Research Article

A Strategy of Cluster-Based Distributed Location Service

Peng Wang,^{1,2} Jing Yang¹ , and Jianpei Zhang¹

¹College of Computer Science and Technology, Harbin Engineering University, Harbin 150000, China

²College of Information Engineering, Suihua University, Suihua 152000, China

Correspondence should be addressed to Jing Yang; yangjing@hrbeu.edu.cn

Received 13 September 2018; Accepted 7 April 2019; Published 2 May 2019

Academic Editor: Claudio Agostino Ardagna

Copyright © 2019 Peng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new strategy of cluster-based distributed location service was proposed to solve communication bottlenecks and vulnerability in centralized system structure location service. In the proposed strategy, the density-first clustering method was used to establish location information profiles for users, and neighboring user groups satisfied the (k, d) anonymous model. When there is a location service request, a number of neighbor user groups that meet the density metric condition were determined firstly, then the location profile similarity of a service requester and the neighbor user was calculated, and finally the collaborative filtering recommendation method was used to recommend the corresponding location service. The experimental results on the location dataset show that the proposed strategy can provide users with a sufficient number of location services and the (k, d) privacy security definition can guarantee the privacy of user location information.

1. Introduction

With the rapid development of mobile Internet technology, various types of mobile terminal apps are integrated into people's daily lives to provide fast, simple, and satisfactory services, for example, real-time navigation, route planning, and inquiring about nearby food, attractions, and convenience stores [1]. The working process of the existing LBS application is as follows: (1) the mobile terminal device sends the location information and the location service request to the server and (2) the server recommends the location services candidate set according to the location information of the user. Most of the existing location service architectures adopt a centralized location server model, which sorts location services according to various evaluation criteria, such as volume, score, and distance and then recommends the sorted location service candidate sets to the users for secondary screening. However, the centralized location service model has some shortcomings. First, the location service results are closely related to the sorting of the priority condition, resulting in the users with same screened conditions will get the same location service recommendation results. Secondly, the centralized location

service model faces huge privacy security threats. For example, some apps collect and publish user location information and custom preferences, resulting in leakage of user privacy information. Thirdly, the centralized location service system structure has the disadvantage of high load on the central node and vulnerability to malicious attacks. Therefore, it becomes a research hotspot that how to overcome the shortcomings of the centralized location service mode, provide users with personalized, high-quality location services, and ensure the privacy of user information in the field of location services.

At present, most of the location services applications recommend the hotspot location service by only considering the location information of the user. The method is simple, and the location service results are popular and singular, which cannot meet the individual needs of the users. The main reason for this phenomenon is that the existing location service recommendation mechanism ignores the user's own attribute preferences and does not consider the potential personal differences of the users. In order to take into account the user's own attribute and personalized preferences in the location service recommendation mechanism, this paper established the user

location information profile for each user and introduced a scoring mechanism to describe the user's personalized preferences. Collaborative filtering algorithms are widely used in e-commerce and other fields. This algorithm was used to recommend suitable products according to the user's individual needs [2]. And a collaborative filtering algorithm was designed for location service recommendation based on the user location information profile model. Moreover, while providing high-quality location services, how to ensure the privacy of users' location information has always been a research hotspot in the field of location services. Grutester and Grunwald firstly proposed the k -anonymous location service privacy protection method [3], and there are many privacy protection methods for location services, such as differential privacy protection methods, perturbation, false location, and generalization [4, 5]. Based on the k -anonymous privacy protection method, the (k, d) privacy protection model is designed for providing high-quality and personalized location service and ensuring the user's location privacy security in this paper. Besides, the centralized location service system structure has the disadvantages of heavy central node load and low privacy security. Therefore, the distributed location service system structure is gradually applied in the field of location services. The literature [6] proposed a k -anonymous location privacy protection method under the peer-to-peer structure using neighboring user location information, but this paper ignored the privacy security problem between user nodes. Shokri proposed a user LBS sharing mechanism (MobiCrowd) to reduce the chances of users' own location being exposed to the server but faced the cold leak and the privacy leakage of the initial LBS request [7]. Therefore, this paper designed a new location service recommendation strategy for the existing location service system. Firstly, the location information profile with user's preference was considered to ensure the quality and personalization requirements of location service. Secondly, the neighboring user group that satisfies the (k, d) privacy model was constructed to ensure the information privacy. Finally, based on the distributed structure, a collaborative filtering recommendation algorithm was designed to recommend location services, which avoids the problem of heavy load on key nodes in centralized location servers.

Our main contributions can be summarized as follows:

- (1) The user location information profile model was proposed, and a scoring mechanism was designed. The process considered the density of the user's location, the user's social attributes, and user preferences and satisfied the personalized needs of the user's location service.
- (2) The (k, d) privacy security model was designed and used to determine the neighboring user group to ensure the information privacy.
- (3) The location service collaborative filtering recommendation method based on the distributed system structure was designed to avoid the computational load and privacy security threats existing in the centralized location server.

The organization structure of the article is as follows: Section 2 describes the related work; Section 3 describes the main content of the recommended location service strategy; Section 4 verifies the feasibility of the proposed strategy by experiment and compares the performance of the algorithm; and Section 5 gives the conclusion and the research outlook.

2. Related Work

Most of the existing location service application models adopt a centralized anonymous server architecture, which faces serious privacy security threats, such as server attacks, data leakage, or link attacks against the user location information, background knowledge attacks, and homogeneous attacks. Therefore, how to provide high-quality location services while ensuring the privacy of users has become a research hotspot of many scholars.

In order to provide the location services, some literatures combined community computing with location services, used clustering algorithms to identify users with common interests, and collected user-related data as much as possible to estimate user preferences. At the same time, the application released as much information as possible to help users choose location services, such as the scores of users and drivers in the taxi software [8]. Although collecting and publishing more user information can improve the quality of location services, it also further increases the risk of user privacy information disclosure, such as more recent malicious crimes against car drivers and taxi users. Therefore, if the centralized storage and transmission of user information can be reduced in the location service, the probability and scale of leakage of user privacy information can be reduced. The literature [6] proposed the location service application under the distributed system structure and used the k -anonymity method to protect the user's location information security based on the peer-to-peer architecture; the literature [7] used the location service result caching mechanism to recommend the location based on the distributed structure; the literature [9, 10] used the neighbor location information to recommend the location service based on the distributed architecture. In addition, there are some location privacy protection methods for distributed location service architecture [11–13]. For example, the literature [11] used mobile client cache location data to provide location services based on distributed systems to avoid exposing location information to the server; the literature [12] used the combination of caching and user collaboration to protect location privacy; and the literature [13] used the P2P structure to protect location privacy and reduced the possibility of location privacy leakage through the capabilities of mobile devices and collaborative recommendation methods. The main contribution of these research results is how to reduce the risk of user privacy leakage under distributed systems.

There are also many research results on the privacy protection methods for location data. Among them, the k -anonymous data protection method was first applied in the field of data publishing. Its main idea is to hide the user's real

data in a collection containing K data [14]. Now the research results based on the k -anonymous privacy protection method are widely applied to a variety of data and application areas. The research results based on the k -anonymity model are also rich. For example, Machanavajjhala et al. proposed the l -diversity model [15] and Liu et al. proposed a (α, k) -anonymity model, which is required to satisfy k anonymity and ensure its diversity α [16]. Based on the idea of k -anonymity and l -diversity models, this paper proposed a (k, d) privacy model for location service applications and established a neighbor user group based on clustering. The neighbor user group satisfied (k, d) privacy security requirements. That is, at least k users and d location service categories were included in the neighbor user group, and there is no requirement for the number of location services. The model is resistant to both link attacks and background knowledge attacks and homogeneous attacks, while providing a sufficient number of location services. Clustering algorithms are widely used in the research of location data privacy protection, especially the K-top and K-means clustering methods [17, 18]. In the literature [19], the density-based clustering method is proposed for the first time. In the research background of our work, the user location point density in the region is used to measure the activity of the user in the region. A higher density of location points indicates that the user is more active in the area, and the higher the probability of recommending a high-quality location service. Therefore, this paper introduced a density-based measurement method in the construction of user location information profiles and the establishment of neighboring user groups.

The collaborative filtering recommendation algorithm is widely used in the fields of network services and e-commerce. The algorithm uses the similarity of information between users to recommend services that users may like. The more detailed the information collected, the more accurate results would be obtained. But the algorithm is still not mature in location service field, and it also faces the risk of data information privacy leakage. At present, the privacy protection methods for collaborative filtering recommendation algorithms mainly include random interference methods and group anonymization methods [20–22]. The literature [9] considered location similarity to recommend location services based on distributed systems, but it is difficult to get higher quality location services. The literature [10] considered the similarity of user social attributes and used the collaborative recommendation algorithm to recommend location services, but it ignored the privacy protection among users. Therefore, this paper designed a new distributed location service recommendation strategy based on the clustering method. In this strategy, users used the density metric-based clustering method to establish their own location information profiles. And then the neighboring user groups were established according to geographic regions so that neighboring user groups can satisfy (k, d) privacy security definitions and utilize collaborative filtering. The collaborative filtering recommendation method was used to recommend the corresponding location service.

Compared with the existing scheme, the proposed scheme used the density clustering method to construct the user location information profile so that the user's own location information is clustered as much as possible; the neighboring user groups were established according to clustering method so that the groups satisfy the (k, d) privacy protection model. The proposed scheme not only ensured the user's location information privacy security but also provided a sufficient number of user location service information. When there was a location service request, the user location area and the neighboring user group location profile similarity were calculated, and the collaborative location recommendation method was used to recommend the corresponding location service. Compared with literatures [9, 10], the proposed scheme can provide users with a sufficient number of high-quality location services while, at the same time, providing more stringent security of user location information privacy.

3. Collaborative Filtering Recommended Location Service Method

3.1. Basic Description. In view of the overloaded and privacy leakage problems of centralized anonymous servers, some scholars have proposed LBS systems based on distributed system structure in recent years. In the literature [6], Chow et al. proposed the k -anonymity algorithm under peer to peer, which used the user information in the ad hoc network and its own cache information to construct the k -anonymous model. This method still needs to access the location server and does not consider the location information security issues among nodes. Shokri proposed the LBS sharing mechanism to obtain the LBS in the literature [7], but it faces the cold start problem, and the initial stage still relies on the centralized anonymous server. The literature [9] recommended location service using the neighboring user group based on the distributed system structure without accessing the location server. But when selecting neighboring users, the overlapped user profiles in the geographical location were considered, the clustering property of the user location information was neglected resulting in the low quality of the recommended location service results. The literature [10] considered the clustering of user location information, adopted collaborative filtering recommendation method to recommend location service, and improved location service quality. However, the single location information privacy security protection mechanism and homomorphic encryption method still caused the risk of location information leakage. Therefore, this paper proposed a new cluster-based distributed location service recommendation strategy. In the strategy based on the distributed system structure, user used the density-prioritized clustering method to establish their own location information profiles and then established a neighboring user group that satisfies the (k, d) privacy security definition. When there was a location service request, a number of neighboring user groups that meet the conditions were determined firstly, then the similarity between the service requesting user and the neighboring user was calculated, and finally the collaborative filtering

recommendation method was used to recommend the corresponding location service.

The higher the density of the location points in the user activity area illustrates the user is more familiar to the area and recommend location service with higher quality. Therefore, the density prioritization clustering algorithm was used to determine the location profile with higher density and the user location evaluation mechanism was designed. In addition, when the clustering algorithm was used to determine the neighboring user group, the type and quantity of the location service information included in the user in the group were increased as much as possible so as to ensure the success rate of the recommended location service. At the same time, combined with the user location information privacy security problem, the neighbor user group (k, d) privacy protection model was designed. And the designed neighbor user group contained at least k users and no less than d location service types. The neighboring user group that satisfied the (k, d) privacy protection model had the structural characteristics of overall decentralized and local clustering and could reduce the risk of users being attacked in the user group.

3.2. Profile Construction. Based on the position profile model in [9], this paper designed a new position information profile construction method. Each user's position information profile contains the position information of the node itself, category label, and score composition. When constructing the location profile, the density measure of the location point was prioritized and the score for the location point was generated according to the location point dwell time and the access frequency. Finally, a position information profile capable of representing the user's active area is formed. The user location information profile model description is shown in the following equation:

$$\begin{cases} L = \{l_1, l_2, l_3, l_4, \dots, l_n\}, \\ l_i = \{(x_i, y_i), t_i, c_i, s_i\}, \\ D = \sum_{k=1}^i \frac{1 + s_i}{L_{\text{area}}}, \end{cases} \quad (1)$$

where L represents the user's location information profile, including n location information, the i -th location information represents the location coordinate l_i , t_i is the timestamp at the location (x_i, y_i) , c_i is the category label of the location (x_i, y_i) , s_i is the score of the location, D is the absolute density of the location information of the user, L_{area} is the area of the location profile, and the value of D is proportional to the familiarity of the user in the area. At the same time, s_i represents the score of (x_i, y_i) , and the scoring function is designed as follows:

$$\begin{cases} \gamma = \alpha \Delta_t + \beta z, & (\alpha > 0, \beta > 0, \alpha + \beta = 1), \\ s(\gamma) = \frac{\gamma}{a + b\gamma}, & (a > 0, b > 0), \end{cases} \quad (2)$$

where Δ_t represents the dwell time of the user at the location point, z represents the number of visits by the user to the location point, and α and β , respectively, represent the weighting ratio coefficient of the dwell time and the number of visits in the scoring result. $s(\gamma)$ represents the score of the location point, the score is monotonically increasing on $[0, +\infty)$, and the range is $[0, 1/b)$, so the b value is used to control the upper limit of the score. And the a value is used to control the rate of score growth. In this paper, $a = 0.8$ and $b = 0.2$, that is, the position score takes the range $[0, 5)$.

In order to ensure the availability of the user location information and improve the clustering of the location points inside the location profile, the density clustering algorithm [19] was adopted to further preprocess for the user location information profile. In the user location information, the location points with higher privacy requirements were blurred, such as family and work unit. Then, the divergence points with lower scores were deleted because these divergence points would greatly reduce the absolute density D of the location information profile resulting in affecting the clustering of the location information profile. Therefore, the preprocessing mainly includes identifying and deleting the divergence point, determining the centroid position of the user position information profile, and calculating the local density parameter ρ_i and the neighbor distance parameter δ_i for each location point in the profile. The formula of the local density parameter ρ_i is as follows:

$$\rho_i = \sum_j \chi(d_{ij} - d_c), \quad (3)$$

where $\chi(x) = 1, x < 0$ and $\chi(x) = 0, x > 0$, d_c is the distance threshold, and the local density ρ_i indicates that the distance from the node i is less than the number of d_c nodes. In addition, the formula of the neighbor distance parameter δ_i is as follows:

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}). \quad (4)$$

Besides, the neighbor distance parameter δ_i for the special node with the highest local density is calculated as follows:

$$\delta_i = \max_j (d_{ij}). \quad (5)$$

Those location points with a relatively large local density ρ_i and a large δ_i were considered to be centroid positions of the information profile, and the point with small local density ρ_i and large δ_i was an outlier.

3.3. Construction of the Neighboring User Group. The k -anonymity technology was first applied to the field of data publishing, which could effectively prevent link attacks but does not prevent background knowledge attacks and homogeneous attacks. However, the principle of l -diversity can overcome the shortcomings of k -anonymity technology. Based on the existing k -anonymity and l -diversity privacy

protection methods, this paper proposed the neighboring user group that satisfied the definition of (k, d) for the distributed location service application problem, and then the neighboring user group was used to recommend the location service for the user. The relevant definitions were as follows.

Definition 1 (k -anonymity). $G(L_1, L_2, L_3, \dots, L_{n-1}, L_n)$ is set as a group of neighboring users, where L_i is the positional profile of the neighboring user, expressed as $\{(x_1, y_1, t_1, c_1, s_1), (x_2, y_2, t_2, c_2, s_2), \dots, (x_m, y_m, t_m, c_m, s_m)\}$ and (x_i, y_i) is the position coordinate identifier. For any (x_i, y_i) , G is divided into clusters $\{g_1, g_2, \dots, g_p\}$, $|g_i| \geq k, 1 \leq i \leq p$; the elements in g_i have the same (x_i, y_i) value, and then the neighbor user group G satisfies k -anonymity.

Definition 2 (d -diversity). $G(L_1, L_2, L_3, \dots, L_{n-1}, L_n)$ is set as a group of neighboring users, where L_i is the location profile of the neighboring user, expressed as $\{(x_1, y_1, t_1, c_1, s_1), (x_2, y_2, t_2, c_2, s_2), \dots, (x_m, y_m, t_m, c_m, s_m)\}$, c_i is the category label of the location service candidate, and $C_{\text{statistic}_G}(c_i)$ is the classified statistic of the location service category in the neighbor user group. Then, if $C_{\text{statistic}_G}(c_i) \geq d$, the neighboring user group G satisfies d -diversity.

If the neighboring user group G satisfies both Definitions 1 and 2, the G satisfies the (k, d) privacy protection definition. The k -anonymity requires that each coordinate identifier of the neighboring user group G corresponds to at least k different neighboring users so that it is difficult to identify the location sensitivity information, and the attacker cannot have a probability of less than $1/k$ to derive individual privacy information even if the target exists in the user group. The d -diversity requires that there are at least d numbers of location information categories of users in the neighboring user group G , and there is no requirement for the number of location service candidates in the group. The (k, d) neighbor user group model was proposed combining the characteristics of k -anonymity and d -diversity, and it has the characteristics of k -anonymity against connection attacks and the characteristics of resisting background knowledge attacks and homogeneous attacks.

The steps for constructing (k, d) neighboring user groups were as follows:

- (1) The user with the highest absolute density was selected as the central node of the neighboring user group in the area and set the centroid position of the user profile to be the centroid position of the neighboring user group.
- (2) Calculate the relative density of neighboring users and sort by relative density.
- (3) Gradually add neighboring users to the neighboring user group and detect if the (k, d) definition is met.
- (4) If the (k, d) requirement is met, finally recalculate the centroid position of the neighboring user group.

- (5) Delete the generated neighbor user group from the user collection.
- (6) Re-execute steps 1–5 for the remaining user collection.

The relative density measurement formula is as follows:

$$\mu = \sum_i \chi(d_{ij} - d_c), \quad i = \{1, 2, 3, \dots, k\}, \quad (6)$$

where $\chi(x) = 1, x < 0$ and $\chi(x) = 0, x > 0$, d_c is the position distance measure, and d_{ij} is the distance from the position point l_i to the centroid position point l_j . The relative density μ corresponds to the number of position points in the location profile where the distance to the centroid position point l_j is less than d_c . The greater the relative density μ illustrates the denser the location profile of the neighboring user relative to the centroid position l_j , and the higher the quality of the location service would be provided.

In addition, when transmitting location service information among neighboring users, it also faces the risk of location information leakage. Therefore, this paper used the Paillier encryption scheme to encrypt location information [23].

3.4. Recommendation of Location Service Using Collaborative Filtering Method. In the location service application scenario designed in this paper, if the user A needs to request the LBS at the location coordinate l_a , the user A requests information from the broadcast location service; then, the neighboring user group that meets the relative density requirement were filtered, the location profile similarity between user A and group members was calculated, and the similarity and recommended location services were sent to user A. Finally, the user A calculates the rating of the recommended location service using the similarity for the received location service candidate set and selects the location service with higher rating.

3.4.1. Relative Density Calculation of Neighbor User Groups. The neighboring user group G calculated the relative density μ of the location information in the group relative to the l_a of the user A according to equation (6). If the density μ is greater than the preset density measure threshold, the location service request is responded; otherwise, the service request information is not responded.

3.4.2. User Location Profile Similarity Calculation. The similarity of the position contour of the neighboring user and the requesting user is calculated when receiving the location profile of the requester. The higher the similarity causes the higher quality of recommended location service. The similarity between users is calculated according to Euclidean similarity measure. Let there be a total of k identical position points in the location profiles, where in the scores of the i -th position point are x_i and y_i , respectively, and the Euclidean distance of the two location profiles is as shown in the following equation:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}. \quad (7)$$

The similarity of the location profiles is as shown in the following equation:

$$\text{sim}(x, y) = \frac{1}{1 + d(x, y)}. \quad (8)$$

Finally, the recommended location service and the corresponding location profile similarity are sent to the user A, and then a location service candidate set including the location service and the similarity is generated.

3.4.3. Location Service Result Score Calculation. The location service candidate set contains a total of n location service results. For any position point l_i , there are m recommended users and the set of attributes (s_j, sim_j) , where $j = 1, 2, 3, \dots, m$, s_j is the score of the A_j user to l_i and sim_j is the similarity of the A_j user and service requester B. For the service requester user B, the location service prediction score calculation of the l_i location point is as follows:

$$s_i = \frac{\sum s_j \text{sim}_j}{\sum \text{sim}_j}, \quad i = 1, 2, 3, \dots, n, j = 1, 2, 3, \dots, m. \quad (9)$$

Finally, the location service with the highest predicted score in the service candidate set was selected as the final location recommendation result.

4. Experiment and Result Analysis

In order to verify the feasibility and effectiveness of the proposed strategy, the CRAWDAD data set was used for simulation and analysis. The data set with location data for 536 taxis within one month [24, 25] is widely used in location data analysis and research. In this section, the characteristics of the three stages of the profile generation algorithm including the location service request, response process, and collaborative filtering recommendation location service would be verified and analyzed based on the CRAWDAD data set. The evaluated indicators mainly include the location profile feature, the number of users in the neighboring user group, the number of service candidate sets, and the probability of recommendation service success. At the same time, the experimental analysis was carried out from the aspects of the system architecture, algorithm efficiency, and communication cost. Besides, the proposed strategy (CDCLS) was compared with the algorithms proposed in [7] (MobiCrowd) and [9] (DCRLS).

4.1. Verification of Position Information Profile Generation Algorithm. In the strategy proposed in this paper, the location information of each node was collected to construct a location information profile. At the same time, location points were rated and location service candidate sets were generated. At last, data preprocessing was performed on the location information profile based on the density clustering method and the divergence point, and the lower score point were deleted. Because the user location service candidate set

had an important impact on the quality of the LBS recommendation result, the relevant features of the user profile were fully verified and analyzed on the real data set. In this process, the parameters of the position nodes, the density measure, and other parameters are analyzed in detail, and the algorithm was compared with DCRLS.

In the experiment, the profile generation algorithm was used to process the position information of 536 taxi objects and the corresponding location information profile was generated, and then the position information profile was preprocessed. The local density parameter and the neighbor distance parameter for each dwell point position were calculated, and the two parameters were used to determine the centroid position and delete the divergence point. Subsequently, the final position service candidate set would be determined. For the convenience of comparison in analysis, the six taxis with the same literature as [9] were selected for case analysis, and finally the data of all objects were obtained.

The 6 taxi objects recorded as A~F were selected randomly, and their original location information plan is shown in Figure 1.

The location profile generation algorithm in the proposed strategy was used to process the moving objects A~F original position information set. The main content included calculating the local density and the neighbor distance value of the position point, thereby deleting the outlier point and determining the location information profile. According to the dwell time and the number of visits, the score and the corresponding location information profile were generated as shown in Figure 2.

The location profile construction method designed in this paper deleted the outlier point based on the density measure and performed effective data preprocessing on the location information profile. Comparing Figure 1 with Figure 2, it can be clearly seen that the processed location information profile had better aggregation and higher density. In addition, the location service candidate set was scored by using the scoring function designed in this paper. And the parameters $\alpha = 0.5$, $\beta = 0.5$, $a = 0.8$, $b = 0.2$, $d_c = 100$; the score range was $[0, 5]$; and location service score statistics of the last six selected mobile objects are shown in Figure 3.

Besides, the location information of the same moving object was compared with the data of the DCRLS, and the comparison is shown in Figure 4.

Compared with the DCRLS algorithm, less location information was used to construct location information profiles in CDCLS and absolute density was higher. From the comparison chart, it shows that CDCLS had better clustering and higher data quality. Therefore, in the experimental environment, it was feasible to recommend the LBS service for users by using the location information profile generation algorithm proposed in this paper, which avoided the risk of privacy leakage.

Finally, the location information profile generation algorithm was used to process the profile information of all moving objects in the data set. The data set contained 536 moving objects, a total of 11,219,955 position points, and an average of 20,930 position points per moving object. In the generated profile, there were a total of 312,827 position

FIGURE 1: The original locations of objects. Scatterplot of (a) object A, (b) object B, (c) object C, (d) object D, (e) object E, and (f) object F.

points, an average of 601 position points per moving object, and the average point position ratio is 2.8%. According to the neighboring user group construction algorithm proposed in this paper, all moving objects in the data set are processed. And $k=10$ and $d=30$ were set, and 53 neighboring user groups were generated. The proposed location information profile algorithm could discard the outliers with shorter moving time and fewer access times so that the location data in the generated location information profile had higher quality utilization. In addition, the generated neighboring user group satisfied the (k, d) privacy security

definition to ensure the user's location information privacy security and high data availability.

4.2. Analysis of Privacy Protection Efficiency. In order to further analyze the privacy protection efficiency of the (k, d) privacy protection model, different k and d values were set for the neighboring user group on the CRAWDAD data set, and the generated neighbor user group satisfied the (k, d) privacy security definition. There were 536 moving objects in the data set, and each neighboring user group contained at

FIGURE 2: The location profile of objects in CDCLS. Scatterplot of (a) object A', (b) object B', (c) object C', (d) object D', (e) object E', and (f) object F'.

least k moving objects when generating the neighboring user group. The moving objects in the group were selected according to the density clustering metric, and the group contained at least d location service categories. The change trend of privacy leakage risk of mobile objects and location data with different k and d values is shown in Figure 5.

As shown in Figure 5, it could be seen that the leakage risk of mobile object location data decreased with the increase of the k value. It was clearly shown that the leakage risk tended to be flat while the values of k were set over 10. Thus, the value of k was set 10 in the experimental process of this paper. In addition, when k was set the same value, the larger value of d caused less privacy leakage risk and the risk of privacy leakage does not change much while the value of d over 30.

Thus, the value of d was set 30 in recommended location service. When constructing a neighbor user group with different (k, d) parameter values, the changes of execution time of the algorithm are shown in Figure 6.

With comprehensive consideration of the effect of different k and d values on the degree of privacy protection and algorithm execution time, k was set as 10 and d was set as 30 on the data set to generate optimal location information profile and constructed the neighbor user group. According to the proposed neighbor user group construction algorithm, all moving objects in the data set were processed and 53 neighbor user groups were generated correspondingly. In our work, by evaluating and analyzing the effect of different k and d values on the privacy risk of the privacy protection

FIGURE 3: The location service score statistics of objects in CDCLS.

FIGURE 4: The comparison of location profile: (a) position number, (b) absolute density, and (c) position number ratio.

FIGURE 5: Location data leakage risk trend chart.

model and the execution time of the algorithm, the following conclusions could be drawn: adopting the (k, d) privacy protection model in the distributed location service recommendation strategy could increase the execution time and the larger value of k and d would cause the longer

FIGURE 6: Comparison of execution time under different (k, d) values.

execution time and lower the privacy leakage risk. Therefore, the balance between privacy protection and time cost could be obtained by appropriately setting the values of k and d .

4.3. Verification of Location Service Request and Corresponding Process. The location service request and response process were verified and performance was analyzed, and the parameter settings were the same as in the previous section. The 100 user location points in the data set were selected randomly as the LBS requester in analysis experiment. The theme of the service request content was simplified to “Where do I go next time?” It should be noted that in practical applications, service topics and related parameters could be flexibly set according to specific actual conditions.

The LBS request and response process was as follows: (a) when the user requests an LBS, the distance truncation parameter d_c was set. (b) The neighboring user groups that meet density requirements were filtered in the data set. (c) The location service was recommended by the neighboring user group according to the collaborative filtering method. d_c was set as 100, 200, and 300 meters for 3 sets of experiments. In each experiment, the number of response neighboring user groups in each group of random 100 service requests is shown in Figure 7.

In each time the service request, the location service candidate set was recommended to the service requester according to the collaborative filtering recommendation method. In the experiment, the number of service candidate set elements was obtained in each service request when the r values were 100, 200, and 300 meters, respectively. And the results are shown in Figure 8.

When the r values were 100, 200, and 300, respectively, the specific statistics of the number of responding users are shown in Table 1.

It is easy to see from the above data that as the value of r increases, the number of responding neighboring user groups and the service recommendation results increase. When the r value was 100 meters, there were fewer cases in response to the neighboring user group and the number of recommended services. When the r value was 200 or 300 meters, the number of responding users and recommended service results tends to be stable and better.

4.4. Comparison among CDCLS, DCRLS, and MobiCrowd. The proposed algorithms CDCLS, DCRLS, and MobiCrowd are based on the distributed system structure of the location service system model, but they have significant differences in the degree of dependence on centralized servers, location information privacy protection, execution efficiency, and communication costs. The above performance indicators were compared and analyzed on the CRAWDAD dataset, and the results are shown in Table 2.

The MobiCrowd algorithm relied heavily on centralized location servers, and the server’s access frequency could be reduced when users accumulate a certain location service result. The DCRLS algorithm mainly used its own location information to provide location services and need to access centralized location servers when it failed. Compared to the two algorithms, the CDCLS algorithm had the least dependence on centralized location servers. When the d_c value was properly set, the chance of

FIGURE 7: The number of response groups.

FIGURE 8: The number of service results.

TABLE 1: Number of service result sets.

d_c (m)	Minimum	Maximum	Average	Failure rate (%)
100	1	752	153.91	0
200	5	1294	340.56	0
300	13	1485	498.55	0

TABLE 2: The comparison of the architecture among different algorithms.

	CDCLS	DCRLS	MobiCrowd
Architecture tiers	2 tiers	2 tiers	3 tiers
Dependence on trusted third party	Low	Low	Medium
Privacy protect among peers	Good	Medium	Weak
Location service quality	Good	Medium	Low

accessing the centralized location server was small. When the three algorithms randomly request 100 location services, the number of access location servers was compared with different values of d_c . The result is shown in Figure 9.

In the same experimental environment, the average communication cost, the average time spent on the server, and the client of the three algorithms are shown in Figure 10.

FIGURE 9: The comparison for the number of access servers among different algorithms. (a) $d_c = 100$. (b) $d_c = 200$. (c) $d_c = 300$.

FIGURE 10: Performance comparison among different algorithms with $d_c = 100, 200$, and 300 . (a) Communication cost. (b) Client cost. (c) Server side cost.

And the communication cost was measured by the number of TCP/IP packets.

It can be seen from Figures 10(a) and 10(b) that the communications load of the CDCLS algorithm and the time cost of the client were higher than the DCRLS and MobiCrowd algorithms. The main reason is that the neighborhood user group partitioning was performed by using density and absolute density with the CDCLS algorithm. Comparing the other two algorithms, the more content was calculated in the CDCLS algorithm. The time spent on the server side of the three algorithms is compared in Figure 10(c). The MobiCrowd gradually increased the time-based service information in the cache, and the time cost on the server side was gradually reduced, but still higher than CDCLS and DCRLS. From the analysis results, the data communication load of the CDCLS algorithm mainly occurred on the client side. But if the d_c value was set reasonably, the time cost of the server side could be ignored. Thus, the analysis results achieved the design goal of the algorithm. It could be confirmed that the proposed algorithm provided better location service quality based on distributed systems, balanced communication load, reduced access to location servers, and avoided the risk of

privacy leakage in existing centralized location service models.

5. Conclusion

In this paper, a cluster-based distributed collaborative filtering recommendation location service strategy was proposed to overcome the low quality of location service and the risk of privacy leakage in existing location service system. In our work, the user location information profile based on the density metric and the (k, d) neighbor user group were constructed, the collaborative filtering method was used to recommend the location service, and finally the algorithm was analyzed on the real data set. The analysis showed that the proposed strategy could provide higher quality location service results, do not rely on third-party centralized location servers, overcome the communication bottlenecks and attack defects of the existing centralized system architecture, and ensure privacy security. In the future work, we will pay more attention to the location information privacy security issue and intend to provide more stringent data privacy security assurance without increasing the data computing load.

Data Availability

The original data set used in this paper is the Cab mobility traces. The Cab mobility traces are provided by the Exploratorium—the museum of science, art, and human perception through the cabspotting project: <http://cabspotting.org>. “Each San Francisco based Yellow Cab vehicle is currently outfitted with a GPS tracking device that is used by dispatchers to efficiently reach customers. The data are transmitted from each cab to a central receiving station and then delivered in real-time to dispatch computers via a central server. This system broadcasts the cab call number, location, and whether the cab currently has a fare.” Cab mobility traces can be collected by following the instructions from <http://cabspotting.org/api>. The relevant data sets supporting the conclusions of this work are available in the manuscript or at the following URL: <https://github.com/wangpeng68/Cab-mobility-traces>.

Conflicts of Interest

The authors have declared that no conflicts of interest exist.

Authors' Contributions

Peng Wang was responsible for conceptualization, data curation, formal analysis, validation, visualization, original draft writing, reviewing, and editing; Jing Yang and Jianpei Zhang were involved in funding acquisition; Peng Wang and Jing Yang were responsible for study methodology; Peng Wang, Jing Yang, and Jianpei Zhang were involved in resource acquisition.

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China under grant nos. 61672179, 61370083, and 61402126; the Natural Science Foundation Heilongjiang Province of China under grant no. F2015030; the Youths Science Foundation of Heilongjiang Province of China under grant no. QC2016083; and the Heilongjiang Postdoctoral Science Foundation under grant no. LBH-Z14071.

References

- [1] M. P. Singh, B. Yu, and M. Venkatraman, “Community-based service location,” *Communications of the ACM*, vol. 44, no. 4, pp. 49–54, 2001.
- [2] J. Bobadilla, F. Serradilla, and J. Bernal, “A new collaborative filtering metric that improves the behavior of recommender systems,” *Knowledge-Based Systems*, vol. 23, no. 6, pp. 520–528, 2010.
- [3] M. Grutester and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” in *Proceedings of the First International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pp. 31–42, San Francisco, CA, USA, May 2003.
- [4] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: a survey of recent developments,” *ACM Computing Surveys*, vol. 42, no. 4, pp. 1–53, 2010.
- [5] H. To, G. Ghinita, L. Fan, and C. Shahabi, “Differentially private location protection for worker datasets in spatial crowdsourcing,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 4, pp. 934–949, 2017.
- [6] C.-Y. Chow, M. F. Mokbel, and X. Liu, “Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments,” *Geoinformatica*, vol. 15, no. 2, pp. 351–380, 2011.
- [7] R. Shokri, “Hiding in mobile crowd: location privacy through collaboration,” *IEEE Transactions on Dependable Secure Computer*, vol. 11, no. 3, pp. 266–279, 2014.
- [8] M. Zhang, J. Liu, Y. Liu et al., “Recommending pick-up points for taxi-drivers based on spatio-temporal clustering,” in *Proceedings of the International Conference on Cloud & Green Computing*, pp. 67–72, IEEE Computer Society, Hunan, China, November 2012.
- [9] P. Wang, J. Yang, and J. P. Zhang, “Protection of location privacy based on distributed collaborative recommendations,” *PLoS One*, vol. 11, no. 9, Article ID e0163053, 2016.
- [10] P. Wang, J. Yang, and J. Zhang, “A strategy toward collaborative filter recommended location service for privacy protection,” *Sensors*, vol. 18, no. 5, p. 1522, 2018.
- [11] A. Shahriyar, L. Janne, I. Hong Jason et al., “Cache: caching location-enhanced content to improve user privacy,” in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, Bethesda, MD, USA, June–July 2011.
- [12] R. Shokri, P. Papadimitratos, G. Theodorakopoulos, and J. P. Hubaux, “Collaborative location privacy,” in *Proceedings of the IEEE 8th International Conference on Mobile Adhoc and Sensor Systems*, Valencia, Spain, October 2011.
- [13] C. Chow, M. F. Mokbel, and X. Liu, “A peer-to-peer spatial cloaking algorithm for anonymous location-based services,” in *Proceedings of the ACM Symposium on Advances in Geographic Information Systems (ACM GIS'06)*, pp. 171–178, Arlington, VA, USA, 2006.
- [14] L. Sweeney, “*k*-anonymity: a model for protecting privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.
- [15] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkatasubramanian, “*L*-diversity,” *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, 2007.
- [16] Y. Liu, B. Yang, and L. I. Guangyuan, “A personalized privacy preserving parallel (α , k)-anonymity model,” *International Journal of Advancements in Computing Technology*, vol. 4, no. 5, pp. 265–271, 2012.
- [17] R. J. Bayardo and R. Agrawal, “Data privacy through optimal *k*-anonymization,” in *Proceedings of the International Conference on Data Engineering (ICDE 2005)*, pp. 217–228, IEEE, Tokyo, Japan, April 2005.
- [18] G. Kou, Y. Peng, and G. Wang, “Evaluation of clustering algorithms for financial risk analysis using MCDM methods,” *Information Sciences*, vol. 275, pp. 1–12, 2014.
- [19] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [20] A. Boutet, D. Frey, R. Guerraoui, A. Jégou, and A.-M. Kermarrec, “Privacy-preserving distributed collaborative filtering,” *Computing*, vol. 98, no. 8, pp. 827–846, 2016.
- [21] K. Chen and L. Liu, “Privacy-preserving multiparty collaborative mining with geometric data perturbation,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 12, pp. 1764–1776, 2009.

- [22] T. Zhu, Y. Ren, W. Zhou, J. Rong, and P. Xiong, "An effective privacy preserving algorithm for neighborhood-based collaborative filtering," *Future Generation Computer Systems*, vol. 36, pp. 142–155, 2014.
- [23] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes" in *Advances in Cryptology EURO-CRYPT99*, vol. 1592 of Lecture Notes in Computer Science, J. Stem, Ed., pp. 223–238, Springer, Berlin, Germany, 1999.
- [24] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *Proceedings of the First International Communication Systems and Networks and Workshops (COMS-NETS 2009)*, pp. 1–10, Bangalore, India, January 2009.
- [25] J. Domingo-Ferrer and R. Trujillo-Rasua, "Microaggregation-and permutation-based anonymization of movement data," *Information Sciences*, vol. 208, pp. 55–80, 2012.