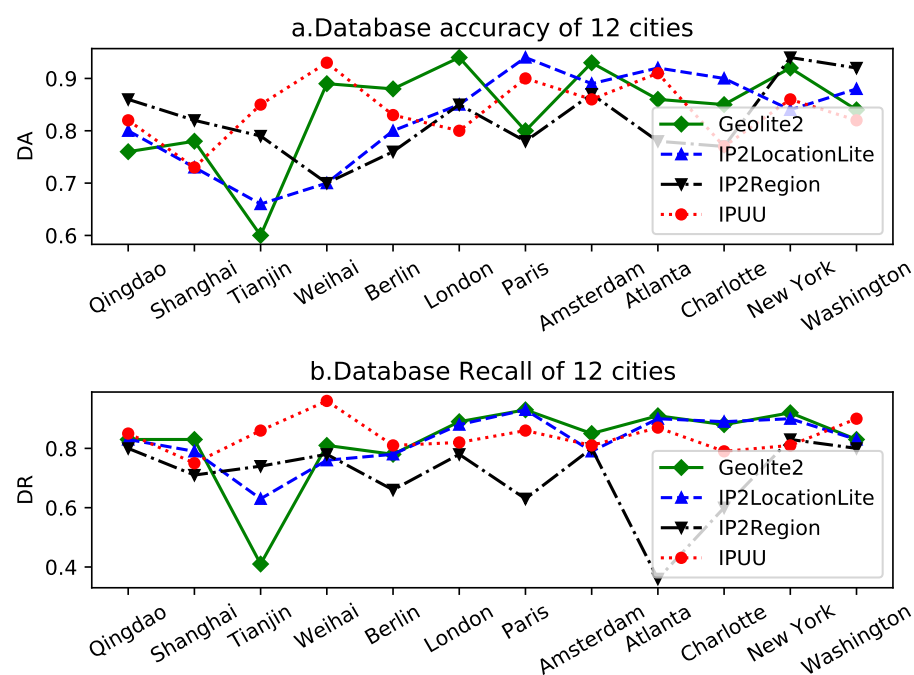


Figure 13. (a) Schematic of the database accuracy results for the experimental data. (b) Schematic of the database recall results for the experimental database.

The results delineated in Figure 13 illustrate that all IP geolocation libraries exhibit exemplary performance in the majority of cities, with high scores for both data accuracy and data recall. This indicates that these four experimental IP geolocation libraries can confidently localize IP addresses in most cities. Interestingly, we observed that the database accuracy and database recall curves of the IP geolocation database generally follow similar trends in most cities, underscoring the impressive results achieved by this IP geolocation database in city-level geolocation endeavors.

However, in a select few cities, the *DA* and *DR* of the IP geolocation database do not exhibit congruence. For instance, in Atlanta, the IP2Region database boasts a *DA* of 78% and a *DR* of 36%. This significant difference implies that in Atlanta, there is a substantial discrepancy between the IP information provided by IP2Region and that provided by the CDCDB: IP2Region provides only a limited number of IP addresses located in Atlanta, but the vast majority of these IP addresses are locationally accurate IPs. While there is a similar trend observed between the database accuracy curve and the database recall curve, a closer examination reveals distinct differences. The data recall curve clearly indicates

that the IP2Region database is less proficient at geolocating cities in Europe and the US. However, this conclusion is not as evident when examining the database accuracy curve. This discrepancy underscores the existence of differences between the reference database CDCDB constructed based on the minimum segment matching mechanism and city delay features and the original IP geolocation database. The reference database improves the original IP geolocation database, and the degree of reliance on the original database varies depending on the city of location.

The close alignment between database accuracy and database recall for the IP geolocation database signifies a high degree of congruence between the IP geolocation database and the reference database. This observation underscores the broad reliability of current IP geolocation databases for city-level geolocation endeavors. However, there are also instances where the *DA* and *DR* of the IP geolocation repository diverge significantly in a few cities. This situation validates the reference databases we constructed on one hand, and on the other hand, it also reveals that the same IP geolocation database may perform differently in different cities.

We further scrutinize the performance of different IP geolocation databases in the same city. From Figure 13, it can be discerned that the performance of different IP geolocation databases does not differ significantly in most cities, and the major differences are confined to a few cities. For instance, in Tianjin, the database recall and database accuracy of GeoLite2 are significantly lower than other databases. The results of this experiment demonstrate that the performance of different IP geolocation databases in the same city varies; though, in most cities, the differences between different IP geolocation databases are marginal. Combined with our previous finding that the same IP geolocation database also has reliability differences in different cities, we can draw a potential conclusion that IP geolocation databases exhibit a certain degree of city bias, and that the reliability of IP information provided by different IP geolocation databases varies according to the city in which they are located.

Figure 14 presents the database reliability scores of the four experimental IP geolocation libraries across 12 specific cities. A significant observation from Figure 14 is that different IP geolocation libraries exhibit different reliability in locating different cities. Overall, IPUU outperforms the others, demonstrating a high overall database reliability score with minimal fluctuations. IP2LocationLite follows, leading in cities such as Qingdao, Shanghai, Paris, Atlanta, and Charlotte, despite its subpar performance in Tianjin and Weihai. GeoLite2 experiences considerable fluctuations in its database reliability score due to its poor performance in Tianjin. The least effective overall is IP2Region, with a score significantly lower than the other three databases. Interestingly, Figure 14 does not reveal any specific country characteristics for the cities that the IP geolocation databases excel at locating. Instead, their strengths and weaknesses appear to be more city-specific.

We computed the mean and variance of the database reliability scores of the four experimental IP geolocation databases, as delineated in Table 5. In Table 5, we observe that IPUU, GeoLite2, and IP2LocationLite exhibit similar average *DRS* scores, with IPUU achieving the optimal result with a *DRS* score of 84%. This suggests that, on the whole, the performance of each IP geolocation database is comparable for city-level geolocation work, with the exception of IP2Region, which performs slightly subpar.

Table 5. The mean and variance of *DRS* score for experimental IP geolocation database.

	GeoLite2	IP2LocationLite	IP2Region	IPUU
Average	83.3%	83.2%	78.99%	84%
Variance	0.009	0.006	0.0063	0.0028

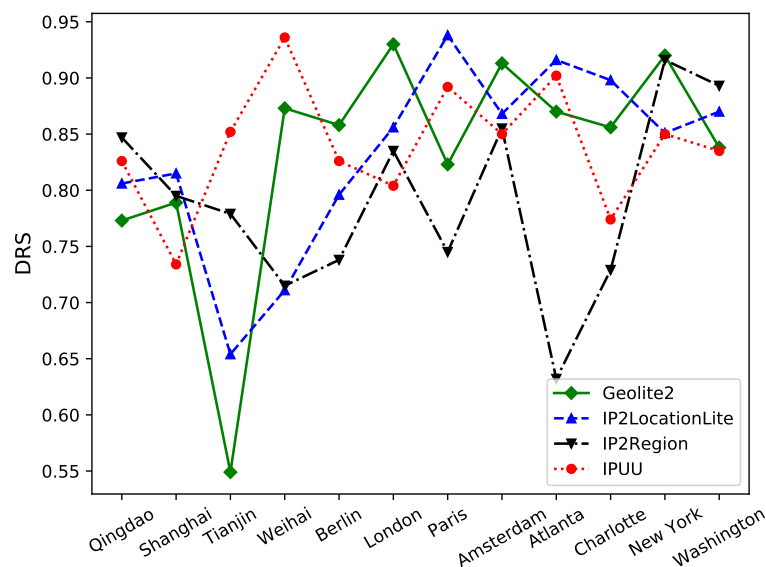


Figure 14. Schematic of the database reliability score results for the experimental data.

However, the *DRS* variances of the four databases exhibit noticeable differences. GeoLite2, which has the largest *DRS* variance, is 3.2 times larger than IPUU, which has the smallest *DRS* variance. The *DRS* variance of IP2LocationLite is akin to that of IP2Region, with a minor difference of 0.0003 only. In terms of *DRS* variance, IPUU's location results are more stable, indicating that it is a database that can provide stable location results more independent of the city where it is located. Conversely, GeoLite2 exhibits more city preference, and in some of the "bad" cities, the positioning information provided by it may be suboptimal.

Of course, since IPUU is a paid database, we cannot ascertain whether the *DRS* variance difference is due to the difference between free and paid databases, so we refrain from evaluating the overall quality of the databases. However, from the performance of the three free databases, the generally high *DRS* mean and generally small *DRS* variance demonstrate that the IP geolocation database has been relatively perfected in city-level geolocation, but there are still a few "bad" regions. Such regional geolocation weaknesses will be the main reason for IPUU's success in city-level geolocation. This kind of regional geolocalization weakness will be the focus of IP geolocalization database in city-level geolocalization work.

Figures 13 and 14 collectively indicate that different IP geolocation repositories do indeed exhibit city preferences. While the reliability of different IP geolocation databases within the same city is often similar, significant performance differences emerge in a few cities. This suggests that the primary distinctions between IP geolocation databases lie at the city level. In certain specific cities, different IP geolocation databases provide markedly different services. Therefore, we recommend online service providers to select an appropriate IP geolocation database based on the cities they serve to enhance service quality.

6. Conclusions

In response to the absence of an accurate and comprehensive method to evaluate the quality of commercial IP geolocation databases, we introduce a novel database evaluation method—the evaluation of IP geolocation databases predicated on urban delay characteristics. Initially, we encapsulated three evaluation strategies based on previous work as the theoretical foundation for database evaluation. Subsequently, we validated and optimized the IP address localization cities by employing the minimum segment matching mechanism and the similarity between the segment delay distribution and the city delay distribution, culminating in the construction of the reference fusion database CDCDB. The city-level localization accuracy of CDCDB reaches an average of 94%, and the city

coverage rate attains 99.99%, providing the foundation for the effective and comprehensive evaluation of IP geolocation databases. Compared with traditional evaluation methods, our algorithm exhibits significant advantages in both evaluation accuracy and database construction efficiency.

Finally, we evaluated four IP geolocation databases using the reference database CD-CDB. The evaluation results reveal that different IP geolocation databases exhibit different city preferences, i.e., different IP geolocation databases excel at locating for different cities. Also, for the same city, the reliability of IP location information provided by different IP geolocation databases varies. However, these preferences do not extend to the national level; instead, the strengths and weaknesses of IP geolocation databases appear to be more city-specific. Therefore, we posit that the future focus of city-level geolocation efforts by IP geolocation databases should be on addressing “weak” cities and reducing the number of cities that are “not good” at geolocation. Concurrently, we also suggest that online service providers should select appropriate IP geolocation databases according to the cities they serve in order to enhance service quality.

Author Contributions: Y.X.: conceptualization, methodology, writing—original draft preparation, validation, visualization; Z.Z.: supervision, project administration; N.L.: writing—reviewing and editing, supervision; Y.L.: software, data curation, validation; E.C.: investigation, data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Shandong Province Key R&D Project 2020CXGC10103; the National Natural Science Foundation of China (62101159) and the Natural Science Foundation of Shandong Province (ZR2021MF055).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/YuanchengXie/-Fusion-IP-GeoDatabase> (accessed on 26 October 2023).

Conflicts of Interest: There are no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CDCDB	City delay characterization database
CBDR	City boundary delay range
CDC	City delay characteristics
NETDC	Network segment delay characteristics
DA	Database Accuracy
DR	Database Recall
DRS	Database Reliability Score

References

- Li, Q.; Wang, Z.; Tan, D.; Song, J.; Wang, H.; Sun, L.; Liu, J. GeoCAM: An IP-Based Geolocation Service through Fine-Grained and Stable Webcam Landmarks. *IEEE/ACM Trans. Netw.* **2021**, *29*, 1798–1812. [\[CrossRef\]](#)
- Zhang, Y.; Deng, R.H.; Bertino, E.; Zheng, D. Robust and universal seamless handover authentication in 5G HetNets. *IEEE Trans. Dependable Secur. Comput.* **2019**, *18*, 858–874. [\[CrossRef\]](#)
- Liu, C.; Luo, X.; Yuan, F.; Liu, F. Rnbg: A ranking nodes based ip geolocation method. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 6–9 July 2020; pp. 80–84.
- Wang, Z.; Niu, Y.; Chen, H.; Cheng, G.; Cui, J.; Zhang, Z. Target driven IP Geolocation Algorithm. *J. Phys. Conf. Ser.* **2021**, *1861*, 012002. [\[CrossRef\]](#)
- Rodriguez Garzon, S.; Deva, B. Geofencing 2.0: Taking location-based notifications to the next level. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 921–932.
- MaxMind. GeoIP2. 2022. Available online: <https://www.maxmind.com/en/geoi2-databases/> (accessed on 30 May 2022).
- IP2Location. IP2Location. 2022. Available online: <https://www.ip2location.com/database/> (accessed on 30 May 2022).
- Callejo, P.; Gramaglia, M.; Cuevas, R.; Cuevas, A. A deep dive into the accuracy of IP Geolocation Databases and its impact on online advertising. *IEEE Trans. Mob. Comput.* **2022**, *22*, 4359–4373. [\[CrossRef\]](#)

9. Gouel, M.; Vermeulen, K.; Fourmaux, O.; Friedman, T.; Beverly, R. IP geolocation database stability and implications for network research. In Proceedings of the Network Traffic Measurement and Analysis Conference, Virtual, 14–15 September 2021.
10. Poese, I.; Uhlig, S.; Kaafar, M.A.; Donnet, B.; Gueye, B. IP geolocation databases: Unreliable? *ACM SIGCOMM Comput. Commun. Rev.* **2011**, *41*, 53–56. [\[CrossRef\]](#)
11. Gharaibeh, M.; Shah, A.; Huffaker, B.; Zhang, H.; Ensafi, R.; Papadopoulos, C. A look at router geolocation in public and commercial databases. In Proceedings of the 2017 Internet Measurement Conference, London, UK, 1–3 November 2017; pp. 463–469.
12. Du, B.; Candela, M.; Huffaker, B.; Snoeren, A.C.; Claffy, K. RIPE IPmap active geolocation: Mechanism and performance evaluation. *ACM SIGCOMM Comput. Commun. Rev.* **2020**, *50*, 3–10. [\[CrossRef\]](#)
13. Shavitt, Y.; Zilberman, N. A geolocation databases study. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 2044–2056. [\[CrossRef\]](#)
14. Jiang, H.; Liu, Y.; Matthews, J.N. IP geolocation estimation using neural networks with stable landmarks. In Proceedings of the 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), San Francisco, CA, USA, 10–14 April 2016; pp. 170–175.
15. Komosny, D.; Voznak, M.; Rehman, S.U. Location accuracy of commercial IP address geolocation databases. *Inf. Technol. Control* **2017**, *46*, 333–344. [\[CrossRef\]](#)
16. Saxon, J.; Feamster, N. GPS-based geolocation of consumer IP addresses. In *Passive and Active Measurement: Proceedings of the 23rd International Conference, PAM 2022, Virtual Event, 28–30 March 2022*; Springer: Cham, Switzerland, 2022; pp. 122–151.
17. Huffaker, B.; Fomenkov, M.; Claffy, K.C. *Geocompare: A Comparison of Public and Commercial Geolocation Databases*; Technical Report; Cooperative Association for Internet Data Analysis (CAIDA): San Diego, CA, USA, 2011.
18. Li, H.; Zhang, P.; Wang, Z.; Du, F.; Kuang, Y.; An, Y. Changing IP geolocation from arbitrary database query towards multi-databases fusion. In Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 3–6 July 2017; pp. 1150–1157.
19. Bo, X.; Han, L.; Yong, W. An IP geolocation database evaluation and fusion model based on data correlation and delay similarity. In Proceedings of the 2nd International Conference on Telecommunications and Communication Engineering, Beijing, China, 28–30 November 2018; pp. 231–236.
20. Zu, S.; Luo, X.; Zhang, F. IP-geolocator: A more reliable IP geolocation algorithm based on router error training. *Front. Comput. Sci.* **2022**, *16*, 161504. [\[CrossRef\]](#)
21. Zu, S.; Luo, X.; Du, S.; Wang, L. A delay deviation tolerance IP geolocation method with error estimation. *Sci. Rep.* **2022**, *12*, 13919. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Ma, Z.; Zhang, S.; Li, N.; Li, T.; Hu, X.; Feng, H.; Zhou, Q.; Liu, F.; Quan, X.; Wang, H.; et al. GraphNEI: A GNN-based network entity identification method for IP geolocation. *Comput. Netw.* **2023**, *235*, 109946. [\[CrossRef\]](#)
23. Cozar, M.; Rodriguez, D.; Del Alamo, J.M.; Guaman, D. Reliability of IP geolocation services for assessing the compliance of international data transfers. In Proceedings of the 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, 6–10 June 2022; pp. 181–185.
24. Livadariu, I.; Dreiholz, T.; Al-Selwi, A.S.; Bryhni, H.; Lysne, O.; Bjørnstad, S.; Elmokashfi, A. On the accuracy of country-level IP geolocation. In Proceedings of the Applied Networking Research Workshop, Virtual, 27–30 July 2020; pp. 67–73.
25. Ganelin, D.; Chuang, I. IP geolocation underestimates regressive economic patterns in MOOC usage. In Proceedings of the 11th International Conference on Education Technology and Computers, Amsterdam, The Netherlands, 28–31 October 2019; pp. 268–272.
26. Xu, W.; Tao, Y.; Guan, X. Experimental comparison of free IP Geolocation services. In *Security with Intelligent Computing and Big-Data Services: Proceedings of the Second International Conference on Security with Intelligent Computing and Big Data Services (SICBS-2018), Guilin, China, 14–16 December 2018*; Springer: Cham, Switzerland, 2020; pp. 198–208.
27. Sommers, J. A web client perspective on ip geolocation accuracy. In Proceedings of the 2020 International Symposium on Networks, Computers and Communications (ISNCC), IEEE, Montreal, QC, Canada, 20–22 October 2020; pp. 1–8.
28. Dan, O.; Parikh, V.; Davison, B.D. IP Geolocation through Geographic Clicks. *ACM Trans. Spat. Algorithms Syst. (TSAS)* **2022**, *8*, 1–22. [\[CrossRef\]](#)
29. Li, R.; Xu, R.; Ma, Y.; Luo, X. LandmarkMiner: Street-level network landmarks mining method for IP geolocation. *ACM Trans. Internet Things* **2021**, *2*, 1–22. [\[CrossRef\]](#)
30. AIWEN-TECH. IPUU. 2022. Available online: <https://mall.ipplus360.com/pros/IPVFourGeoDB/> (accessed on 27 April 2022).
31. IP2Region. IP2Region. 2022. Available online: <https://github.com/zoujingli/ip2region/> (accessed on 30 May 2022).
32. Gharaibeh, M. Characterizing the Visible Address Space to Enable Efficient Continuous IP Geolocation. Ph.D. Thesis, Colorado State University, Fort Collins, CO, USA, 2020.
33. Gan, Y.; Zhang, H.; Liu, Y.; He, L. IP Geolocation Method Based on Neighbor IP Sequences. In Proceedings of the 2020 6th International Symposium on System and Software Reliability (ISSR), IEEE, Chengdu, China, 24–25 October 2020; pp. 46–51.
34. Zhao, Q.; Wang, F.; Huang, C.; Yu, C. Improving IP geolocation databases based on multi-method classification. In Proceedings of the 2020 IEEE 14th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China, 30 October–1 November 2020; pp. 44–48.

35. Marchetta, P.; Botta, A.; Katz-Bassett, E.; Pescapé, A. Dissecting round trip time on the slow path with a single packet. In *Passive and Active Measurement: Proceedings of the 15th International Conference, PAM 2014, Los Angeles, CA, USA, 10–11 March 2014*; Springer: Cham, Switzerland, 2014; pp. 88–97.
36. Zhao, F.; Luo, X.; Gan, Y.; Zu, S.; Cheng, Q.; Liu, F. IP Geolocation based on identification routers and local delay distribution similarity. *Concurr. Comput. Pract. Exp.* **2019**, *31*, e4722. [[CrossRef](#)]
37. Van Brummelen, G.; Hamm, E.A. Heavenly mathematics: The forgotten art of spherical trigonometry. *Aestimatio Sources Stud. Hist. Sci.* **2014**, *11*, 127–130. [[CrossRef](#)]
38. CAIDA. Macroscopic Internet Topology Data Kit. 2022. Available online: <https://www.caida.org/catalog/datasets/internet-topology-data-kit/> (accessed on 14 June 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.