WILEY | Hindawi

Research Article

An Efficient Geolocation Method for Malicious LBSD Users Based on Dynamic Adjustment of Probes

Wenqi Shi [ID],[1,2] **Xiangyang Luo** [ID],[1,2] **Jiashan Guo** [ID],[1,2] **Fenlin Liu** [ID],[1,2] **and Lingling Li** [ID][3]

[1]*State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China*
[2]*Key Laboratory of Cyberspace Situation Awareness of Henan Province, Zhengzhou 450001, China*
[3]*School of Intelligent Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450046, China*

Correspondence should be addressed to Xiangyang Luo; luoxy_ieu@sina.com

The integration of the Internet of Things (IoT) and social networks is a promising trend of network technology. However, the diversity of social networks also poses potential risks to IoT security. Researching on the geolocation of social network users can verify the effectiveness of location protection mechanisms adopted by service providers, as well as provide a means for geolocating miscreants in social networks. Most current research focuses on how to infer the true location of a target within a specific region, such as within a city, while less research has been done on how to achieve fast and accurate localization of targets under long-range conditions. In this manuscript, an efficient localization method for LBSD users at long distances based on dynamic adjustment of probes (DAPL) is proposed. Based on the analysis of factors that affect the accuracy and efficiency of the target location approximation, DAPL can approach the real location sustainability of the target by dynamically generating probe locations. By identifying abnormal fluctuations of the target's reported distance, timely corrections of probe location are made to improve efficiency. In experimental results for Momo, a global popular LBSD social platform with more than 115 million active users show that even the initial probe is thousands of miles away from the target, DAPL can geolocate the target with a success rate close to 100% (99.5%), which is much higher than 70.6% of the existing method. Only about 12 times of LBSD service queries are needed, and DAPL can geolocate 88.9% of targets within 40 meters with an average error of 22.1 meters, which has higher efficiency and approximate accuracy compared with the existing typical method.

1. Introduction

With the development of information technology, the integration of Internet of things (IoTs) and social networks has become an important development trend [1, 2]. Global popular social platforms such as WeChat and Facebook have released IoT platforms [3]. On the one hand, this integration further accelerates the development of IoT. On the other hand, the illegal behavior of malicious social network users, such as spreading rumors and disseminating viruses, poses a great threat to IoT security too. Broomium, a network security company, released a survey report in 2019 which shows that the global social network crime scale is as high as $3.2 billion per year [4]. Research on social network user

geolocation can provide positive means for crime investigation of malicious social network users [5, 6]. Meanwhile, the location privacy is an essential concern for both IoT and social network security. Research on social network user geolocation can verify whether the location privacy of social network users is effectively protected [7–10]. In this manuscript, we focus on an important location-based service of social networks [11, 12], location-based social discovery (LBSD), and explore whether the location of social network users can be inferred based on the public services provided by social networks. LBSD is a popular type of LBS, which enables users to find their nearby users, thus facilitating the establishment of social relationships among users [13]. Most of the worldwide popular mobile social platforms

provide LBSD services such as Facebook, WeChat, Telegram, and Momo. While LBSD services provide convenience to users, there is also a risk of threatening their location privacy. Since most LBSD services do not provide API for obtaining data, such as user text or social connections, most current LBSD user geolocation methods utilize relative distance information to geolocate the target. Existing localization methods for LBSD users can be generally classified into three main types: trilateration-based, number theory-based, and successive approximation-based method.

Trilateration model is a classical localization method which infers the location of the target by obtaining the target's exact distance at three known locations [14, 15]. However, in practical situations, it is not easy to obtain the exact distance of the target user. To address this problem, Ding et al. [16] propose an enhanced trilateration method based on distance segments, which takes the localization results of the target in the intersection of multiple circular regions centered on probes' location. The method can geolocate and track WeChat users within the city. In papers [17, 18], the relationship between nearby user order and their distance is studied. The upper and lower limits of the target's distance are delimited based on the distance of users who are adjacent to the target in nearby user list. Then, the target's location is inferred using the trilateration model. Trilateration-based method is efficient and easy to implement, which has high accuracy if the distance range can be delimited accurately; otherwise, the localization error is large.

The number theory-based method describes the relationship between the actual and reported distance into mathematical models. By strategically deploying probes in a specific area, the location of the target is calculated based on the target's reported distance obtained by each probe. Xue et al. [19] analyze the feature that WeChat report relative distance between nearby users in bands and propose a localization algorithm, which study the localization of the target on a line, then extended it to two-dimensional space. Based on Xue's work, Cheng et al. [20] point out that the initial probe's location will affect the localization accuracy of one-dimensional method and propose a new deployment strategy of the initial probe. Considering that the LBSD service will confuse the nearby user's distance, Peng et al. [21] propose a two-dimensional localization method based on heuristic number theory, which further improves the practicality. The theoretical accuracy of number theory-based methods is high, but it lacks of experimental verification of the location confusion mechanism. The big difference between the mathematical model and the actual distance reported strategies leads to the practical accuracy is hard to achieve the theoretical accuracy.

The successive approximation-based method firstly delineates the initial space where the target may be located based on its reported distance obtained by the first probe; then, the target's real location is gradually approximated based on the changes of the reported distance obtained by probes whose location are adjusted continuously. Li et al. [22] investigate the LBSD services of some popular social platforms, WeChat, Momo, and Skout, and their location confusion strategies. Momo and Skout users are geolocated based on iterative trilateration, which continuously approach the target's true location by iteratively using trilateral localization. Meanwhile, a space partition-based method is proposed to break the limit of the minimum reported distance. Subsequent works further investigate how to improve the localization accuracy within a specific region [23–25]. In response to the location confusion of LBSD services, paper [26] proposes a target localization method based on orientation identification, which estimates the orientation of the target relative to the central probe based on the target's reported distance obtained by surrounding probes and then iteratively approximates the target location. Paper [27] discusses the target localization problem when the reported distance is confused with Gaussian noise, which delimits the small range where the target is located based on iterative trilateration firstly and then estimates target's location within the small range based on maximum likelihood estimation. Successive approximation-based method is more robust which can resist the influence of random noise of LBSD. However, this type of method focuses more on how to improve the localization accuracy based on the premise that the specific area where the target is located is known, while further research is needed to achieve fast and accurate localization of the target when the specific area is unknown.

In sum, most of the existing research focuses on target localization within a specific area (e.g., within a city), while there is a lack of effective means for localization of long-range targets (thousands of kilometers away). Since most LBSD social software restrict access to user profiles, the existing social user location inferring methods, such as text-based or social relationship-based, are not applicable to the localization of LBSD targets [28]. To address this problem, in this manuscript, an efficient long-range LBSD user localization method based on dynamic adjustment of probes (DAPL) is proposed. DAPL achieves fast approximation of the target's real location by deploying probes dynamically and improves the localization efficiency by monitoring the abnormal change of target's reported distance. The main contributions of our work are shown as follows:

(i) Factors that affect the accuracy of target geolocation under long-range conditions are analyzed, and corresponding solutions are given. We analyze the factors affecting the accuracy and the efficiency of target position approximation from two perspectives, respectively: maximum included angle and minimum distance between probes. The effectiveness of the proposed solutions is verified through practical experiments

(ii) A long-range localization method for LBSD users based on the dynamic adjustment of probes (DAPL) is proposed. The proposed DAPL achieves fast approach to the real location of the target by dynamically updating probes' location and performing trilateral localization. What is more, the localization efficiency is improved by detecting and disposing of anomalous probe combinations. Actual experimental

results for Momo user localization show that DAPL has significantly higher success rate, better efficiency, and similar accuracy compared with the typical existing method

The rest of the manuscript is organized as follows. In Section 2, the LBSD service and existing long-range localization methods are a briefly introduced. Section 3 introduces basic principles and main steps of DAPL. In Section 4, we explain and illustrate the key steps of DAPL in more detail. Section 5 describes the real-world experiment. Finally, Section 6 summarizes the manuscript.

2. LBSD Service and Related Work

This section provides a brief introduction to LBSD services and the current privacy protection strategies commonly used by service providers as well as an analysis of the shortcomings of existing methods for long-range localization of LBSD users.

2.1. Introduction of LBSD Service. The LBSD service is a kind of popular LBS, also known as the proximity discovery service. With this service, users can discover other users close to the location of their mobile devices. Many mobile social applications, such as WeChat, Momo, TanTan, QQ, Facebook, and Telegram, provide this service to meet users' needs for making friends. Typical scenario of the LBSD service is shown in Figure 1.

As shown in Figure 1, when a user uses LBSD service, the application will submit a request to the LBSD server for querying nearby users. After receiving the request, the server will screen the database that stores all the users' location information and response to the quarrier with location-related information and profiles of users who are close to the quarrier's location. Then, the application parses out the profile and location information (usually in form of distance) of the nearby users from the response and displays them to the quarrier. When the application sends a request to the server, its accurate location is obtained by calling the embedded GPS interface of the device (e.g., WeChat and QQ) or calling the third-party map API (e.g., Momo and TanTan) [29].

To protect the privacy of normal users, most LBSD services will confuse the location information (or relative distance) of nearby users. The common means of location confusion are as follows.

Hide Coordinates. Instead of displaying the exact coordinates of nearby users or their locations on the map, the application will only display the relative distance of nearby users. According to our investigation, rarely LBSD services will show the exact location of nearby users

Display Distance by Segment. Instead of reporting the precise distance of nearby users, the server will set a minimum distance granularity, and the distance between users displayed is a multiple of the minimum granularity. For example, the minimum distance displayed by WeChat is 100 meters, for QQ and Momo is 10 meters, and for Skout is 0.5 miles

Add Random Noise. Some LBSD services add random noise to the distances, making the distance of a same nearby user different for multiple queries at the same location. In this way, the displayed distance does not directly reflect the accurate distance or narrow distance range of nearby users, thus protecting user privacy. This is a widely used means for distance confusion

Hide Distance. Some LBSD services will not to show distance information of nearby users and only display the nickname, signature, gender, and other profile information of nearby users. For example, Grindr, a popular location based dating App, sorts nearby users by their distance but does not display the user's distance on the APP page

In addition, the LBSD service adopts many measures to prevent the service from being abused. The commonly used measures are as follows

Limit the Query Range. The server only reports nearby users within a specific distance from the querier. If a user is far away from the quarrier, she (he) will not be found by the querier. If LBSD service has this limitation, the target cannot be geolocated by long-range localization method because if the target is too far, his reported distance will not be obtained by probes continuously

Restrict the Query Frequency. Some APPs will limit the number of times an account can use the LBSD service per day. When the number of uses exceeds the threshold, the LBSD service will be shut down for a period of time

Set a Time Domain. The server will periodically clean the database of user location information. When a user is no longer active for a period of time, the server will not recommend her/him to other users

Besides, some LBSD services will only show the relative distance of friends and strangers cannot be found. Typical social platforms that provide LBSD services are shown in Table 1.

2.2. Analysis of Existing Long-Range Localization Methods of LBSD Users. As we mentioned in Introduction, most of the existing studies focus on precise localization of LBSD targets within a specific region. For example, the prior knowledge that the city where the target is located is known. However, the city-level location of LBSD users is not easy to infer. Existing city-level location inference methods for social users rely on the available of abundant user data, such as social relationships and generated text of the target. This kind of method is suitable for open social platforms such as Twitter and Weibo [28]. Unlike Twitter and Weibo, most LBSD service do not provide API interfaces for crawling user data and social relationships between users are unknown, which makes the research on localization of LBSD users under long-range conditions is of practical significance.

By analyzing the accuracy of the reported distance of LBSD services in Momo and Skout, Li et al. propose a long-range target location approximation method based on iterative trilateration (short for ITBL) [22]. The principle of ITBL is shown in Figure 2. ITBL takes advantage of the feature that there is no distance range limitation when Momo and Skout provide LBSD services and uses the trilateration to achieve long-range target localization. Firstly,
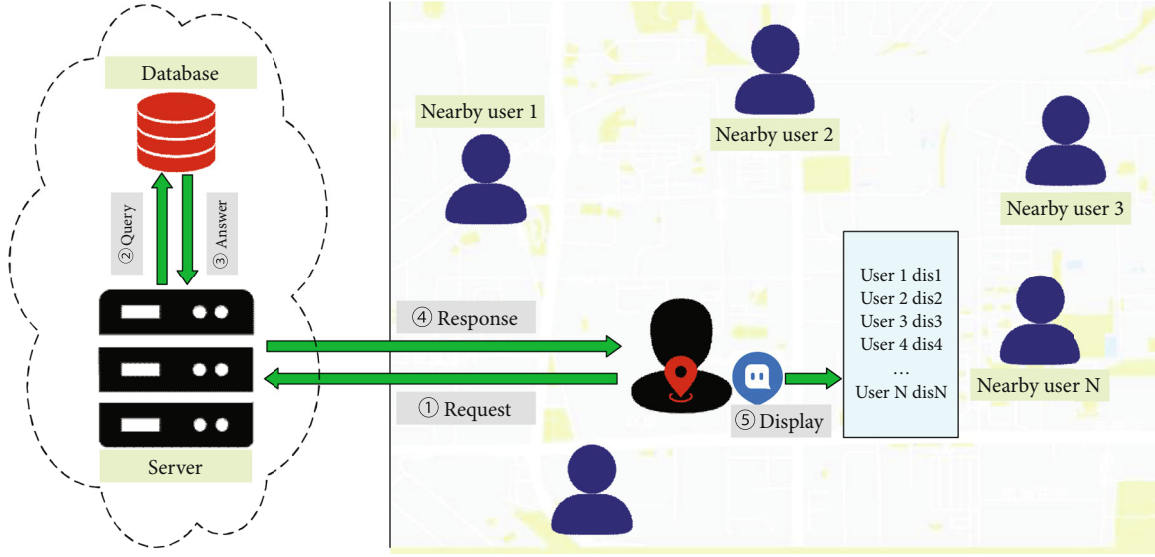
FIGURE 1: Typical scenario of LBSD service.

TABLE1: Typical social platforms with LBSD service.

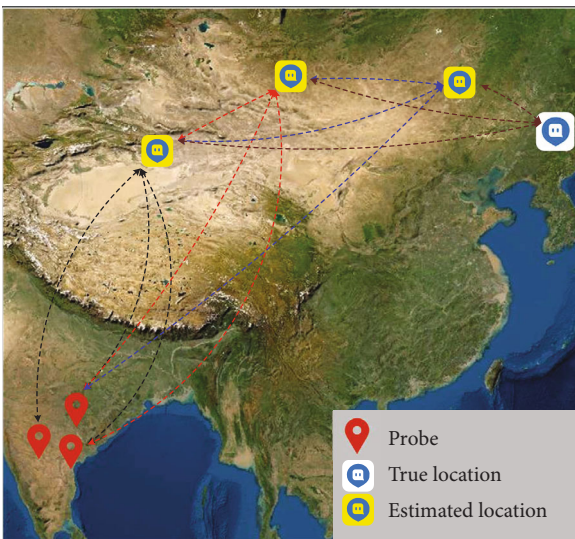| Social platforms	Number of active users (millions)	Distance	Minimum reported distance	Random noise	Restriction
WeChat	1241.6	√	100 m	√	Time/range/frequency
QQ	606.4	√	100 m	√	Time/range
Telegram	500	√	1 m	√	Time/range
Line	220	√	100 m	√	Range/time
Momo	115.5	√	10 m	√	No
TanTan	26	√	100 m	√	Range
Skout	10	√	0.5mile	√	No
Grindr	8	×	N/A	×	Range
SayHi	N/A	√	10 m	×	Range

FIGURE 2: Schematic diagram of geolocating method based on iterative trilateration.

locations of three probes are randomly set which query the target's reported distance, respectively. Then, based on probes' location and the corresponding target's reported distance, estimated location of the target is calculated by the least square method [30]. Thirdly, a new probe is set at the estimated location, which queries the reported distance of the target again. The probe with the largest reported distance is eliminated, and the remaining three probes are used to geolocate the target again. The above process is performed iteratively until the maximum distance between probes is less than a certain threshold. Finally, the position of the probe with the nearest distance is selected as the final result.

ITBL takes the systematic errors in reported distance into account and uses the least square method to estimate the trilateration results, which has low time complexity. However, the effectiveness of ITBL depends on the smaller distance between target's estimated location and its true location. If the probe in the estimated location obtains a greater distance of the target compared with the three initial probes, as shown in Figure 3, ITBL will continue to select previous three probes to perform trilateration. If the target's reported distance obtained by these probes remains

FIGURE 3: Schematic diagram of failed geolocation.

unchanged, the algorithm will not be able to approach the target any more, which results in localization failure. ITBL estimates the trilateration result based on least square method, which output a location that minimizes the square sum of the difference between the real distance and the reported distance. Influence of noise on the accuracy of different distance range is not considered, and it cannot guarantee that the estimated location is closer to the target. Our practical tests also verify that ITBL has a high failure probability, which will be introduced further in Section 4.

Different from the existing ITBL method, DAPL determines the location of probes for trilateration based on the target's reported distance obtained by the previous probes, so the probe distribution that may lead to failed localization can be avoided. With the dynamic adjusting of probes' location, the localization success rate is effectively improved without enlarging the time complexity and localization errors.

3. Proposed method

In response to the fact that most existing methods study target localization within a specific area and lack of effective target geolocation means under long-distance conditions, a long-range localization method for LBSD users based on the dynamic adjustment of probes is proposed. DAPL dynamically generates probe positions by reconstructing the coordinate system to ensure the sustainability of the localization process. In addition, location of probes is timely corrected based on the change of target's reported distance, thus avoiding the increase of time overhead due to distance oscillation.

3.1. Background Knowledge. To facilitate understanding, some background knowledge and definitions of terms involved in the algorithm are first introduced.

Probe: it is essentially social accounts whose location is known. A probe obtains the distance of the target by querying the LBSD service. Depending on the difference of location determination, probes are divided into anchor probes and generated probes. The location of the anchor probe is set randomly at the beginning of localization process or calculated from the previous round of trilateration. The location of the generated probe is generated based on the anchor probe's location, which is on the axis of the coordinate system with the anchor probe as the origin

Reported Distance: the distance of nearby users is displayed by the LBSD service, which is related to the actual distance. But the existence of systemic noise makes the reported distance to be inaccurate

Potential Area: the potential area is the smallest geographic area where the target may be located, which is determined by prior knowledge. It should be noted that our work discusses target localization when the location-related prior knowledge of the target (e.g., provincial location) is unknown. If the potential area of the target is known, the method proposed in this manuscript is still valid and will perform better

Anomalous Probe Combination: when three probes are used for iterative trilateration, the localization process may be interrupted due to conditions such as probe location distribution and reported distance confusion. When this condition happens, the combination of three probe locations is called the anomalous probe combination. A typical

anomalous probe combination is that when the estimated location of trilateration is farther away from the target than the three probes, if the same three probes with the smallest reported distance of the target are selected to perform next-round trilateration, the same situation will appear again and the geolocation procedure will be interrupted. If we want to achieve the continuous approximation of the real location of the target, the anomalous probe combination should be discovered and deciphered in time

Distance Oscillation: distance oscillation refers to the phenomenon that in the process of geolocating the target, the target's reported distance queried by the anchor probe gradually increases for several rounds of trilateration, and then, the anchor probe gradually converges to the target. The distance oscillation phenomenon is caused by the large difference between the estimated and the actual location of the target, which can significantly enlarge the time consumption. In our method, we mitigate distance oscillation by inverting the coordinates of the generated probes

Least Square Method: Least square method (LS for short) is a kind of statistical method. By minimizing the square sum of errors, LS can find the best matching function of data [31]. It is widely used in the field of sensor node positioning [30, 32, 33]. Setting the location of probe p_i is (x_i, y_i), and the real position of the target is (x, y). d_i represents the distance of the target obtained by probe p_i. Due to the existence of systemic noise, d_i is not accurate. Error equation v_i can be expressed as

$$v_i = \left[(x_i - x)^2 + (y_i - y)^2 \right]^{1/2} - d_i, i = 0, 1, \cdots, N, \quad (1)$$

where N represent the number of probes. The purpose of the least square method is to find X that minimizes the square sum of v_i, that is

$$(\text{LS}): \quad \min_X \sum_{i=0}^{N} \left(v_i^2 \right). \quad (2)$$

The error equation v_i is nonlinear constraints, which should be converted to the form of linear equation based on its Taylor expansions, as shown in the following:

$$v_i = (l_i, m_i) \left(\delta_x, \delta_y \right)^T + d_0^i - d_i, i = 0, 1, \cdots, N, \quad (3)$$

where $l_i = (x_0 - x_i)/d_0^i$, $m_i = (y_0 - y_i)/d_0^i$, and $d_0^i = \left[(x_i - x_0)^2 + (y_i - y_0)^2 \right]^{1/2}$. So, the constraints above can be expressed in the following matrix form:

$$V = AX + \boldsymbol{b}. \quad (4)$$

In equation (4), $V = (v_1, v_2, \cdots, v_N)$, and $X = (\delta_x, \delta_y)^T$ represents the deviation of the real position of the target

from (x_0, y_0), $\boldsymbol{b} = (d_0^1 - d_1, d_0^2 - d_2, \cdots, d_0^N - d_N)^T$, and

$$A = \begin{pmatrix} l_1 & m_1 \\ l_2 & m_2 \\ \cdots \\ l_N & m_N \end{pmatrix}. \quad (5)$$

The solution of LS is $X = (A^T A)^{-1} (A^T \boldsymbol{b})$, and the final evaluation result of the target location is $\hat{X} = (X_0 + X)$

The LBSD user localization problem can be regarded as the problem of finding the optimal solution under distance constraints. More clearly, on the basis of the location of multiple probes and the imprecise distance between the target and the probes are known, we want to infer the accurate location of the target. For nonlinear objective function, Newton method or quasi-Newton method can also be used to estimate the least square optimal solution [34].

3.2. Basic Principle and Main Steps of DAPL.
In this manuscript, DAPL, a LBSD user geolocation method based on the dynamic adjustment of probes, is proposed. We analyze the location obfuscation strategies in typical LBSD services and point out anomalous probe combination and distance oscillation problems in existing methods, which lead to low success rate and high time complexity. To address this, we alleviate the anomalous probe combination problem by introducing generated probes and solve the distance oscillation problem by inverting the coordinates of generated probes.

The basic principle of DAPL is shown in Figure 4, which mainly consists of two parts: target location approximation and probe location correction. The first part uses multiple rounds of trilateration to geolocate the target. In each round trilateration, the location of probes is determined by continuously reconstructing the coordinate system and the intermediate result in each round is estimated by the least square method. The second part identifies the anomalous probe combinations in each round of trilateration by monitoring the change of target's reported distance obtained by the anchor probe. The coordinates of generated probes are adjusted if the abnormal change of target's reported distance appears, so as to avoid the distance oscillation phenomenon and improve the efficiency.

The main steps of DAPL are as follows.

Step 1 : Initial probe generation. Firstly, an initial anchor probe P_1 is deployed at a randomly selected location within the potential area, whose location is noted as L_1. If the potential location area of the target is unknown, the initial anchor probe's location is randomly selected globally. Then, the initial anchor probe queries the LBSD service to obtain the target's reported distance, which is noted as R_1.

Step 2 : Establish the coordinate system. A Euclidean coordinate system is established with L_1 as the origin, and probes P_2 and P_3 are set on the coordinate axes whose locations are $(R_1/2, 0)$ and $(0, R_1/2)$, which are recorded as L_2 and L_3, respectively. If L_2 and L_3 are outside the potential area,
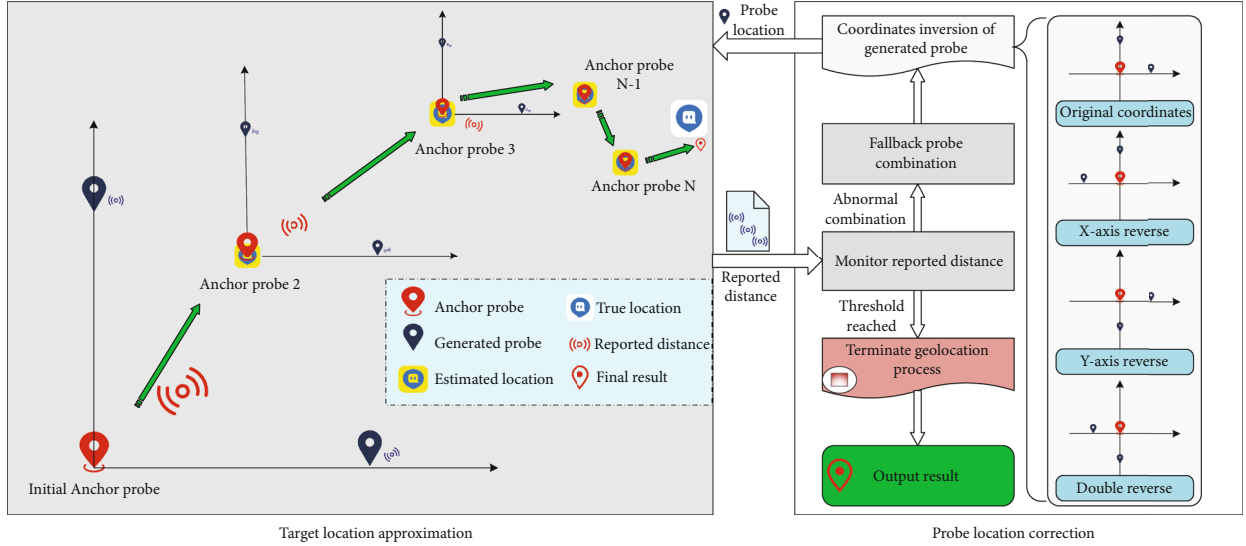
FIGURE 4: Diagram of the proposed method.

the coordinates of P_2 and P_3 are set to crossover points of the coordinate axes and the potential area. The reported distances of the target obtained by P_2 and P_3 are recorded as R_2 and R_3, respectively.

Step 3 : Estimate the target's position. With $\mathbf{L} = \{L_i | 1 \leq i \leq 3\}$ and $\mathbf{R} = \{R_i | 1 \leq i \leq 3\}$ as parameters, the LS method is used to calculate the estimated location of the target, which is noted as L_e. Objective equation f is formulated as

$$f(X) = \sum_{i=0}^{m} (Tdis(L_i, X) - R_i)^2, \qquad (6)$$

where $Tdis(P_1, P_2)$ represents the actual distance between two points P_1 and P_2. Since complex operations are required to calculate the distance based on latitude and longitude, it is difficult to convert the objective equation f into matrix form. So, the optimal value X corresponding to the minimum of the objective equation f is calculated by using Newton's method [34].

$$L_e = \min_X f(X). \qquad (7)$$

The anchor probe is deployed at position L_e, and the anchor probe queries the target's reported distance R_e.

Step 4 : Anomalous probe combination identification and probe location correction. After R_e is acquired, Algorithm 1 is utilized to verify whether the probe for trilateration is an anomalous combination. The algorithm takes $\mathbf{L}$, $\mathbf{R}$, L_e, R_e, and the threshold ϵ as inputs and outputs the execution code of the next step and a list containing the estimated position of the target in the coordinate inversion process and the corresponding target's reported distance.

Step 5 : Target localization. If the code that returned by Step 4 is 1, it means that the threshold value ϵ is achieved in the

process of coordinate inversion and the localization is success. The location of probe which obtains minimum target's reported distance is output as the final localization result.

If the code that returned by Step 4 is 2, it indicates that the anomalous probe combination still exists after three times of coordinate inversion. The output EL contains estimated results of target position $EL[i][0]$ obtained by different combinations of probe coordinates and the corresponding reported distance of the target $EL[i][1]$. $EL[i][0]$ corresponding to the smallest $EL[i][1]$ is selected as the final result of this round of trilateration, and the anchor probe is set at $EL[i][0]$, as shown below, then jumping to Step 2.

$$L_1 = \left\{ \begin{array}{c} EL[i][0] | EL[i][1] = \min (EL[j][1]), \\ 0 \leq j < len(EL) \end{array} \right\}. \qquad (8)$$

If code return by Step 4 is 0, it indicates that after Step 3, the target's reported distance obtained at the estimated position is greater than the maximum reported distance obtained by the three probes. The probes are treated as anomalous probe combination. Generated probes have performed coordinate inversion in Step 4, and the estimation result is successfully obtained which is closer to the target. Then, the location of the anchor probe is updated as below, and jump to Step 2.

$$L_1 = \{EL[i][0] | i = len(EL) - 1\}. \qquad (9)$$

In the above process, how to dynamically set the probes and cope with the anomalous probe combination are the key of DAPL, which will be introduced in detailed in Section 4.

4. Target Location Approximation Based on Dynamic Adjustment of Probes

As introduced in Section 2, most LBSD services obfuscate the reported distance to protect user privacy. Existing

```
Input: L, R, R_e, L_e, ε
Output: execution code.
1 code = 0; //Initial value of execution code.
2 EL = (L_e, R_e); //List of L_e to be evaluated.
3 i = 1; //Times of coordinate inversion.
4 axis = X;
5 ifR_e ≤ εthen://Judge whether the localization is successful.
6 code = 1;
7 break; //End the localization.
8 end.
9 whileR_e > max (R)do://Detect anomalous probe combination.
10 ifi ≥ 3 then://All coordinate inversions are tried.
11 code = 2;
12 break; //End inversion process.
13 end.
14 if axis == X then://Invert coordinates of x-axis generated probe.
15 L_2 = (−R_1/2, 0);
16 L_e = LS(L, R); //Estimate target location based on new probes.
17 update R_e; //Obtain target distance at the evaluation location.
18 EL.append((L_e, R_e));
19 i = i + 1;
20 axis = Y;
21 continue; //Verification.
22 end.
23 else.
24 if axis == Y then://Invert coordinates of the probe on the y-axis.
25 L_2 = (R_1/2, 0);
26 L_3 = (0, −R_1/2);
27 L_e = LS(L, R); //Estimate target location based on new probes.
28 update R_e; //Obtain target distance at the evaluation location.
29 EL.append((L_e, R_e));
30 i = i + 1;
31 axis = XY.
32 continue; //Verification.
33 end.
34 else.
35 L_2 = (−R_1/2, 0); //Invert coordinates of 2 generated probes.
36 L_3 = (0, −R_1/2);
37 L_e = LS(L, R); //Estimate location based on new probes.
38 update R_e; //Obtain target distance at the evaluation location.
39 EL.append((L_e, R_e));
40 i = i + 1;
41 continue; //Verification.
42 end.
43 end
44 return:EL, code.
```

ALGORITHM 1: Dispose of anomalous probes combinations based on coordinate inversion.

literature states that the randomness of the reported distance in most LBSD services increases as the reported distance increases [18, 23, 26]. This situation should be taken into account when geolocating a target at a long distance, so as to the selection of probes' location can contribute to a continuous approximation of the target's true location.

4.1. Improving Localization Success Rate by Reestablishing the Coordinate System. Trilateration is a classical localization method with high efficiency. Existing LBSD user localization method based on iterative trilateration can achieve fast approximation of the target's true position at long distances,

but it is susceptible to reported distance obfuscation adopted by the service provider. Existing studies show that in most LBSD services which do not restrict the query range, the systemic noise in the reported distance increases with the actual distance [22, 26, 27]. When users are thousands of kilometers apart, there may be tens or even hundreds of kilometers of errors in the reported distance. Meanwhile, when users are within a few kilometers of each other, the error range of the reported distance may be hundreds of meters. When the least square method is used to estimate the target location without considering the variability of such systematic errors over different distance ranges, there is a possibility

that the estimated location is farther from the target compared with probes. In such case, the existing ITBL method will continuously select the previous three probes to geolocate the target, resulting in localization failure. Therefore, it is necessary to update the probe positions after each round of trilateration to ensure the sustainability of localization.

The location obfuscation strategy used by the LBSD server is unknown for normal users. If we want to realize a reasonable combination of probes' location, it is more practical to manage it from the perspective of the probe deployment. Based on an intuitive idea, we analyze the deployment of probes that may affect the success rate of trilateration in terms of two factors: interprobe distance and interprobe pinch angle.

Interprobe Angle: when three points are on a straight line and three circles are drawn centered on the point, respectively, the graph is symmetric about the straight line. It is known from the basic principle of trilateration that there will be at last two estimation results that satisfy the least square constraints. The probability is high that the distance between two possible estimation results is greater than the maximum reported distance of the target obtained by probes. To avoid this situation and facilitate the calculation, we reestablish the coordinate system to redefine the probe location for the next round of trilateration. With the anchor probe as the center, generated probes' location is selected in its due north and due east directions, respectively, and the maximum angle between the three probes is 90°. Thus, the problem of high localization failure rate caused by too large maximum angle can be avoided. The maximum included angle between the probes is shown in Figure 5. Obviously, the maximum angle between the probes θ is between 60° and 180°

Interprobe Distance: the distribution of probes' location can greatly affect the efficiency of target geolocation. When we want to geolocate the target from a long distance, the target's reported distance obtained by the initial anchor probe is more than thousands of kilometers. At this time, if the distance between probes which performing trilateration is small, the geolocation result is more vulnerable to the confusion of the reported distance, which will lead to a farther distance between the estimated location and target's real location, as shown in Figure 6(a). When the anchor probe is close to the target and the generated probe is far away from the target, since the confusion of the reported distance is stronger with the distance increases, the distance between the target's real location and the estimated location may also be farther than the anchor probe, as shown in Figure 6(b). Both these situations can have a significant negative impact on geolocating efficiency. Therefore, based on empirical values from actual tests, we determine the coordinates of the generated probe based on the target's reported distance obtained by the anchor probe

By reestablishing the coordinate system and determining the coordinates of the generated probe based on targets' reported distance acquired by the anchor probe, the problem of localization failure due to distance confusion in long-range conditions can be effectively avoided. In Section 5, the effectiveness of our method in terms of localization success rate will be verified based on practical tests.
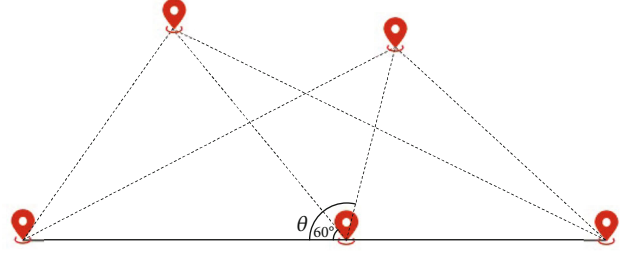

Figure 5: Schematic diagram of maximum included angle between probes.

4.2. Improving Localization Efficiency by Probe Coordinate Inversion. By reestablishing the coordinate system and controlling probe position distribution, the success rate of localization can be improved. However, abnormal probe combinations may still exist, which may lead to a decrease in geolocation efficiency due to the increase of LBSD service query. A countermeasure deals with the reduced efficiency based on the inversion of probe coordinates is proposed.

When the estimated position of a round of trilateration is farther from the target, a possible reason is that the generated probe is farther from the target than the anchor probe, which may lead to distance oscillations and increase the time consumption. If it is found that the target's reported distance obtained by the anchor probe is greater than the maximum reported distance of the target in previous round of trilateration, it means that there is still an abnormal probes combination in previous round of trilateration. In this case, the coordinates of the generated probe are inverted, as shown in Figure 7. As we mentioned before, in previous round of trilateration, the coordinates of generated probes which recorded as B and C are determined in due north and due east of the anchor probe which recorded as A. If the abnormal probe combination is detected, the position of one generated probe is inverted. The point in the coordinate system that is symmetrical to the origin (anchor probe) is taken as the new location of generated probe B'. Then, probe B' queries the target's reported distance again and the trilateration is performed again based on probes A, B', and C.

Whether the target is located in any quadrant of the coordinate system, by inverting the coordinates of generated probes, it is possible to make probes for performing trilateration in a reasonable combination. So that the distance oscillation phenomenon can be detected and avoided in time, and the efficiency loss in the localization process can be reduced. Results of practical experiments also verify the effectiveness of our method, which will be described in detail in Section 5.

5. Experiment and Analysis

To verify the effectiveness of proposed DAPL, we conduct practical experiments based on Momo platform and compare the experimental results with the existing methods. It should be noted that because there are few existing studies on long-distance geolocation of LBSD users, we select the

(a) Undersize interprobe distance

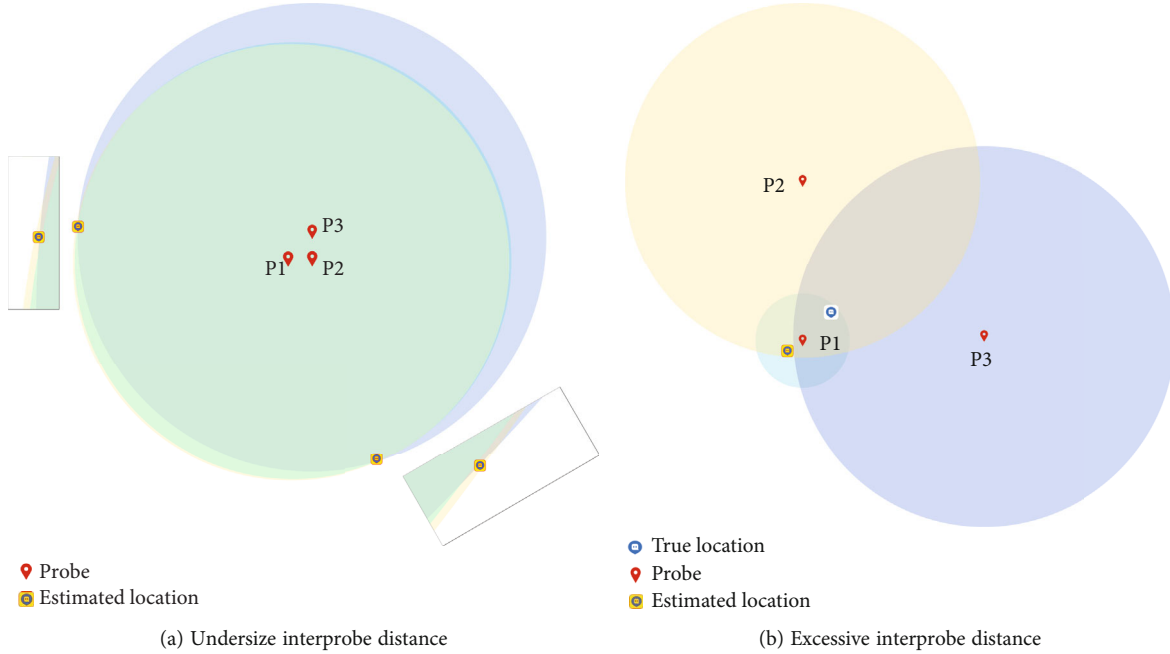(b) Excessive interprobe distance

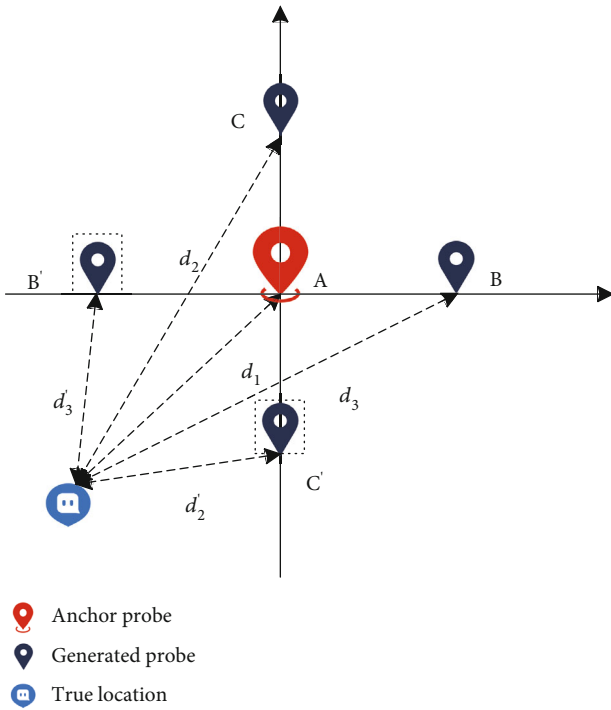FIGURE 6: Abnormal distribution of interprobe distance.

FIGURE 7: Schematic diagram of generated probe coordinate inversion.

most typical existing method based on iterative trilateration for comparison [22].

5.1. Experiments Settings. Momo is a popular LBSD application with hundreds of millions of active users worldwide which has the characteristics of unlimited query times and query range. So, it is a very suitable social platform to verify the proposed method. The experimental environment is

TABLE 2: Experimental environment.

Items	Setup
Social platform	Momo
CPU	Intel Core i7-7700 3.60GHz
RAM	16GB
Operating system	Windows 10 X64
Mock location	Nox player V7.0.0.8
Android version	Android 6.0.0
Automated testing tools	Appium
Least square solution	Python/SciPy/optimize/BFGS

shown in Table 2. We use the computer (CPU: Intel Core i7-7700, RAM: 16GB) that runs the Android simulator Nox player. Nox player can provide the virtual location function and can arbitrarily modify the location of most mobile apps, including Momo. Appium, an automatic test tool, can realize the automatic operation of software, such as startup, click, drop-down, and return. The least square method is realized by the optimize function in the SciPy package in Python, and the BFGS function is selected to calculate the optimal solution of the least square.

Based on the above experimental environment, the experiment lasted about 10 months. The experiment settings are shown in Table 3. Only two different Momo accounts are used for the experiment, one as a probe and the other as a target. Firstly, by setting the location of the simulator, the location of the target is set in any area of Heilongjiang Province at the northernmost of China, and the longitude and latitude of the target are recorded. Then, the probe's coordinates are randomly set in other provinces 3000-6000 kilometers away from the target. To verify the target geolocating effect of DAPL in long-distance conditions, the distance

TABLE 3: Experimental settings.

Items	Setup
Probe distribution	3000-6000 km from the target
Targets' location	Random distribution in Heilongjiang Province, China
Target quantity	5000
Threshold for ITBL	50 m
ϵ for proposed DAPL	20 m
Experimental time	February 2021 to November 2021

TABLE 4: Comparison of geolocation success rate.

Method	Items	3000-4000 km	4000-5000 km	5000-6000 km	Totally
	Target number	1373	2031	1596	5000
Proposed DAPL	Successful number	1369	2020	1587	4976
	Successful rate	99.6%	99.5%	99.4%	99.5%
	Target number	317	1757	2926	5000
ITBL [22]	Successful number	224	1272	2033	3529
	Successful rate	70.7%	72.4%	69.5%	70.6%

between probes and the target is set more than 3000 km, which is greater than the geographical range of most countries. The distance from the equator to the South and North Poles is about 6000 km, which is "long-range" enough. Even the target is geolocated globally, we can find a location within 6000 km from the target through a few queries. For this reason, we set the location of probes in the geographic area of 3000-6000 km to the target. The geolocating results are compared with the target's real location to calculate the geolocating error, and the query times to LBSD service in the geolocating process are recorded. DAPL only needs one anchor probe's position as input, and the probes' positions needed in the subsequent geolocating process are generated by the initial anchor probe. In contrast, the existing ITBL will randomly generate three initial probe positions and then perform geolocating. ITBL method ends geolocating when the maximum distance between probes is less than the threshold and selects the location of the probe closest to the target among the three probes as the geolocating result. The DAPL method ends the positioning when the reported distance of the target obtained by the probe is less than the threshold and selects the location of the probe corresponding to the minimum reported distance as the final result. Within a certain range, the smaller the threshold is, the higher the geolocating accuracy. Because the selection of thresholds of the two methods is different, based on the intuitive idea, in order to ensure similar accuracy, the threshold of ITBL method should be greater than that of DAPL method. In this manuscript, we set the thresholds of the two methods to 50 meters and 20 meters, respectively.

We have geolocated the target 5000 times in the real environment. It should be noted that Momo will set a time limit for location information update. More clearly, after a user queries the distance information of nearby people, server will lock the user's location status for a period of time. When the user moves to another location during this period

of time and queries the nearby people again, the server still returns the distance of the nearby people queried by the user at the previous location. After a large number of tests, we found that the time limit is about 1 minute. However, we can effectively shorten the time limit by killing the application and then restarting it, and the limit can be reduced to about 20 seconds.

5.2. Experimental Results and Analysis. The effectiveness of DAPL is analyzed from three aspects: success rate, geolocation accuracy, and service query times.

5.2.1. Geolocation Success Rate. In order to unify the standard, when the method fails to reach the preset threshold after more than 200 times of LBSD service query, it is regard as a failed geolocation. Under actual conditions, it will take several hours to carry out 200 times of service query since Momo provide no API to accelerate the process. During this time period, if the target user updates his location, the previous queries will be invalid, which will affect the practicability of the method.

Table 4 shows the success rate of the two methods when geolocating Momo users. ITBL successfully geolocates 3529 Momo targets, and the comprehensive success rate is about 70.6%. By comparison, the proposed DAPL geolocates 4976 Momo targets successfully, whose comprehensive success rate is 99.5%, which is much higher than ITBL. According to the distance between the target and the initial probe, the whole geolocating process is divided into three categories, 3000-4000 km, 4000-5000 km, and greater than 5000 km, respectively. The geolocation success rates in different distance ranges are discussed, respectively. For ITBL, the geolocation result is classified according to the distance of the farthest of the three probes. It can be seen from the table that under different categories, the success rates are similar for both two methods and the distance between the
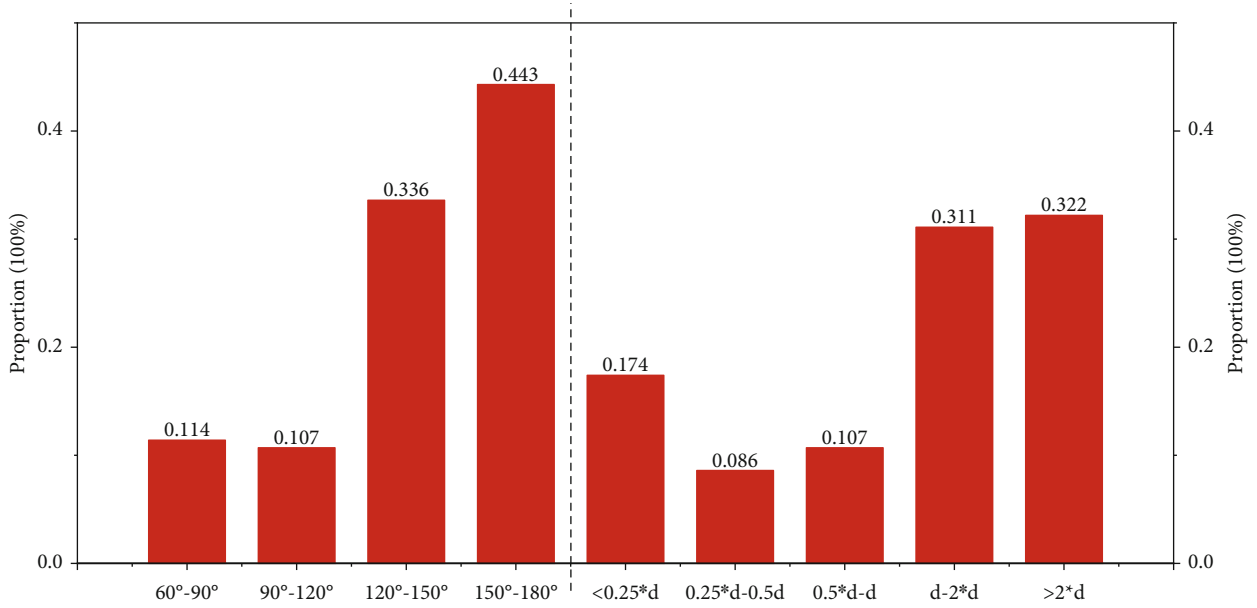
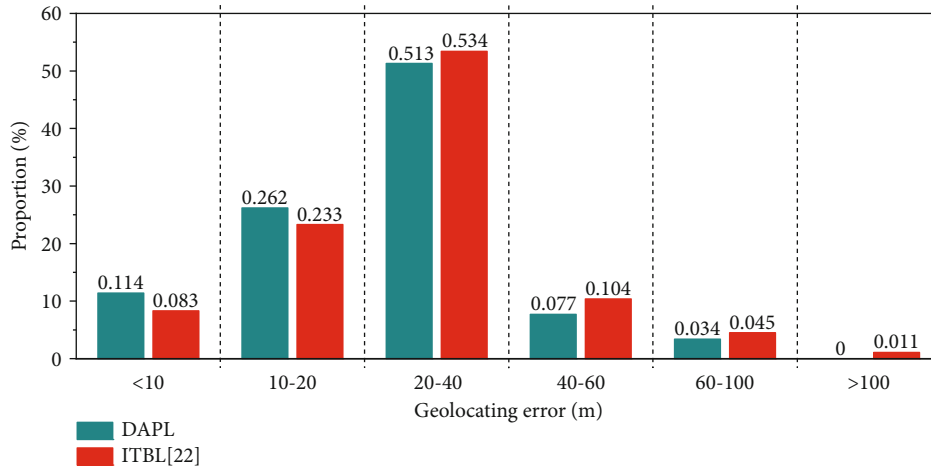FIGURE 8: Maximum included angle and distance distribution of initial probes.

FIGURE 9: Distribution of geolocating errors.

target and the initial probe has limited influence on the geolocation success rate.

In order to verify the factors affecting the geolocation success rate discussed in 4.1, we make statistics on the maximum included angle and distance distribution of the initial probe combinations with geolocating failure of ITBL, and the results are shown in Figure 8. Figure 8(a) shows the distribution of the maximum included angle between the initial probes in the case of geolocating failure. As we discussed in Section 4, the maximum included angle between the initial probes ranges from 60° to 180°. It can be seen from the figure that when the maximum included angle is greater than 120°, the proportion of failed geolocation increases significantly. Figure 8(b) shows the distribution of the relationship between the maximum interprobe distance and the maximum target's reported distance (d in the figure) obtained by the initial probes. When the distance between the probes is too large or too small ($<0.25 * d$ or $> d$), the proportion of

failure cases is higher. Experimental result verifies that the analysis of the factors affecting the geolocation success rate in Section 4.1 is reasonable to a certain degree.

5.2.2. Geolocation Accuracy. For all successfully geolocated targets, the minimum error, the average error, and the median error of DAPL are 0.19 m, 22.1 m, and 21.7 m, respectively, which are 0.45 m, 23.8 m, and 20.6 m, respectively, for the ITBL. Figure 9 shows the distribution of geolocating errors. From the figure, the proposed DAPL method has similar geolocating accuracy performance compared with ITBL, both of which can geolocate about 80% of the targets within 40 m, which is enough to determine the target's location in a very small range.

When the threshold is smaller, the two methods both can get higher accuracy. However, a smaller threshold will increase the time cost. Our work focuses on how to geolocate the target in long-distance conditions, rather than achieving
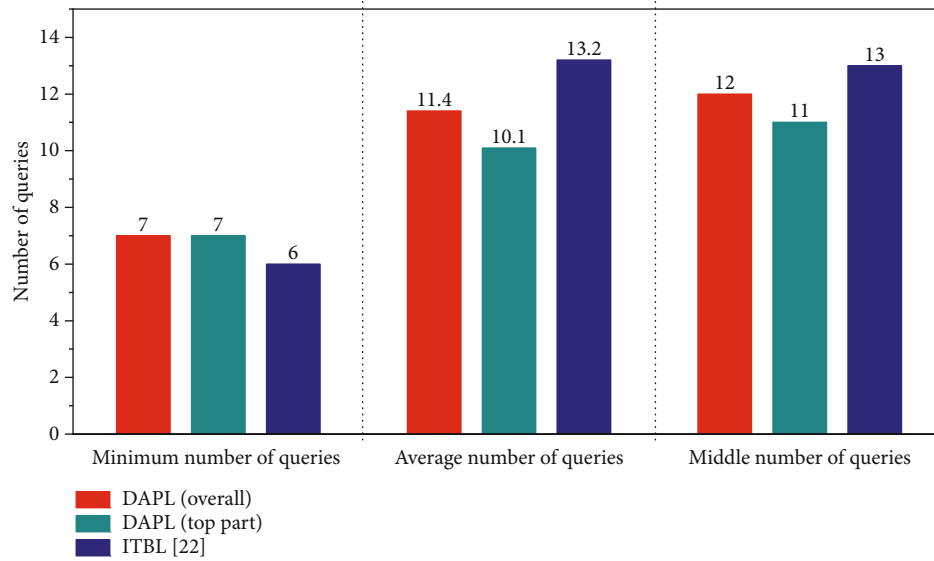
Figure 10: The comparison of service query times.

higher precision regardless of cost. Therefore, we believe that the two methods are commensurate in geolocating accuracy.

5.2.3. Service Query Times. Service query times are the number of times LBSD service is used in the geolocating process, which reflects the time consumption performance of the method. To ensure the fairness of the experiment comparison, we do not consider cases of failed geolocation and only compare the consumption of query times in the case of successful geolocation. The results are shown in Figure 10. The average number of queries of DAPL is about 11.4 times, which is lower than 13.2 times of ITBL. The fastest geolocation requires only 7 queries of service of DAPL. Because the number of successfully geolocated targets by DAPL is more, if only 3529 targets with the best performance in DAPL are considered, as represented by the top part in the figure, our method has better performance.

Although DAPL will add new probes in each round of trilateration which will need more queries to the server, the more reasonable combination of probe positions of DAPL will help to approach the real position of the target faster. Therefore, it has better efficiency performance in general.

In summary, the proposed method can achieve reliable and accurate geolocation of Momo target users under long-distance conditions. Compared with the existing method based on iterative trilateration, it can effectively improve the success rate, reduce the time consumption, and have similar accuracy.

6. Conclusion

In this manuscript, an efficient localization method for LBSD users at long distances based on dynamic adjustment of probes (DAPL) is proposed. DAPL achieves continuous approximation to the target's location by reestablishing the coordinate system and adjusting probes' location dynamically. In this process, the distance oscillation phenomenon

is avoided based on coordinate inversion of the probe to ensure the efficiency. The actual experiment based on Momo shows that the proposed method achieves nearly 100% geolocating success rate for the target without knowing the potential area information of the target. At the same time, compared with the existing representative method, it has less time overhead and higher success rate on the premise of ensuring the positioning accuracy. The motivation of our work is to verify whether the malicious social network users can be geolocated based on the public service provided by the social platforms. The actual experiment shows that the current location protection mechanism adopted by service providers cannot prevent the user's real location from being inferred. On the premise of ensuring the service quality, spatial confusion and other privacy enhancing mechanisms need to be used to increase the random confusion of reported distance [9, 35, 36]. In the future work, we will put more effort into the probability of distance oscillation caused by different location confusion mechanisms and the corresponding countermeasures.

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The Social Internet of Things (SIoT) - when social networks meet the Internet of Things: concept, architecture and network characterization," *Computer Networks*, vol. 56, no. 16, pp. 3594–3608, 2012.

[2] L. Atzori, A. Iera, and G. Morabito, "From "smart objects" to "social objects": The next evolutionary step of the internet of things," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 97–105, 2014.

[3] "WeChat hardware platform," http://iot.weixin.qq.com/.

[4] M. Michael, "Social media platforms and the cybercrime economy," 2019, https://www.bromium.com/wp-content/ uploads/2019/02/Bromium-Web-of-Profit-Social-Platforms-Report. pdf.

[5] L. Zhang, J. Wang, W. He, P. Zhang, and S. Zhang, "WeChat rumor analysis and governance based on big data," in *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, pp. 1–4, New York, 2019.

[6] R. R. Darby, A. Horn, F. Cushman, and M. D. Fox, "Lesion network localization of criminal behavior," *Proceedings of the National Academy of Sciences*, vol. 115, no. 3, pp. 601–606, 2018.

[7] J. Hua, W. Tong, F. Xu, and S. Zhong, "A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1155–1168, 2018.

[8] D. Jurgens, T. Finethy, J. McCorriston, Y. Xu, and D. Ruths, "Geolocation prediction in Twitter using social networks: a critical analysis and review of current practice," in *Proceedings of the 9th International Conference on Web and Social Media*, pp. 188–197, Oxford, UK, 2015.

[9] L. Kong, Z. Liu, and Y. Huang, "SPOT," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1681–1684, 2014.

[10] Y. Yao, X. Zhang, H. Wu, Z. Wang, and J. Wang, "A novel location privacy protection algorithm for social discovery application," *IETE Technical Review*, vol. 38, no. 1, pp. 82–92, 2021.

[11] H. Wang, Y. Li, Y. Chen, and D. Jin, "Co-location social networks: linking the physical world and cyberspace," *IEEE Transactions on Mobile Computing*, vol. 18, no. 5, pp. 1028–1041, 2019.

[12] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1082–1090, New York, 2011.

[13] K. P. N. Puttaswamy, S. Wang, T. Steinbauer et al., "Preserving location privacy in geosocial applications," *IEEE Transactions on Mobile Computing*, vol. 13, no. 1, pp. 159–173, 2014.

[14] Z. Yang and Y. Liu, "Quality of trilateration: confidence-based iterative localization," *IEEE Transactions on Parallel & Distributed Systems*, vol. 21, no. 5, pp. 631–640, 2010.

[15] M. S. Huang and R. M. Narayanan, "Trilateration-based localization algorithm using the lemoine point formulation," *IETE Journal of Research*, vol. 60, no. 1, pp. 60–73, 2014.

[16] Y. Ding, S. T. Peddinti, and K. W. Ross, "Stalking Beijing from Timbuktu: a generic measurement approach for exploiting location-based social discovery," in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pp. 75–80, New York, 2014.

[17] N. P. Hoang, Y. Asano, and M. Yoshikawa, "Your neighbors are my spies: location and other privacy concerns in dating apps," in *Proceedings of the 18th International Conference on Advanced Communication Technology*, pp. 715–721, PyeongChang, Korea (South), 2016.

[18] W. Shi, X. Luo, J. Guo, C. Liu, and F. Liu, "Where are WeChat users: a geolocation method based on user missequence state analysis," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 2, pp. 319–331, 2021.

[19] M. Xue, Y. Liu, K. W. Ross, and H. Qian, "I know where you are: thwarting privacy protection in location-based social discovery services," in *Proceedings of the IEEE Conference on Computer Communications Workshops*, pp. 179–184, Hong Kong, China, 2015.

[20] H. Cheng, S. Mao, M. Xue, and X. Hei, "On the impact of location errors on localization attacks in location-based social network services," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, Springer, Cham, 2016.

[21] J. Peng, Y. Meng, M. Xue, X. Hei, and K. W. Ross, "Attacks and defenses in location-based social networks: a heuristic number theory approach," in *Proceedings of the 36th International Symposium on Security and Privacy in Social Networks and Big Data*, pp. 64–71, Hangzhou, China, 2015.

[22] M. Li, H. Zhu, Z. Gao et al., "All your location are belong to us: breaking mobile social networks for automated user location tracking," in *Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 43–52, New York, 2014.

[23] W. Shi, X. Luo, F. Zhao, Z. Peng, Q. Cheng, and Y. Gan, "Geolocating a WeChat user based on the relation between reported and actual distance," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, 2018.

[24] J. Wang, H. Cheng, M. Xue, and X. Hei, "Revisiting localization attacks in mobile app people-nearby services," in *Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, Springer, Cham, 2017.

[25] T. L. Lin, H. Y. Chang, and S. L. Li, "A location privacy attack based on the location sharing mechanism with erroneous distance in geosocial networks," *Sensors*, vol. 20, no. 3, pp. 916–918, 2020.

[26] G. Wang, B. Wang, T. Wang, A. Nika, H. Zheng, and B. Y. Zhao, "Whispers in the dark: analysis of an anonymous social network," in *Proceedings of the 14th ACM Conference on Internet Measurement Conference*, pp. 137–150, New York, 2014.

[27] I. Polakis, G. Argyros, T. Petsios, S. Sivakorn, and A. D. Keromytis, "Where's Wally?: precise user discovery attacks in location proximity services," in *Proceedings of 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 817–828, 2015.

[28] X. Luo, Y. Qiao, C. Li, J. Ma, and Y. Liu, "An overview of microblog user geolocation methods," *Information Processing & Management*, vol. 57, no. 6, pp. 102318–102375, 2020.

[29] P. A. Zandbergen, "Accuracy of iPhone locations: a comparison of assisted GPS, WiFi and cellular positioning," *Transactions in GIS*, vol. 13, no. 1, pp. 5–25, 2009.

[30] J. Liu, Y. Zhang, and F. Zhao, "Robust distributed node localization with error management," in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 250–261, New York, 2006.

[31] H. C. So and L. Lin, "Linear least squares approach for accurate received signal strength based source localization," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 4035–4040, 2011.

[32] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, 2008.

[33] Y. Shang, W. Rumi, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 11, pp. 961–974, 2004.

[34] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, Reading, MA: Addison-Wesley, 1984.

[35] M. Siddula, Y. Li, X. Cheng, Z. Tian, and Z. Cai, "Privacy-enhancing preferential LBS query for mobile social network users," *Wireless Communications and Mobile Computing*, vol. 2020, 13 pages, 2020.

[36] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services: a comprehensive survey," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–36, 2021.