Received April 10, 2019, accepted April 24, 2019, date of publication May 6, 2019, date of current version May 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914972

Street-Level Geolocation Based on Router Multilevel Partitioning

FAN ZHAO[1, 2], RUI XU[3, 4], RUIXIANG LI[1, 2], MA ZHU[1, 2], AND XIANGYANG LUO[1, 2]

[1]State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China
[2]Zhengzhou Science and Technology Institute, Zhengzhou 450001, China
[3]Cyberspace Security Key Laboratory of Sichuan Province, Chengdu 610000, China
[4]China Electronic Technology Cyber Security Co., Ltd., Chengdu 610000, China

Corresponding author: Xiangyang Luo (luoxy_ieu@sina.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0801303 and Grant 2016QY01W0105, in part by the National Natural Science Foundation of China under Grant U1636219, Grant 61602508, Grant 61772549, Grant U1736214, and Grant 61572052, and in part by the Plan for Scientific Innovation Talent of Henan Province under Grant 2018JR0018.

ABSTRACT The high-precision geolocation of Internet hosts plays an important role in many applications, such as online advertising and deception detection. The existing typical high-precision geolocation algorithms usually utilize single-hop or relative delay to geolocate an Internet host at street-level granularity. However, it is difficult to accurately measure the single-hop or relative delay within a city. This challenge sometimes results in large geolocation errors. To solve this problem, a street-level geolocation algorithm based on router multilevel partitioning is proposed. Unlike existing typical algorithms, the proposed algorithm makes a credible hypothesis that each router has a relatively stable service object for a period of time. By analyzing the connection between routers and landmarks, the possible geographic service ranges of routers are inferred from the geographic distribution of landmarks. Then, distance constraints arising from routers' service ranges are formed to estimate the geographic location of the target IP. Theoretical analysis of the geolocation error shows that the maximum and average errors of the proposed algorithm are less than those of existing typical algorithms. The proposed algorithm is evaluated by geolocating a total of 12,152 target IP addresses located in four cities in different regions. The experimental results show that, compared with the existing typical street-level geolocation algorithms SLG and NC-Geo, the average median error of the proposed algorithm decreases from 4.735 km and 3.776 km to 3.25 km, representing error reductions of approximately 31.36% and 13.96%, respectively.

INDEX TERMS IP geolocation, delay-distance correlation, multilevel partitioning of routers, service range calculation.

I. INTRODUCTION

IP geolocation technology aims to obtain the geographical location of an IP address, such as country, city, longitude and latitude [1]. It is widely used in business marketing and network security. For example, after obtaining the user's location, Internet service providers can design targeted advertisements, intelligently adjust the language and content of web pages according to local laws, and push local weather forecasts and news information. It is no exaggeration that the vast majority of online services can benefit from identifying users' locations [2]. IP geolocation also plays an important role in network security. For instance, cheating behaviors could be detected by verifying a user's identity based on his location. Therefore, research on IP geolocation technology is of great significance.

Existing IP geolocation methods include 3 categories: database-based, data mining-based and network measurement-based. Database-based methods are currently widely used. Many kinds of IP location databases exist on the Internet, such as MaxMind [3], Quova [4], IP2Location [5], and NetAcuity [6]. Reference [1] evaluated the accuracy of IP2Location, MaxMind and NetAcuity for routers. The results showed that these databases are unsatisfactory, and there is considerable room for improvement. Reference [7] evaluated 3 popular databases and noted that the average accuracy at the city level of these databases is less than 70%.

The associate editor coordinating the review of this manuscript and approving it for publication was Shagufta Henna.

Data mining-based methods map IP addresses to geographic locations through association analysis of a large quantity of data. Typical methods include Structon [8] and Checkin-Geo [9]. Structon extracts IP address-related locations from web pages. Then, some strategies are taken to infer whether the IP address corresponds to the extracted location. However, due to the large number of shared hosts, server hosting and other situations, only a small number of IP addresses can be geolocated at street-level granularity. The Checkin-Geo algorithm mainly includes two parts. One is to extract user's precise location from the user's ''mobile social network check-in''; the other is to extract the IP address from a ''fixed login device''. Then, this information is combined through the same account to achieve precise geolocation of the IP address. This algorithm has high geolocation accuracy, but a large quantity of user-related data must be obtained from third parties.

Network measurement-based methods are current popular research topics that have attracted extensive attention from scholars. Typical algorithms include GeoPing [10], Constraint-based Geolocation (CBG) [11], Topology-based Geolocation (TBG) [12], Octant [13], Learning-based Geolocation (LBG) [14], Point of Presence (PoP) Analysis-based Geolocation [15], Geo-Cramér–Rao [16], Street-Level Geolocation (SLG) [17], Smartphone-based Geolocation [18], Geo-PoP [19], and NC-Geo [20]. These methods first measure the delay or the topological structure of landmarks and then estimate the target's location by analyzing the relationship between measured results and geographical location. However, most of these algorithms based on network measurement can only geolocate the target at region- or city-level granularity, and only a few algorithms, such as the SLG and NC-Geo algorithms, can geolocate the target at street-level granularity. In the SLG algorithm, the location of the landmark with the minimum relative delay to the target is selected as the estimated location of the target. The relative delay refers to the delay between the target and common router, plus the delay between the landmark and common router. The NC-Geo algorithm analyzes the inaccuracy of selecting the nearest landmark according to the minimum relative delay in SLG. Moreover, NC-Geo calculates the geographic location of the nearest common router and takes it as the target's estimated location. Compared with other existing geolocation algorithms, the SLG and NC-Geo algorithms significantly improve geolocation accuracy and are used in more applications.

The measured delay consists of propagation delay, transmission delay, processing delay and queuing delay, but only propagation delay is related to geographical distance. However, the propagating delay is a very small proportion of the measured delay between nodes within a city. It is difficult to convert the measured delay into an effective geographical distance. In addition, the relative delay in SLG is usually obtained by calculation, which may introduce more errors than direct measurement. Therefore, the landmark closest to the target is less likely to be selected according to the minimum relative delay. In the NC-Geo algorithm, the location of the nearest common router is calculated with a single-hop delay. Similarly, it can only determine a large geographical area due to the inaccurate delay. It is also difficult to determine which common router is closest to the target when there is more than one nearest common router.

To solve these problems, this paper proposes a street-level geolocation algorithm based on router multilevel partitioning. Unlike existing typical algorithms, the proposed algorithm does not utilize the delay within a city. Instead, it makes a credible hypothesis that each router has a relatively stable service object for a period of time. By analyzing the connection between routers and landmarks, the possible geographic service ranges of routers are inferred from the geographic distribution of landmarks. Then, distance constraints arising from routers' service ranges are formed to estimate the geographic location of the target IP. Both theoretical analysis and the experimental results demonstrate that the proposed algorithm reduces the geolocation error of existing street-level geolocation algorithms.

The organization of the paper is as follows. The defects of the street-level geolocation algorithms SLG and NC-Geo are analyzed in Section 2. Section 3 describes the framework and main steps of the proposed algorithm, with particular emphasis on two core parts of the algorithm. In Section 4, we analyze the performance of the proposed algorithm. Section 5 presents and discusses the experimental results. Section 6 summarizes the paper and highlights the main problems to be studied in the future.

II. PROBLEM DESCRIPTION

Existing IP geolocation algorithms based on network measurement usually attempt to describe the conversion or statistical relationship between delay and geographical distance. The geolocation accuracy of most of these algorithms is only tens of kilometers or even hundreds of kilometers. Only a few algorithms, such as SLG and NC-Geo, can geolocate the target IP at street-level granularity (several kilometers). The SLG algorithm uses a three-tier geolocation process to gradually reduce the possible location of the target. In the first tier, the delay between the probing hosts and the target IP is converted into geographical distance, and the target is geolocated to a larger area based on multilateration. In the second tier, the relative delay between the landmarks and the target is converted into distance; then, the target is geolocated to a smaller area via multilateration. A schematic diagram of the geolocation process of the third tier is shown in Fig. 1. The common routers are R_A, R_B and R_C. In the third tier, the location of the landmark with the minimum relative delay of the target is taken as the estimated location of the target, for example, the landmark L_b.

Reference [20] analyzed the relationship between the relative delay and geographical distance of 176 landmarks located in Zhengzhou city. In [20], the ''Dist-rank-of-shortest-delay'' method proposed in [2] was used to analyze the relationship between the minimum relative delay and the

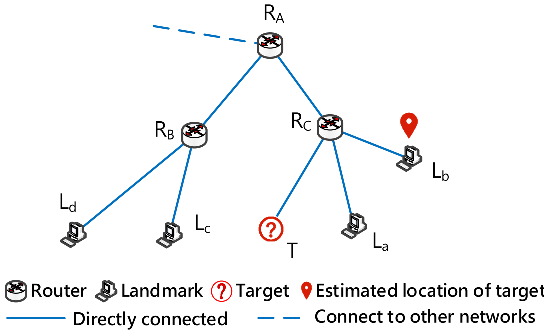

FIGURE 1. Schematic diagram of the third-tier geolocation process in the SLG algorithm.

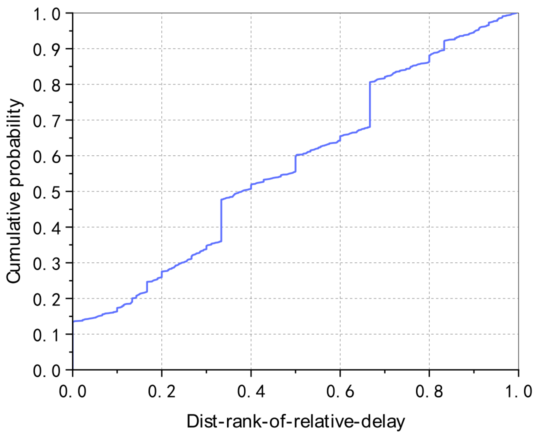

FIGURE 2. CDF of dist-rank-of-relative-delay for 6,861 landmarks.

shortest geographical distance. That is, for landmark A, find all relative (relative delay, distance) pairs and rank them from smallest to largest. The distance ranking corresponding to the shortest relative delay of A is equal to the order of the distance in the ranking divided by the total number of (relative delay, distance) pairs. The analysis result indicates that the ratio of minimum relative delay mapping to minimum geographical distance is less than 30%. This result illustrates that, in the SLG algorithm, the probability of successful selection of the nearest landmark is less than 30% according to the minimum relative delay. Considering that the number of landmarks used in this experiment is small, the results may be unreliable. We repeated the experiment using 6,861 landmarks located in Hong Kong. The cumulative probability distribution of the ranking is shown in Fig. 2. It shows that the SLG algorithm can only geolocate less than 15% of the targets to the nearest landmarks. Therefore, it is difficult for the SLG algorithm to find the nearest landmark in the third level.

In [20], an IP geolocation algorithm based on the nearest common router (NC-Geo) was proposed. A schematic diagram of the geolocation process of the NC-Geo algorithm is presented in Fig. 3. The algorithm geolocates the target in the following way. First, the nearest common router R_N between the target and the landmarks is found from the topology measurement results. Then, the cosine theorem is used to calculate the conversion coefficient from the single-hop delay to the geographic distance between the landmarks and the common router. When there are more than three landmarks connected to the common router, the geographical location of the nearest common router is geolocated based on multilateration. In addition, the location of the nearest common router (R_N) is taken as the target's location.

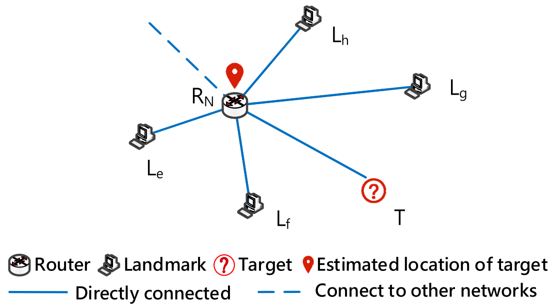

FIGURE 3. Schematic diagram of the geolocation process in the NC-Geo algorithm.

The algorithm considers that the principle of the minimum relative delay corresponding to the shortest distance in the city is invalid, but this principle as proposed in [2] is still valid. However, the propagating delay related to distance in the measured delay is very small between nodes in a city. In particular, the delay of the "last mile" is affected by many factors. These effects lead to a smaller proportion of propagation delay in the measured delay. Furthermore, the delay between common routers and landmarks cannot be measured directly and is often obtained indirectly by calculation. Therefore, it is difficult to convert the single-hop or relative delay into appropriate geographical distance. As a result, when geolocating a common router based on multilateration, the intersection area is large, as is the geolocation error. In addition, if multiple common routers are nearest to the target, we will not know which router should be geolocated.

III. PROPOSED ALGORITHM

To solve the above problems, this paper proposes a street-level geolocation algorithm based on router multilevel partitioning. The algorithm assumes that the location of the next hop node on the routing path of each router is relatively fixed in a certain period of time; that is, the distribution range of these nodes (which is called the service range of the router in this paper) is relatively fixed. After obtaining the topological connection among landmarks, targets and routers, the possible location of targets could be estimated according to the service range of common routers. Since the relative or single-hop delay is difficult to measure accurately, we do not utilize it in our geolocation process. This avoids the incorrect choice of the closest landmark in the SLG. Similarly, it also avoids the inaccurate calculation of the nearest common router's location in the NC-Geo. If there is more than one nearest common router, we take the intersection of their service ranges as the target's location. This approach makes up for the deficiency of the NC-Geo algorithm in this case.

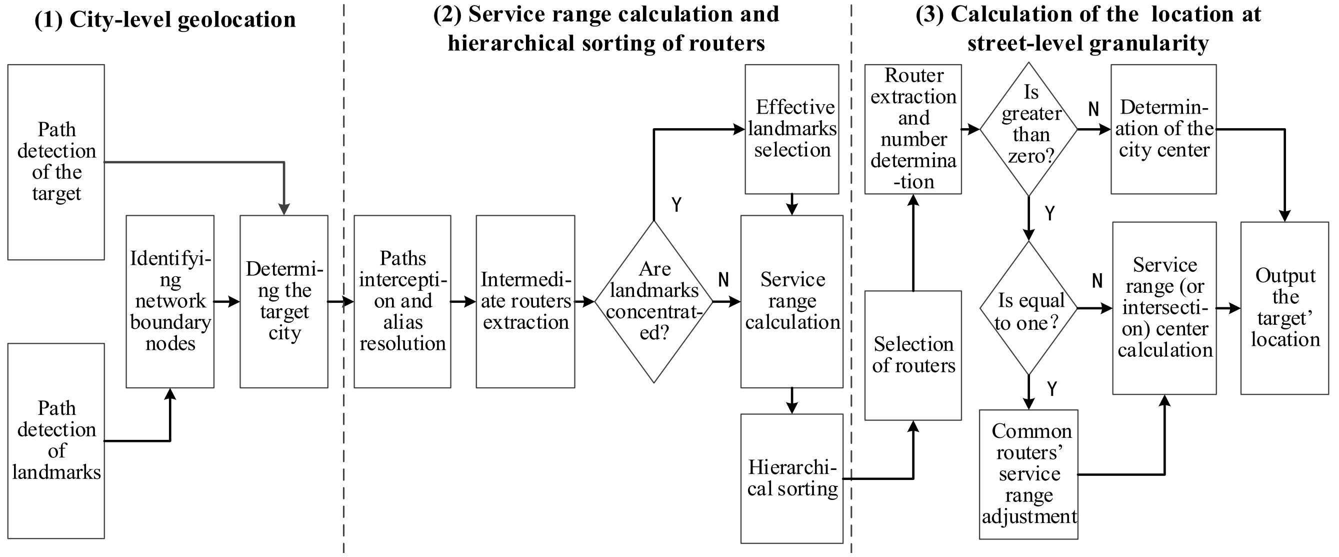
(1) City-level geolocation
(2) Service range calculation and hierarchical sorting of routers
(3) Calculation of the location at street-level granularity
Path detection of the target
Path detection of landmarks
Identifying network boundary nodes
Determining the target city
Paths interception and alias resolution
Intermediate routers extraction
Are landmarks concentrated?
Y
N
Effective landmarks selection
Service range calculation
Hierarchical sorting
Router extraction and number determination
Selection of routers
Is greater than zero?
N
Y
Determination of the city center
Is equal to one?
N
Y
Service range (or intersection) center calculation
Common routers' service range adjustment
Output the target' location

FIGURE 4. Framework of the proposed algorithm.

A. FRAMEWORK AND MAIN STEPS

Many factors are considered by Internet service providers when deploying network devices, such as delay and bandwidth. Currently, Open Shortest Path First (OSPF) is one of the most widely used network protocols. To be applicable to large-scale networks, the protocol divides the hierarchical structure into regions and logically divides a larger autonomous system into smaller areas. In practical applications, this division tends to be consistent with geographical region divisions. The consistency makes it possible for us to determine the target's location at city-level granularity based on its routing path. At the same time, when approaching end users, Internet service providers usually take into account many factors to provide high-quality services. For example, the deployed devices need to be maintained and managed easily, and network delay caused by overload should be reduced. For this reason, the distance between the deployed access router and the end users is usually small [21]. Each access router has a limited number of users and a relatively fixed geographical distribution.

Under the guidance of the above ideas, this algorithm first initiates path detection for landmarks and identifies the boundary routers of cities according to the routing paths. By comparing the identified boundary routers and the target's routing path, the target city is determined. Next, the initial service ranges of the routers are inferred according to the geographical location distribution of the connected landmarks. Then, according to the number of hops between the routers and the landmarks, the routers are sorted hierarchically. Finally, the finer-grained location of the target is calculated according to the service range of one or more routers connected to the target IP. The framework of the proposed algorithm is shown in Fig. 4.

The algorithm mainly consists of 3 parts: city-level geolocation (steps 1-3), service range calculation and hierarchical sorting of routers (steps 4-8), and estimation of the street-level location of the target (steps 9-11). The complete steps are as follows.

Input: target IP and landmarks

Output: geographical location of the target IP

Step 1: Measurement of the target and landmarks' routing paths. The IP addresses of a series of routers on the routing path from each probing host to the target and landmarks are obtained by path detection. Then, the topological connections between these nodes are constructed.

Step 2: Identification of network boundary nodes of candidate cities. The IP addresses of the routers that only forward packets to a single city are found and are regarded as the network boundary IP addresses of the corresponding city.

Step 3: Determination of the target city. The routing path of the target is compared with the network boundary nodes of each candidate city. The city whose network boundary node is included in the target's routing path is selected as the target city (i.e., the city where the target IP is located).

Step 4: Interception of landmark paths and alias resolution of routers. The routing paths of the landmarks that are between the probing hosts and the target city's boundary nodes are deleted. That is, only the IP address on the routing paths between the target city's boundary nodes and the landmarks is reserved. Then, multiple aliases belonging to the same router are resolved.

Step 5: Extraction of intermediate routers and judgment of the distribution of connected landmarks. The number of landmarks connected to each intermediate router and the number of geographical locations of these landmarks are counted. When the number of landmarks is greater than the number of geographical locations, step 6 is executed; otherwise, step 7 is executed.

Step 6: Selection of effective landmarks. For different landmarks located in the same location, we use subnet analysis tools (such as the TreeNET tool in [22]) to determine whether these landmarks belong to the same subnet. If multiple landmarks belong to the same subnet, only one of them is retained.

Step 7: Calculation of the intermediate router service range. For the landmarks (assume that the number is k) connected to the intermediate router R_i, the center of all landmarks is calculated. The circle formed by the radius of the distance between the center and the farthest landmark serves as the initial service range of R_i. The radius and center of the circle are denoted as r and O, respectively. Then, the geographic attribute quaternion of R_i in the form of (R_i, O, r, k) is constructed.

Step 8: Hierarchical sorting of intermediate routers. According to the hop distance between the intermediate routers and the connected landmarks, the intermediate routers are hierarchically sorted. That is, the router with l hops from the landmark is marked as level l. At the same level, intermediate routers are sorted according to the number of connected landmarks such that the intermediate routers with more landmarks are closer to the front in the sequence.

Step 9: Selection of routers at the same level on the target path. Starting from the router closest to the target, the router is examined to determine whether it is at the same level as the router sequences constructed in step 8. If the router does not exist in the constructed sequence, it indicates that there is no common router between the target and all landmarks. Then, the central longitude and latitude of the target city are calculated as the estimated location of the target and the algorithm terminates. Otherwise, step 10 is executed.

Step 10: Router extraction and number determination. If the number of common routers is only one, then the center of the router's service area is taken as the estimated location of the target and the algorithm terminates; otherwise, step 11 is executed.

Step 11: Adjustment of common routers' service range and calculation of the intersection center. The total number of landmarks connected to the target through the nearest common routers is denoted as k_{sum}. Combined with the geographic attribute quaternion of each router established in step 7, each nearest common routers' service range is gradually expanded in the proportion k/k_{sum} until the intersection is not empty. Finally, the center of the intersection is calculated as the target's estimated location

and the algorithm terminates.

Next, we will elaborate on the two core parts of the algorithm: the service range calculation and hierarchical sorting of routers (steps 4-8), and the estimation of the street-level location of the target (steps 9-11).

B. ROUTER SERVICE RANGE CALCULATION AND HIERARCHICAL SORTING

Routers are important interconnection devices between hosts on the Internet. To provide users with low-latency network services and high-quality network services, Internet service providers often deploy routers according to the size and distribution of networked users. This approach makes each router have a specific range of services, some of which are relatively small or centralized. The service range calculation and hierarchical sorting of routers could provide a basis for calculating the target's location in the future. The detailed process is as follows.

1) CALCULATION OF THE ROUTER SERVICE RANGE

Because of the limited number of landmarks, it is difficult to obtain the connection information between the router and all nodes in its service range. Therefore, there is a certain error in inferring the service range of the router according to only the distribution of the connected landmarks. We describe the reliability of the inferred results by the number of landmarks distributed in different locations connected to routers. That is, when more landmarks in different locations are used to calculate the service range of a particular router, the reliability of the calculated service range increases. Likewise, for different routers with the same hop distance from the landmarks, using more landmarks in different locations increases the reliability of the calculation results.

The geographic locations of landmarks belonging to the same subnet are usually adjacent and are connected to the same router. Therefore, the number of landmarks should not be directly used to measure the reliability of the router service range calculation. To exclude this effect of this situation, we first inspect the distribution of the connected landmarks. If the distribution of the landmarks (or portion of landmarks) is very centralized, TreeNET is used to analyze whether these landmarks belong to the same subnet. If so, only one landmark is retained to participate in the calculation.

For the convenience of the following narrative, we make the following assumptions. After intercepting the paths, the routers in the routing paths of all landmarks are represented as $\mathbf{R} = \{R_1, R_2, \cdots, R_n\}$. By subnet analysis and filtering, there are $count_k$ landmarks (denoted as $\{L_1, L_2, \cdots, L_{count_k}\}$) that are connected to the k-th router (denoted as R_k). The center of the $count_k$ landmarks is calculated and recorded as $center_k$. The distance from the farthest landmark to $center_k$ is marked as rad_k, and the geographic attribute quaternion of R_k is constructed as $(R_k, center_k, rad_k, count_k)$.

2) HIERARCHICAL SORTING OF ROUTERS

Usually, the further the hops are between the landmark and the router, the wider the router service range (detailed in Section 4.1). When geolocating the target, the larger the service range of the router, the less help it provides. For intermediate routers, if they have the same number of hops

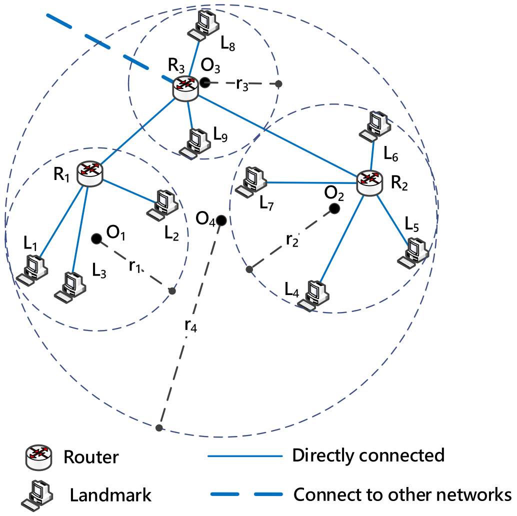

FIGURE 5. Examples of multilevel partitioning of intermediate routers.

from the landmarks, connecting to more landmarks increases the reliability of the calculated service range. Therefore, it is necessary to classify routers according to their distance from the landmarks and rank the routers according to the number of landmarks connected at the same level. In this paper, the router at a distance of i hops from the landmark is called the i-th level.

When $i = 1$, the first-level sequence constructed is as follows:

$\mathbf{S}_1 = [(R_1^1, center_1^1, rad_1^1, count_1^1), (R_2^1, center_2^1, rad_2^1, count_2^1), \cdots, (R_x^1, center_x^1, rad_x^1, count_x^1)]$, where x is the number of routers in the first-level sequence and $count_1^1 \geq count_2^1 \geq \cdots \geq count_x^1$.

If $i = 2$, the second level sequence is as follows:

$\mathbf{S}_2 = [(R_1^1, center_1^1, rad_1^1, count_1^1), (R_2^1, center_2^1, rad_2^1, count_2^1), \cdots, (R_y^1, center_y^1, rad_y^1, count_y^1)]$, where y is the number of routers in the second level sequence and $count_1^2 \geq count_2^2 \geq \cdots \geq count_x^2$.

By analogy, the i-th level sequence is constructed as follows:

$\mathbf{S}_i = [(R_1^1, center_1^1, rad_1^1, count_1^1), (R_2^1, center_2^1, rad_2^1, count_2^1), \cdots, (R_z^1, center_z^1, rad_z^1, count_z^1)]$, where z is the number of routers in the i-th level sequence and $count_1^i \geq count_2^i \geq \cdots \geq count_z^i$.

Finally, the sequence denoted as $[\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_i]$ can be constructed for all routers.

The following example illustrates the construction process of the above router multilevel sequence. As shown in Fig. 5, $L_1, L_2, \cdots, L_9$ are landmarks and R_1, R_2, R_3 are routers. We assume that all landmarks do not belong to the same subnet, and the connection between them is shown in Fig. 5. In this scenario, the construction process of the multilevel sequence is as follows.

① Calculation of the router service range. The service range of each router is calculated according to the distribution and number of connected landmarks. The quaternion of each router is represented as $(R_1, O_1, r_1, 3)$, $(R_2, O_2, r_2, 4)$, $(R_3, O_3, r_3, 2)$ and $(R_3, O_4, r_4, 9)$.

② Hierarchical sorting of routers. Fig. 5 shows that the routers directly connected to the landmarks are R_1, R_2 and R_3; thus, the first-level sequence is $[(R_2, O_2, r_2, 4)$, $(R_1, O_1, r_1, 3)$, $(R_3, O_3, r_3, 2)]$. The second level sequence is $[(R_3, O_4, r_4, 9)]$ because the router connected to the landmarks via 2 hops is R_3.

This example shows that a router may be arranged into several hierarchical sequences. The lower the level of the sequence, the further the router may be from the landmark, and the wider the calculated service range is. When the router service range is used to infer the location of the target, the constraints on the possible location of the target are weak. Therefore, when geolocating a target, we first look for the nearest common router from the highest-ranking sequence (see Section III(C) for details).

C. GEOLOCATION FOR THE TARGET IP

In this section, we describe in detail how to geolocate the target IP. We recorded the intercepted path of the target IP_T as $(h_{p-e}^T, \cdots, h_{p-1}^T, h_p^T, IP_T)$ and the constructed comparison sequence as $[\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_i]$. A detailed comparison of the calculation methods is as follows.

(1) h_p^T is matched with the IP address of the router in sequence $\mathbf{S}_1$. The set of geographic attribute quaternions of the routers that are the same as h_p^T in $\mathbf{S}_1$ are denoted as $\mathbf{G}_1 = \{(R_1^1, center_1^1, rad_1^1, count_1^1), (R_2^1, center_2^1, rad_2^1, count_2^1), \cdots, (R_q^1, center_q^1, rad_q^1, count_q^1)\}$, (where $count_1^1 \geq count_2^1 \geq \cdots \geq count_q^1$), and the total number of landmarks connected by these routers is $count_{sum}^1 = count_1^1 + count_2^1 + \cdots + count_q^1$. If $\mathbf{G}_1 \neq \phi$:

1) If $q = 1$, $center_1^1$ is taken as the estimated position of the target.

2) If $q > 1$:

a. If the circles $(center_1^1, rad_1^1)$, $(center_2^1, rad_2^1)$, $(center_q^1, rad_q^1)$ intersect each other, the center of their intersection is taken as the target's location.

b. If the circles $(center_1^1, rad_1^1)$, $(center_2^1, rad_2^1)$, $(center_q^1, rad_q^1)$ do not intersect or partially intersect, the radius of the service range of $R_1^1, R_2^1, \cdots, R_q^1$ is gradually expanded in the proportion of $count_1^1 \big/ count_{sum}^1$, $count_2^1 \big/ count_{sum}^1$, $\cdots$, $count_q^1 \Big/ count_{sum}^1$ until each circle intersects. Then, the center of the intersection is taken as the estimated location of the target.

(2) If $\mathbf{G}_1 = \phi$, h_{p-1}^T is matched with the IP addresses of the routers in the sequence $\mathbf{S}_2$, and the comparison and calculation method are the same as those above.

(3) By analogy, if the set of geographic attribute quaternions of the IP address as h_{p-e}^T is still empty, the center of the target city is taken as the estimated position of the target.

Fig. 6 shows an example of the geolocation process when $e = 1$. We assume that the nearest common router is found in the first-level sequence. We discuss the geolocation process of the target T in 3 cases.

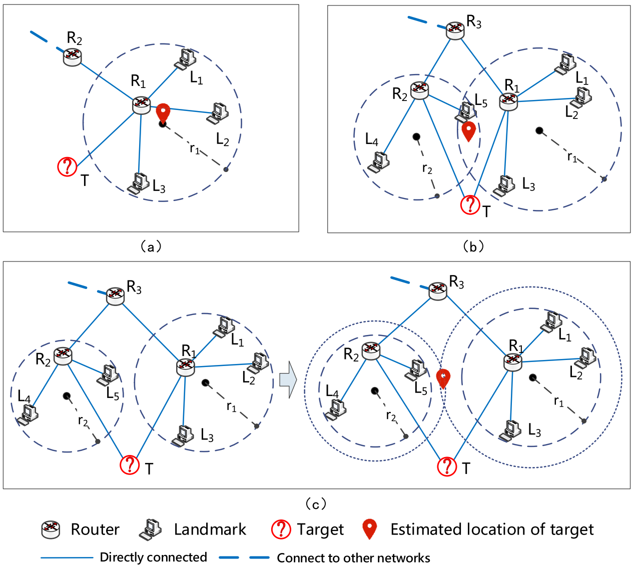

FIGURE 6. Example of the geolocation process.

(1) If there is only one nearest common router, as presented in Fig. 6 (a), the center of the router's service range is taken as the estimated location of the target, that is, the location of the red water droplet icon in the figure.

(2) If the number of nearest common routers is greater than one, the intersection of the service ranges of these routers is first checked to determine whether the intersection is empty. If the intersection is not empty, as indicated in Fig. 6 (b) (we take 2 routers as an example), the center of the intersection is taken as the target's location. Otherwise, the radii of the service range of routers r_1 and r_2 are enlarged by proportions of 2/5 and 3/5, respectively, until intersection occurs. Then, the center of the intersection is taken as the estimated location of the target, as shown in Fig. 6 (c).

IV. PERFORMANCE ANALYSIS OF THE PROPOSED ALGORITHM

In this section, the principle of the algorithm is analyzed in terms of two aspects: the effectiveness of router multilevel partitioning and the error comparison with the existing typical geolocation algorithms.

A. ANALYSIS OF THE EFFECTIVENESS OF ROUTER MULTILEVEL PARTITIONING

The basic idea of this algorithm is to estimate the possible location of the target by determining the service range of the router connected to the target. Therefore, the service range directly affects the geolocation error for the target. If the service range is small, the target can be geolocated accurately. The overall performance of the algorithm is determined by the existence of such routers with a small service range and the proportion of these routers among all routers. From the point of view of Internet service providers, to provide users with high-quality network services, it is better to deploy network devices closer to users than to randomly deploy them in the whole city. Therefore, there should be a considerable proportion of routers close to network terminals for specific geographical areas.

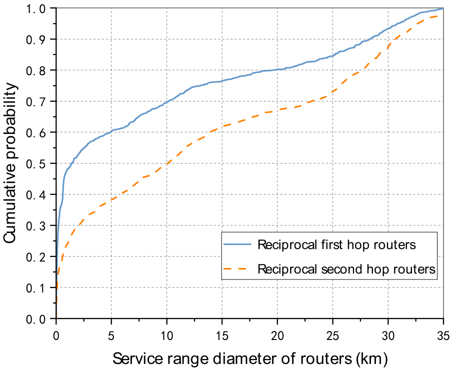

FIGURE 7. CDF of the service range diameter of some routers located in Hong Kong.

The above hypothesis is verified by the following experiments. We measure the routing paths of 6,861 landmarks located in Hong Kong and obtain the connections between these landmarks and intermediate routers. Based on the statistics of the measurement results, 809 routers directly connected with at least 5 landmarks located at different locations are selected. According to the geographical distribution of these landmarks, the service range of each router is obtained by using the calculation method in Section 3.2. The blue solid line in Fig. 7 presents the cumulative probability distribution of the service ranges of these routers. As shown in the figure, more than 50% of routers have a service range of less than 2 km, and more than 60% of routers have a service range of less than 5 km. In addition, 634 routers connected with at least 5 landmarks at different locations through 2 hops are selected. The cumulative probability distribution of the service range of these routers is shown by the red dotted line in Fig. 7. This figure indicates that the service range of routers with 2 hops from landmarks is obviously larger, but nearly 40% of routers still have a service range of less than 5 km. The existence of these routers makes it possible to achieve high-precision geolocation results.

Fig. 7 also shows that some routers have a large service range, and this may be because some of the landmarks are close to the backbone network. Although the service range of these routers is large, their range is still significantly smaller than the maximum diameter (approximately 70 km) of Hong Kong. When geolocating a target, it is often possible to find several nearest common routers connecting the target and the landmarks. As mentioned in Section 3.3, the target is still expected to geolocate to a relatively small area.

B. ANALYSIS OF GEOLOCATION ERROR COMPARED WITH EXISTING TYPICAL GEOLOCATION ALGORITHMS

This paper compares the performance of the SLG algorithm, NC-Geo algorithm and the proposed algorithm in terms of the two aspects of maximum error and average error.

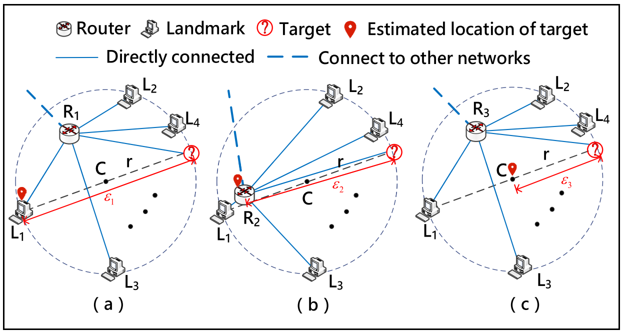

FIGURE 8. Scenarios for maximum error analysis.

1) ANALYSIS OF THE MAXIMUM ERROR

To analyze the maximum geolocation error of the 3 algorithms, we take as an example the scenario in which a common router is directly connected to the target and landmarks. The geographical center of the landmarks connected to the common router, the distance between the center and the farthest landmark, and the distance between the target and the common router are recorded as C, r and d_e, respectively. Assuming that the target and the nearest common router have the same probability of existence at any point in the circle, the following is true:

For the SLG algorithm, the location error is the distance between the target and the selected landmark based on the minimum relative delay. The maximum error occurs when the target and the selected landmark are located at both ends of the circle's diameter, as shown in Fig. 8 (a). Thus, the maximum error is $\varepsilon_1 = 2r$(the orange solid line in Fig. 8 (a)).

In the NC-Geo algorithm, the location of the nearest common router is taken as the estimated location of the target. Therefore, the geolocation error is the distance between the common router and the target. The maximum error also occurs when the nearest common router and target are located at both ends of the circle's diameter. Since the algorithm assumes that the nearest common router is located in the interior of multiple landmarks, the maximum error is slightly less than the diameter (as shown in the red solid line in Fig. 8 (b)). We denote the maximum error as $\varepsilon_2 = r + \mu$ (where $0 \leq \mu < r$); then, $r < \varepsilon_2 < 2r$.

For the proposed algorithm, in this scenario, the center of the circle is taken as the estimated position of the target; thus, the geolocation error is the distance between the center and the target. The maximum error occurs when the target is on the circle (as shown in Fig. 8 (c)), and the maximum error is the radius of the circle, that is, $\varepsilon_3 = r$(as shown by the red solid line in the figure).

Consequently, the comparison results of the maximum geolocation errors of the 3 algorithms are as follows: $\varepsilon_1 > \varepsilon_2 > \varepsilon_3$; that is, the maximum error of the proposed algorithm is the smallest.

2) ANALYSIS OF AVERAGE ERROR

The SLG algorithm geolocates the target at the location of a landmark falling within the circle. Referring to the average error analysis method in [20], we assume that there are w landmarks in the circle; they are denoted as $L_1, L_2, \cdots, L_w$. The distance between the i-th landmark (denoted as L_i) and the target T is recorded as $d_{dis}(L_i, T)$, and the distribution of the landmark in the circle is regarded as the uniform distribution. Then, it is true that $d_{dis}(L_i, T) \in [0, \varepsilon_1]$. We consider $d_{dis}(L_i, T)$ as a set of uniform values in $[0, \varepsilon_1]$, and the probability of each value is $1/w$. Then, the average error of the SLG algorithm is as follows.

$$\begin{aligned} &E_{SLG-ave} \\ &= \mathrm{E}\left(d_{dis}\left(L_i, T\right)\right) = \frac{1}{w}\sum_{i=1}^{w} d_{dis}\left(L_i, T\right) = \frac{1}{w}\sum_{i=1}^{w}\frac{i * \varepsilon_1}{w} \\ &= \frac{1}{w}\left(\frac{\varepsilon_1}{w} + \frac{2\varepsilon_1}{w} + \ldots + \varepsilon_1\right) \\ &= \frac{1}{w}\left(\frac{\varepsilon_1}{w}(1 + 2 + \ldots + w)\right) \\ &= \frac{\varepsilon_1(w+1)}{2w} = \frac{r(w+1)}{w} \end{aligned} \tag{1}$$

The NC-Geo algorithm regards the location of the nearest common router as the target's location. Considering that there is a large error in converting the single-hop delay into distance within a city, if w landmarks are evenly distributed in the circle, the location of the nearest common router can also be regarded as uniformly distributed. The distance between the common router and the target calculated by the j-th calculation is $d_{dis}(R_j, T)$ and $d_{dis}(R_j, T) \in (0, \varepsilon_2]$. We consider $d_{dis}(R_j, T)$ as a set of uniform values in $(0, \varepsilon_2]$, and the probability of each value is $1/w$. Then, the average error of the NC-Geo algorithm is as follows.

$$\begin{aligned} E_{NC-Geo-ave} &= \mathrm{E}\left(d_{dis}\left(R_j, T\right)\right) = \frac{1}{w}\sum_{j=1}^{w} d_{dis}\left(R_j,\ T\right) \\ &= \frac{1}{w}\sum_{j=1}^{w}\frac{j * \varepsilon_2}{w} = \frac{\varepsilon_2(w+1)}{2w} \\ &= \frac{(r+\mu)(w+1)}{2w} \end{aligned} \tag{2}$$

The proposed algorithm regards the center of the circle as the estimated position of the target. Since the center of the circle is fixed, the possible position of the target can be regarded as a uniform distribution in the circle. The distance between the common router location and the target calculated in the k-th time is recorded as $d_{dis}(C, T_k)$, and $d_{dis}(C, T_k) \in [0, \varepsilon_3]$. We consider $d_{dis}(C, T_k)$ as a set of uniform values in $[0, \varepsilon_3]$, and the probability of each value is $1/w$. Then, the average error of the proposed algorithm is as follows.

$$\begin{aligned} E_{Pro-Geo-ave} &= \mathrm{E}\left(d_{dis}\left(C, T_k\right)\right) = \frac{1}{w}\sum_{k=1}^{w} d_{dis}\left(C, T_k\right) \\ &= \frac{1}{w}\sum_{k=1}^{w}\frac{k * \varepsilon_3}{w} = \frac{\varepsilon_3(w+1)}{2w} \\ &= \frac{r(w+1)}{2w} \end{aligned} \tag{3}$$

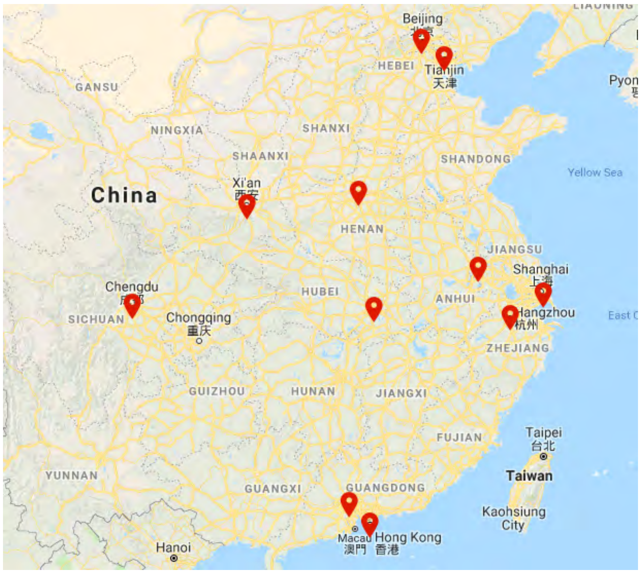

FIGURE 9. Geographical distribution of probing hosts located in China.

In conclusion, the results of comparing the average error of the 3 algorithms are as follows:

$$\begin{aligned} E_{Pro-Geo-ave} &= \frac{r(w+1)}{2w} \leq E_{NC-Geo-ave} \\ &= \frac{(r+\mu)(w+1)}{2w} < E_{SLG-ave} = \frac{r(w+1)}{w}; \end{aligned}$$

that is, compared with the SLG algorithm and NC-Geo algorithm, the average geolocation error of the proposed algorithm is the smallest.

V. EXPERIMENTS

To evaluate the performance of the proposed algorithm, geolocation experiments are performed on 12,152 IP addresses located in China and the United States.

A. EXPERIMENTAL SETTINGS

1) NUMBER AND GEOGRAPHICAL DISTRIBUTION OF DEPLOYED PROBING HOSTS

Considering that the targets in the experiments are located in China and the United States, the probing hosts are deployed in these two countries to reduce the redundancy of detection. Among them, the probing hosts located in China are deployed in 11 different cities, such as Beijing, Shanghai, Guangzhou, Hong Kong and so on. The probing hosts in the United States are also located in 11 cities, including Los Angeles, Washington and New York, among others. Fig. 9 and Fig. 10 show the geographical distribution of these probing hosts.

2) SOURCE AND SIZE OF THE GROUND TRUTH DATA

The ground truth data consisting of IP addresses with known locations used in the experiments are mainly obtained by the following two ways:

The first way mines landmarks from web pages. The detailed acquisition process is as follows. First, the geographical location of an organization is mined from the web page of its homepage; then, the IP addresses are obtained by parsing the domain names of this homepage. We link them together to form a record, such as ⟨geographical location, IP address⟩. Next, the IP address is used to access the web page. If the result returned is inconsistent with the original homepage, the record is considered untrustworthy and is deleted. Finally, we use the landmark evaluation method based on the nearest common router proposed in [23] to evaluate the reliability of the remaining landmarks and retain the credible records.

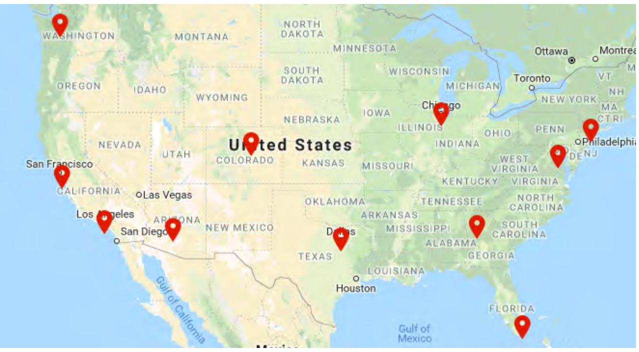

FIGURE 10. Geographical distribution of probing hosts located in the United States.

TABLE 1. The number of IP addresses and geographical locations in the experiment.

City	The number of IP addresses	The number of geographical locations
Beijing (China)	1,849	1,467
Hong Kong (China)	6,861	6,861
Zhengzhou (China)	982	893
Los Angeles (United States)	2,460	1,305

The second way collects an IP with the street-level location in query results from existing public databases. Then, the reliability of the landmarks is evaluated by the method proposed in [23], and the IPs with credible locations after the evaluation are retained.

In the IP addresses obtained by the above two ways, the IP addresses that respond to the requests from probing hosts are used in the experiment. The distribution of the number of landmarks in each city is reported in Table 1.

3) MEASURING TOOLS AND STRATEGIES

In the experiments, the delay and topology of the targets and landmarks are needed. The measurements are initiated from distributed probing hosts, and the traceroute tool is utilized. When measuring a given IP address, the traceroute tool can present a series of IP addresses from the probing host to this IP (the path where the packets are forwarded), in addition to the delay between the IP and the probing host.

To improve the efficiency of measurement and obtain as rich a network topology as possible around the target area, different probing hosts are used for targets and landmarks located in different cities. For targets and landmarks located in China, we measure the delay and topology from 10 probing hosts located in 10 cities in mainland China and 1 probing host located in Washington, DC. For targets and landmarks located in Hong Kong, 6 probing hosts located in Beijing, Shanghai, Guangzhou, Hong Kong, Washington, DC and Los Angeles are used. For targets and landmarks located in the United States, 2 probing hosts located in China (Beijing and Hong Kong) and 11 probing hosts located in the United States are used. Due to the influence of delay expansion, we measure each target and landmark 20 times and take the minimum value for the experiment.

B. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, comparison experiments with the NC-Geo algorithm and the SLG algorithm are carried out to evaluate the performance of the proposed algorithm. We take 12,152 IP addresses with known location as the targets, and we compare the geolocation errors of these algorithms for these targets. In the experiments, leave-one-out cross-validation is adopted; that is, in each geolocation experiment, one IP is taken as the target, and the remaining IP addresses are taken as landmarks.

1) COMPARISON WITH THE NC-GEO ALGORITHM

The NC-Geo algorithm requires that there are more than 3 landmarks connected to the nearest common router; otherwise, the location of the nearest common router cannot be calculated, and the target IP cannot be geolocated. The error analysis shows that the average error and maximum error of the NC-Geo algorithm are higher than those of the proposed algorithm. In this paper, geolocation experiments compared with the NC-Geo algorithm are carried out. The median error and maximum error are used to represent the accuracy of these two algorithms. The experimental results are shown in Table 2 and Table 3.

Table 2 and Table 3 indicate that the median error and maximum error of the proposed algorithm are less than those of the NC-Geo algorithm. The NC-Geo algorithm can geolocate only the target IP addresses where the number of the landmarks connected to the nearest common router is more than three. As a result, the NC-Geo algorithm fails to geolocate 253, 424, 167 and 437 targets in Beijing, Zhengzhou, Hong Kong and Los Angeles, respectively. In contrast, the proposed algorithm can geolocate all targets.

2) COMPARISON WITH THE SLG ALGORITHM

The SLG algorithm is also taken as a comparison algorithm to geolocate the same target. Fig. 11 shows the cumulative probability of the error distance for 982 targets located in Zhengzhou. The median and maximum errors of the SLG algorithm are 2.33 km and 20.48 km, respectively, which are 1.24 km and 15.28 km for the proposed algorithm.

TABLE 2. Test results for number of targets that can be geolocated.

City	Number of targets	Number of targets that can be geolocated	
		NC-Geo[20]	Proposed
Beijing	1,849	1,596	1,849
Zhengzhou	982	815	982
Hong Kong	6,861	6,437	6,861
Los Angeles	2,460	2,023	2,460

TABLE 3. Comparison of geolocation results with the NC-Geo algorithm.

City	Median error (km)		Maximum error (km)	
	NC-Geo[20]	Proposed	NC-Geo[20]	Proposed
Beijing	7.26	6.08	29.82	24.86
Zhengzhou	1.37	1.24	18.39	15.28
Hong Kong	4.25	3.78	31.62	30.61
Los Angeles	2.23	1.9	34.28	33.27

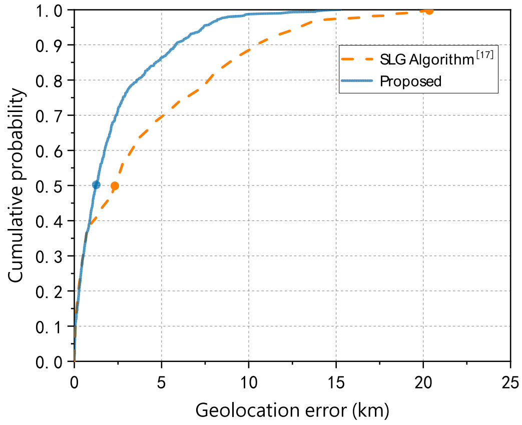

FIGURE 11. CDF of geolocation error for 982 targets located in Zhengzhou City.

The median and maximum errors are reduced by 46.78% and 25.39%, respectively.

The cumulative probability of geolocation errors for 1,849 targets located in Beijing is presented in Fig. 12. As indicated, the median and maximum errors of the SLG algorithm are 8.98 km and 34.89 km, respectively. Furthermore, they are 6.08 km and 24.86 km for the proposed algorithm, respectively, which are approximately 32.29% and 28.75% less than those of the SLG algorithm.

Fig. 13 presents the geolocation results of 2,460 targets located in Los Angeles. The median error and maximum error of the SLG algorithm are 3.16 km and 34.74 km, respectively. The median error and maximum error of the proposed algorithm are 1.9 km and 33.27 km, respectively, which are 39.87% and 4.23% less than those of the SLG algorithm.

The geolocation results of 6,861 targets located in Hong Kong are shown in Fig. 14. The median errors of these two algorithms are 4.47 km and 3.78 km, and the

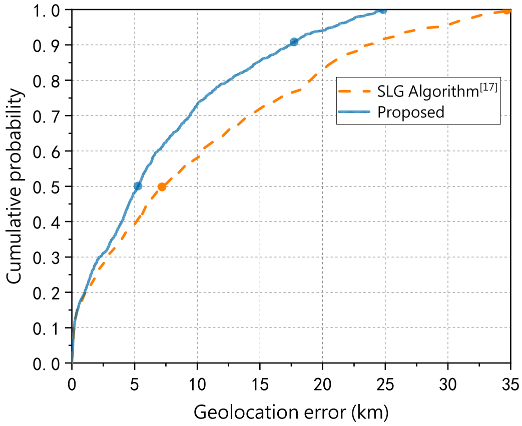

FIGURE 12. CDF of geolocation error for 1,849 targets located in Beijing City.

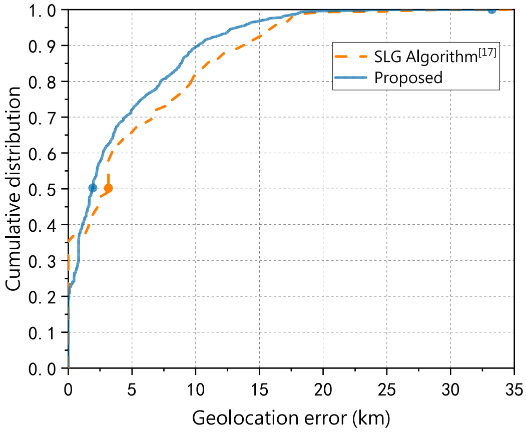

FIGURE 13. CDF of geolocation error for 2,460 targets located in Los Angeles.

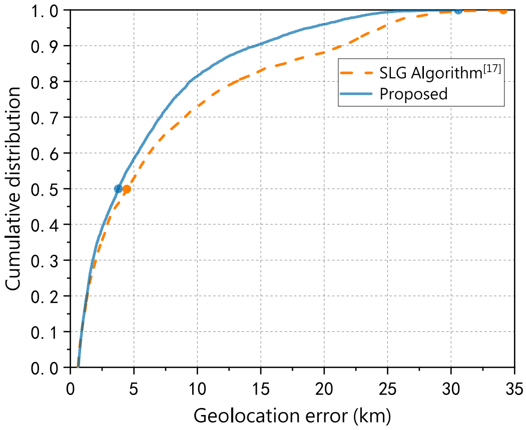

FIGURE 14. CDF of geolocation error for 6,861 targets located in Hong Kong.

maximum errors are 34.12 km and 30.61 km, respectively. Compared with the SLG algorithm, our algorithm reduces the median error and maximum error by approximately 15.44% and 10.29%, respectively.

Fig. 13 also shows that the SLG algorithm and the proposed algorithm geolocate approximately 35% and 23% of the targets to an error of nearly 0 km. This result occurs because the 2,460 targets located in Los Angeles are distributed in only 1,305 different locations; that is, many targets are located in the same location. This situation makes the SLG algorithm have a higher probability of selecting the landmark at the same location as the target, although the minimum relative delay is difficult to measure and calculate accurately. However, our algorithm has higher geolocation accuracy for all targets.

VI. CONCLUSION

IP geolocation technology has played an important role in many fields and has attracted extensive attention from many scholars. The complexity and dynamics of the Internet make it difficult to accurately geolocate target IP addresses. Some scholars have proposed high-precision geolocation algorithms, such as SLG and NC-Geo. However, these algorithms have obvious shortcomings in practical applications because they all need to measure the delay between nodes within a city, which is difficult to measure accurately. This limitation leaves room for improving the geolocation accuracy of the algorithms. In this paper, we explore the high-precision geolocation algorithm and propose a street-level geolocation algorithm based on router multilevel partitioning. This algorithm does not utilize the delay between nodes within a city, but it achieves higher geolocation accuracy than existing algorithms. Both the theoretical analysis and experimental results demonstrate that the proposed algorithm has obvious advantages relative to the street-level geolocation algorithms SLG and NC-Geo.

However, for targets that are not connected to the landmark through a common router, the proposed algorithm geolocates them only in the city center, which may be inaccurate. Therefore, in the future, we will focus on how to make use of more target-related network attributes to achieve high precision in the absence of a common router.

REFERENCES

[1] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, "A look at router geolocation in public and commercial databases," in *Proc. Internet Meas. Conf.*, Nov. 2017, pp. 463–469.

[2] D. Li *et al.*, "IP-geolocation mapping for moderately connected Internet regions," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 381–391, Feb. 2013.

[3] (2019). *Maxmind*. [Online]. Available: www.maxmind.com

[4] (2019). *Quova*. [Online]. Available: www.quova.com

[5] (2019). *IP2location*. [Online]. Available: www.ip2location.com

[6] (2019). *NetAcuity*. [Online]. Available: www.digitalelement.com

[7] O. Dan, V. Parikh, and B. D. Davison, "Improving IP geolocation using query logs," in *Proc. 9th Int. Conf. Web Search Data Mining*, Feb. 2016, pp. 347–356.

[8] C. Guo, Y. Liu, W. Shen, H. J. Wang, Q. Yu, and Y. Zhang, "Mining the Web and the Internet for accurate IP address geolocations," in *Proc. INFOCOM*, Apr. 2009, pp. 2841–2845.

[9] H. Liu, Y. Zhang, Y. Zhou, D. Zhang, X. Fu, and K. K. Ramakrishnan, "Mining checkins from location-sharing services for client-independent IP geolocation," in *Proc. IEEE Conf. Comput. Commun.*, Apr./May 2014, pp. 619–627.

[10] V. N. Padmanabhan and L. Subramanian, "An investigation of geographic mapping techniques for Internet hosts," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 31, no. 4, pp. 173–185, Oct. 2001.

[11] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-based geolocation of Internet hosts," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1219–1232, Dec. 2006.

[12] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards IP geolocation using delay and topology measurements," in *Proc. 6th ACM SIGCOMM Conf. Internet Meas.*, Oct. 2006, pp. 71–84.

[13] B. Wong, I. Stoyanov, and E. G. Sirer, "Octant: A comprehensive framework for the geolocalization of Internet hosts," in *Proc. 4th USENIX Conf. Netw. Syst. Design Implement.*, Apr. 2007, pp. 313–326.

[14] B. Eriksson, P. Barford, J. Sommers, and R. Nowak, "A learning-based approach for IP geolocation," in *Proc. Int. Conf. Passive Act. Netw. Meas.*, Apr. 2010, pp. 171–180.

[15] S. Liu, F. Liu, F. Zhao, L. Chai, and X. Luo, "IP city-level geolocation based on the PoP-level network topology analysis," in *Proc. 6th Int. Conf. Inf. Commun. Manage.*, Oct. 2016, pp. 109–114.

[16] G. Ciavarrini, M. S. Greco, and A. Vecchio, "Geolocation of Internet hosts: Accuracy limits through Camér–Rao lower bound," *Comput. Netw.*, vol. 135, pp. 70–80, Apr. 2018.

[17] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang, "Towards street-level client independent ip geolocation," in *Proc. 8th USENIX Conf. Netw. Syst. Design Implement.*, Mar. 2011, pp. 365–379.

[18] G. Ciavarrini, V. Luconi, and A. Vecchio, "Smartphone-based geolocation of Internet hosts," *Comput. Netw.*, vol. 116, pp. 22–32, Apr. 2017.

[19] S. Zu, X. Luo, S. Liu, Y. Liu, and F. Liu, "City-level IP geolocation algorithm based on PoP network topology," *IEEE Access*, vol. 6, pp. 64867–64875, 2018.

[20] J.-N. Chen, F.-L. Liu, Y.-F. Shi, and X. Luo, "Towards IP location estimation using the nearest common router," *J. Internet Technol.*, vol. 19, no. 7, pp. 2097–2110, Dec. 2018.

[21] A. Prieditis and G. Chen, "Mapping the Internet: Geolocating routers by using machine learning," in *Proc. 4th Int. Conf. Comput. Geospatial Res. Appl.*, Jul. 2013, pp. 101–105.

[22] J.-F. Grailet, F. Tarissan, and B. Donnet, "TreeNET: Discovering and connecting subnets," in *Proc. 8th Int. Workshop Traffic Monit. Anal.*, Apr. 2016, pp. 1–8.

[23] R. Li, Y. Sun, J. Hu, T. Ma, and X. Luo, "Street-level landmark evaluation based on nearest routers," *Secur. Commun. Netw.*, vol. 2018, Jul. 2018, Art. no. 2507293. doi: 10.1155/2018/2507293.

FAN ZHAO received the B.S. degree from the Nanjing University of Posts and Telecommunications, in 2012, and the M.S. degree from the State Key Laboratory of Mathematical Engineering and Advanced Computing, in 2015. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Mathematical Engineering and Advanced Computing, and the Zhengzhou Science and Technology Institute, under the supervision of Prof. X. Luo.

His research interests include network security and IP geolocation.

RUI XU was born in 1977. She received the B.S. degree from Sichuan University, in 2000, and the M.S. degree from the Southwest Institute of Communications, in 2003. She is currently a Researcher with the Cyberspace Security Key Laboratory of Sichuan Province and Cyberspace Security Technology Key Laboratory of CETC (CSTKL and CETC), Chengdu, China.

She was supported by the National Key Research and Development Program of China. Her research focuses on cyberspace security, cyberspace surveying, and mapping.

RUIXIANG LI received the B.S. degree in information security from the Zhengzhou Science and Technology Institute, China, in 2015, where he is currently pursuing the master's degree.

His research interests include network entity geolocation, data analysis, and information security.

MA ZHU received the B.S. degree from Henan Normal University, in 2004, and the M.S. degree from the University of Electronic Science and Technology of China, in 2004. She is currently an Associate Professor with the Zhengzhou Science and Technology Institute.

She was supported by the National Key Research and Development Program of China. Her research interests include network data analysis, network entity geolocation, and cyberspace security.

XIANGYANG LUO received the B.S., M.S., and Ph.D. degrees from the State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China, in 2001, 2004, and 2010, respectively.

He has been with the State Key Laboratory of Mathematical Engineering and Advanced Computing, since 2004. From 2006 to 2007, he was a Visiting Scholar with the Department of Computer Science and Technology, Tsinghua University. Since 2011, he has held a postdoctoral position at the Institute of China Electronic System Equipment Engineering Co., Ltd. He was supported by the National Natural Science Foundation of China, the National Key Research and Development Program of China, and the Plan for Scientific Innovation Talent of Henan Province. He has authored or coauthored more than 100 refereed international journal and conference papers. His research interests include network topology, network security, and network geolocation.