Research Article

RPAR: Location Privacy Preserving via Repartitioning Anonymous Region in Mobile Social Network

Jinquan Zhang,^{1,2} Yanfeng Yuan,¹ Xiao Wang,¹ Lina Ni ^{1,2,3},
Jiguo Yu ^{4,5,6} and Mengmeng Zhang¹

¹College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

²Shandong Province Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science and Technology, Qingdao 266590, China

³Key Laboratory of the Ministry of Education for Embedded System and Service Computing, Tongji University, Shanghai 201804, China

⁴School of Information Science and Engineering, Qufu Normal University, Shandong 276826, China

⁵Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

⁶Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan 250014, China

Correspondence should be addressed to Lina Ni; nln2004@163.com

Received 11 June 2018; Revised 2 October 2018; Accepted 17 October 2018; Published 5 November 2018

Guest Editor: Yan Huo

Copyright © 2018 Jinquan Zhang 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.

Applying the proliferated location-based services (LBSs) to social networks has spawned mobile social network (MSN) services that allow users to discover potential friends around them. While enjoying the convenience of MSN services, the mobile users also are confronted with the risk of location disclosure, which is a severe privacy preserving concern. In this paper, we focus on the problem of location privacy preserving in MSN. Particularly, we propose a repartitioning anonymous region for location privacy preserving (RPAR) scheme based on the central anonymous location which minimizes the traffic between the anonymous server and the LBS server while protecting the privacy of the user location. Furthermore, our scheme enables the users to get more accurate query results, thus improving the quality of the location service. Simulation results show that our proposed scheme can effectively reduce the area of anonymous regions and minimize the traffic.

1. Introduction

Internet of Things (IoT), a trend of future networks, is immersed into many aspects of our personal and working lives and provides more comprehensive intelligent service. Social networks widely used in mobile Internet catalyze mobile social networks (MSNs), and users in MSN can not only acquire their own location information and sign in a location but also find nearby friends and access to location-based services (LBSs) such as finding the nearest hotel, finding directions, sharing action tracks, obtaining information of body area networks, and so on [1–7]. However, when we enjoy the convenience of LBS and MSN services, the mobile users also are confronted with the risk of location disclosure, which is a severe privacy preserving concern [8–12].

MSN services (MSNS) have a wide range of applications in people's daily life, where location-aware information plays

a very important role [13]. In addition to supporting real-time services, location-based applications can predict the behavior of the users by analyzing the user's position traces to obtain the user's interest preferences and make user interest recommendations. With the development of MSNS, location privacy issues are gradually attracting more and more people's attention. The location information includes the user's identity information, location coordinates information, time stamp, and other sensitive information. Although analyzing the users' locations and trajectories can better support MSNS and recommended services, it is easier for attackers to attack the user's location information so as to expose user's privacy. With the use of MSNS, mobile users are increasingly aware of the risk of privacy disclosure when enjoying location services [14–18].

Location-based privacy protection based on LBS is designed to prevent malicious attackers from gaining access

to the mobile user's location (or motion trajectory) to prevent user information from being compromised. Since the concept of LBS has been proposed, location-based privacy protection has quickly been raised and developed into academic research hot spots. The user's location privacy mainly includes location privacy, trajectory privacy, and user identity privacy. Aiming at these three aspects, quite a few privacy protection methods have been proposed, including commonly used dummy location technology, temporal and spatial anonymous technology, pseudonym technology and other methods protecting the users' location privacy, trajectory privacy, and identity information privacy [19–21]. Although the existing location privacy protection methods that apply to MSN can resist common privacy attacks [22], there are still some weaknesses to be resolved.

In this paper, we focus on the location privacy preserving in MSN aiming at larger communication overhead, larger range, and inaccuracy of query results for traditional anonymous schemes. The main contributions of this paper are summarized below.

- (1) We propose a repartitioning anonymous region for location privacy preserving (RPAR) scheme based on the central anonymous location. The anonymous region is divided into several subregions, users' real locations are replaced by the central location, and a repartition is carried out to solve the tail anonymity user set after the anonymous region partition.
- (2) We analyze the superiorities of RPAR algorithm. RPAR minimizes the communication traffic between the anonymous server and the LBS server while protecting the privacy of the user location. Furthermore, our scheme enables the users to obtain the more accurate query results, thus improving the quality of LBS.
- (3) We simulate RPAR algorithm in extensive experiments. Simulation results show that our proposed scheme can effectively reduce the area of anonymous regions and minimize the traffic.

The rest of the paper is organized as follows. Section 2 reviews the related work. In Section 3, we give the preliminaries of location privacy preservation. Following that, in Section 4, the anonymous region repartition algorithm is proposed. In Section 5, simulations are given to verify the effectiveness of our proposed models and algorithms. Finally, we draw our conclusions and give the future work in Section 6.

2. Related Work

Recently, there have been considerable interests in the research of privacy preservation technologies for location-based services (LBSs) [23].

The k -anonymity scheme was first proposed to be Sweeney's application in a relational database [24] to protect sensitive properties by generalizing some of the nonsensitive properties in the database in case of stealing by adversaries. The generalization process makes the correlation between

user information and their related record fuzzy, and every record in data table published is indistinguishable from other $k - 1$ records. The scheme hides the link between the user and its corresponding information, guaranteeing the user's privacy information security.

To protect location privacy, researchers like Gruteser first proposed to apply the k -anonymity model to the location privacy preservation [25]. By constructing a location region composed of a query user and $k - 1$ common users, the k users' geographical position information is generalized to be a k -anonymity region. The users being generalized cannot be distinguished from other $k - 1$ users in terms of identity information, geographical position information, and so on. Even though the adversaries can access a user's location information in the anonymous region, they cannot identify the correspondence between the locations of the query information issued and the relevant user, protecting users' privacy information. The location anonymity scheme, proposed in [26], splits a contiguous anonymous region into several scattered subanonymity regions and demonstrates its effectiveness.

The privacy preserving scheme covers not only k -anonymity technology with strong applicability but also location privacy preserving scheme based on fake information. Bamba adopts virtual location information to hide real information in order to achieve location privacy preserving effect in [27]. Its main idea is as follows: Firstly users put forward location service requests and privacy requirements; then central anonymous servers send their fake position information and the real one to LBS servers; at last the LBS servers return query results back to anonymous servers, and the anonymous servers calculate the correct results and return them back to the corresponding users. Due to mixture of true and fake information, it is difficult for adversaries to distinguish the users' actual query information; thus the users' privacy information is protected.

Duckham et al. proposed to use obfuscation technique as an effective location privacy preserving mechanism [28] as well as position fuzzy region to process fuzzy query, satisfying the needs of privacy protection while sacrificing the quality of the service. Hence, how to strike a balance between location privacy preservation and service queries deserves our further study.

Ni et al. studied location privacy neighbor query based privacy preference in [29]; both the contradiction between the location protection and personal privacy preference and the balance between location privacy protection and query service quality problems are analyzed and summarized. Xu et al. proposed the location privacy region generation algorithm, centroid migration method [30], based on the spatial confusion location privacy protection technology; thus a certain deviation of centroid in the location privacy region was made. Ye et al. put forward a kind of location privacy protection scheme [31] based on active sharing mechanism; the main idea of this method is actively sharing the user's location information by sharing mechanism, so as to reduce the users' dependency on LBS servers and enhance the protective effect of the location privacy information.

FIGURE 1: Architecture of our system model.

In recent years, domestic and foreign researchers and institutions have widespread attention on location-based privacy preservation; increasingly in-depth studies are carried out in location-based privacy. Besides the above-described location privacy preserving technologies, there are a wealth of methods such as location data randomization [32], fuzzification of space or time data [33, 34], methods based on strategies and encryption [35–37], and sensitive semantic based security anonymity mechanism [33, 38–40].

In [41], we have proposed a preliminary location privacy preserving scheme via repartitioning anonymous region in MSN. However, this scheme did not elaborate the design motivation and the algorithm analysis. Also, this scheme did not implement the simulations to verify its performance.

3. Preliminaries

The k -anonymity privacy preserving scheme was first applied to data releasing; extensive research and application are carried out after the application on the fields of location-based privacy preservation. Thereafter, anonymous region segmentation method is proposed to solve the insufficiency on large communication overhead of k -anonymity. However, in many cases, anonymous regions that are partitioned once cannot achieve ideal effect; therefore we put forward the repartitioning anonymous region scheme for location privacy preserving (RPAR).

3.1. System Model. At present, although there are many communication modes that are unsuitable for the central anonymous servers, current location privacy preserving schemes mainly adopt the central server mode [42, 43]. Our scheme uses the classic central server mode, and the anonymous queries are processed through both the central anonymous servers and the LBS servers. As shown in Figure 1, the principles of the model are as follows:

- (1) When the user requests the location query service, all the query contents, location information, and parameters needed to be set are sent to the central anonymous servers.
- (2) After receiving the query information sent by users, according to certain rules, the central anonymous servers will generate an anonymous user set which

meets the requirements, figure out the number of subanonymity regions, and then partition them; a few scattered subanonymity regions are yielded. When the subanonymity regions meet the requirements, their central location is computed to replace corresponding subanonymity regions to send requests to the LBS server.

- (3) The LBS server handles the query information sent by the central anonymous servers and returns the query results.
- (4) After the refinement process, the central anonymous servers return the corresponding results to the users.

3.1.1. Relevant Definitions. To describe RPAR scheme, it is necessary to introduce some relative knowledge of location anonymity in the division process. According to the requirements of the scheme, we refer to the relevant definitions like location k -anonymity, nearest neighbor principles, etc., and several definitions such as central anonymity region location and the tail anonymity user set are also proposed.

Definition 1 ((location k -anonymity) [44]). Suppose that there exists a mobile user whose location coordinates is (x, y) . If the user and at least other $k-1$ users cannot be differentiated by location information after the generalization for this user, we can say that the k users' locations satisfy location k -anonymity.

The k users' information forms a user anonymity set. Note that the least rectangular region which includes all the location K -anonymity users is called the location k -anonymity region. It is not difficult to see in Figure 2 that the rectangular box is an anonymous region with anonymous user set where $k = 12$.

Formally, we use AR to represent a location k -anonymity region. Thus, according to certain rules, AR can be divided into n discrete rectangular anonymity regions, represented by $AR = \{sub_{AR_1}, sub_{AR_2}, \dots, sub_{AR_n}\}$. It should be clear that these small scattered anonymity regions $sub_{AR_i} (i = 1, 2, \dots, n)$ are all the subanonymity regions of continuous anonymity region AR . Figures 2 and 3 depict the status before and after the anonymous regions are repartitioned, respectively.

FIGURE 2: Before partition of k -anonymity region. FIGURE 3: After partition of k -anonymity region.

Definition 2 (central location of anonymous region). The location of the two diagonals of a rectangular subanonymity region is said to be its central location, which is represented by coordinates (x', y') .

We will take central location as a fake location to issue location service requests by replacing the subanonymity regions.

Definition 3 (tail anonymity user set). After partitioning the location k -anonymity region AR , if k/n is not an integer, then the supernumerary $k \bmod n$ users form a tail anonymity user set.

Definition 4 (nearest neighbor principle). Take a location point as the center, and find other location points in accordance with the priority principle of the nearest Euclidean distance from the center point.

4. Location Repartitioning Anonymous Region Scheme

In order to overcome the drawbacks of the exiting schemes, including the oversized subanonymity regions and the inaccuracy of query results during the anonymous region partition, we propose the repartitioning anonymous region for location privacy preserving (RPAR) scheme. The oversized subanonymity regions are further partitioned, and the central locations of the subanonymity regions replace those subanonymity regions to issue queries to the servers, so as to

reduce the communication overhead and achieve relatively accurate query results.

4.1. Motivation. According to the traditional anonymity schemes, when the users request LBSs, their true locations will be replaced with anonymous region and issue requests to LBS servers. However, if the anonymous parameter k is relatively bigger, then the anonymous regions will be relatively larger accordingly, especially apparent in the scenario where the users are sparse. Although the privacy preserving is in a high level, the accuracy of the request results is low, resulting in poor service quality.

Roman et al. [45] proposed anonymous partition algorithm for the first time, splitting the continuous k -anonymity region into several discrete subanonymity regions. Compared with the traditional continuous anonymity regions, the method in [45] narrows the region of the anonymous regions, improving the service quality to a certain extent. However, the final returned candidate result set is still fuzzy by querying through subanonymity regions, and the communication traffic has a certain decrease but the overhead is still large.

In addition, when there are still remaining users after the subanonymity regions are partitioned (i.e., there remains tail anonymity user set), the method in [45] puts all the remaining users in one subanonymity region. As shown in Figure 2, in the anonymous region where $k = 14$, the red solid dots represent the users who issue the query requests, and the white dots represent other real users in the k -anonymity region. We set the segmentation parameter $n = 3$, indicating that the anonymous user set is divided into three subanonymity regions, and the number of users in each subanonymity region is $k' = k/n = 4$, as shown in Figure 3. However, the following problems will occur when the scenario in subanonymity regions is consistent with that in Figure 3:

- (1) The real locations of the users are replaced by the decentralized subanonymity regions in the corresponding regions to issued queries, which can still produce a large candidate result set and require a large amount of communication overhead.
- (2) When the tail anonymity user set, that is, the value of $k \bmod n$, is too large, it will result in the number of users in the subanonymity regions divided by the tail anonymous user set being larger than that in other subanonymity regions, resulting in great difference in query accuracy.
- (3) As shown in Figure 3, if one of the subregions is oversized or even close to the area of the k -anonymity regions, the total area of subanonymity regions is greater than A_{min} , segmentation algorithm will return back to the k -anonymity algorithm with fewer users. The locations of the real users are far from the central region, causing queries to be not accurate.

In order to overcome these weak points, we construct RPAR scheme. Using the central location of the subanonymity regions instead of the subanonymity regions to issue queries to the servers, the communication traffic is

FIGURE 4: Repartition of anonymous region.

greatly reduced. When total area of subanonymity regions is greater than A_{min} , the largest subanonymity region is repartitioned into the other subanonymity regions, as shown in Figure 4, until the total area of the subanonymity regions is not greater than A_{min} anymore.

For the query accuracy problem caused by the oversized granularity of the tail anonymity user set, the users in the tail anonymity user set are partitioned into other subanonymity regions in accordance with the nearest neighbor principle.

4.2. Basic Idea of RPAR. Combined with Figures 2, 3, and 4, the basic idea of RPAR scheme is elaborated as follows:

- (1) The solid red dots represent the users initiating query who find $k - 1$ users with the query users who can form k -anonymity regions according to the nearest neighbor principle, and k users information set is recorded. It can be seen in Figure 2 that $k = 14$.
- (2) According to the parameter n , that is, the number of subanonymity regions, the k users are divided into n subanonymity regions, so the number of users each subanonymity region contains is $k' = k/n = 4$. The mobile users (red dots) are as the center to search other nearest $k' - 1$ users and form the first subanonymity region.
- (3) A user from the rest of the users that is not in the first subanonymity region is randomly selected as the central point. According to the nearest neighbor principle, the subanonymity regions are formed with the user and other 3 users who are not in the first subregion from anonymous region, until the remaining user number is 0 or below k' .
- (4) The tail anonymity user set is repartitioned into other subanonymity regions according to the nearest neighbor principle, and the subanonymity regions are updated.
- (5) The area of the subanonymity regions is calculated. If the total area of the subanonymity regions is greater than A_{min} , users' maximum subanonymity region will be repartitioned until the total area of the subanonymity regions is not greater than A_{min} .

- (6) The central locations of all the subanonymity regions are computed which are used to replace their subanonymity regions to issue queries to the LBS servers.

4.3. Algorithm Design. Let $Q = (id, L, S, M)$ be a user's location service query. For each Q , id represents the user identity, M represents the query information, L represents the location information, such as the user's location coordinates (x, y) , and S is the privacy parameters including anonymous region parameter k , homogeneity parameter l (difference degree in anonymous query target regions), subanonymity region number n , and user's smallest total subanonymity area A_{min} . The parameters like k , l , and n are all user-defined. In practice, users can determine the values of these parameters according to their own requirements and sensitivity. Thus, the algorithm can generate different degrees of anonymity to guarantee the performance.

For minimum area of anonymous regions, calculation probability is inferred through the Bayesian networks and the related background knowledge based on social network model, and minimum anonymous area is obtained by means of the maximum likelihood estimator. Figure 5 is the Bayesian network diagram for anonymous region estimation.

In the Bayesian network diagram, Q and k represent users' query information, A_{min} represents the minimum anonymous area, $User$ is the user's query information, M is the user's query content such as hotel, hospital, *uid* and *position*, respectively, represent the user's identity and location information, *time* represents the query time, and l is the diversity parameter of query content which is not discussed in our paper.

In the process of calculation and estimation, without loss of generality, the relatively weaker nodes are ignored in order to simplify the computation complexity. The reasoning process of node A_{min} is described as follows:

- (1) The probability distribution of the second and third nodes in the second layer is calculated:

$$P(User | e) \propto P(e_{c_j} | User) P(User | k) \propto \prod_{j=1}^3 P(e_{c_j} | User) \quad (1)$$

$$P(User | e) = P\left(\frac{l}{M}\right) P(M | k) \propto P\left(\frac{l}{k} | M\right) \quad (2)$$

- (2) Through the probability distribution of the first node in the first layer,

$$P(k | e) = P(e_{c_j} | k) = \prod_{j=1}^3 P(e_{c_j} | k) \quad (3)$$

By (1) to (3), we have

$$P(A_{min} | k) = \left(\frac{P(k)}{P(User | e)} \right) P(M | e) \quad (4)$$

where e represents all the background knowledge and e_{c_j} represents the background knowledge of the child nodes.

FIGURE 5: The Bayesian network for anonymous region estimation.

Input: $uid, (x, y), k, l, n, A_{min}, M$, other users' location information in anonymous set stored in anonymous servers.
Output: central locations (x_i, y_i) and M .

- (1) $k' = \lfloor k/n \rfloor$
- (2) form the first sub-anonymity region sub_{AR_1} , and calculate its *central location* (x_1, y_1) and area S_1 based on $(x, y), k'$ and other user location information.
- (3) **for** $i = 2$ **to** n **do**
- (4) $m = k - k' * (i - 1)$;
- (5) select user u_i randomly from the rest of the *tail anonymity user set* U_m : $u_i = \text{Random}(U_m)$;
- (6) form sub_{AR_i} and calculate its *central location* (x_i, y_i) and area S_i according to the information of u_i and parameter settings;
- (7) calculate total area of all the sub-anonymity regions: $S_A = S_A + S_i$;
- (8) **end for**
- (9) **if** $k \bmod n == 0$ **then**
- (10) **if** $S_A > A_{min}$ **then**
- (11) repartition the largest sub-anonymity region until $S_A < A_{min}$
- (12) **end if**
- (13) **else**
- (14) repartition the users in the *tail anonymous user set* into sub-anonymity regions according to the *nearest neighbor principle*
- (15) **end if**
- (16) update the *central locations* (x_i, y_i) of the sub-anonymity regions
- (17) return (x_i, y_i) to all users in i -th sub-anonymity region. (x_i, y_i) and M are sent to the LBS servers to query location services.

ALGORITHM 1: RPAR.

Here, the background knowledge refers to maps, historical query records, and historical query probability of a certain area, etc.

When users query location service Q , all the parameters are needed to be sent to the anonymous servers. The parameters include identity information uid , location coordinates (x, y) , anonymous region parameter k , subanonymity region number n , minimum anonymity regions' area A_{min} , query content M , and other users' location coordinates. After the parameters are processed by the anonymous servers, the fake locations are sent to the LBS servers to request service queries. In practice, the area of the anonymous region is calculated based on the coordinates of the position points of the user in the upper left and lower right corners of the anonymous region.

According to the algorithm description, parameter settings, and the algorithm process, the pseudocode of the detailed RPAR algorithm is elaborated in Algorithm 1.

4.4. Algorithm Explication. RPAR is proposed based on the existing k -anonymity algorithm and partitioned sub-anonymity regions to solve the problems of the scattered sub-anonymity regions with total oversized area of anonymous regions and the oversized value of $k \bmod n$.

In the process of the subanonymity region partition, the nearest neighbor principle is used which means searching qualified users according to the distance with nearest and highest priority. The distance in the nearest neighbor principle is Euclidean distance, namely,

$$d = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (5)$$

where (a_1, b_1) and (a_2, b_2) are location coordinates of two points and d is the distance between two points.

According to the locations and parameters of the users who issued the location service requests, the servers find out the other $k - 1$ users nearest to this user, who satisfied the conditions. It is noted that, in the case of the repartitioning

of the subanonymity regions, the nearest neighbor principle is still used to make the request users as the center and form the first subanonymous region by searching the $k_1 - 1$ users, where k_1 is the user number in the first subanonymity region. Then the coordinates (x_1, y_1) of the central position in the first subanonymity region are calculated according to the following formula:

$$\begin{aligned} x_1 &= \frac{a_1 + a_2 + \dots + a_k}{k_1} \\ y_1 &= \frac{b_1 + b_2 + \dots + b_k}{k_1} \end{aligned} \quad (6)$$

where $(a_i, b_i), i = 1, 2, \dots, k_1$, are each user's location coordinates in the first subanonymity region and k_1 is the users' number contained in this region.

After repartitioning the first subanonymity region, a user's location is selected randomly as the center. Other subanonymity regions are formed in accordance with the nearest neighbor principle, and the area and the coordinates of central locations of subanonymity regions are calculated. When n subanonymity regions are divided up, if $k \bmod n = 0$, directly compare whether the total region of subanonymity regions is greater than A_{min} . If the answer is exact, then repartition the biggest anonymous region until the total region is not larger than A_{min} . If $k \bmod n \neq 0$, then the users in the tail anonymity user set are repartitioned into the other subanonymity regions. After all the subanonymity regions are partitioned, the subanonymity regions are replaced with their central location to issue the location service requests to the LBS servers.

4.5. Algorithm Superiority Analysis. The traditional k -anonymity scheme makes k users form an anonymous region to issue queries to the LBS servers. The result of the queries generally is a candidate set. In the scenario with sparse users, the area of the anonymous regions will be large and the searched candidate set will also be correspondingly larger. Li et al. [26] divided the anonymous region into a few scattered subanonymity regions, using each subanonymity region to query; the query candidate set is small accompanied with precision improved; however a lot of communication overhead would cost.

Our RPAR scheme directly splits k users into several distributed subanonymity regions instead of forming a k -anonymity region. The regions are replaced with its center to issue requests to the LBS servers, which greatly reduce the communication overhead, and the accurate query results are obtained which referred to the central location of subanonymity regions, so each user can get their accurate query results by referring to the location. Therefore, compared with the traditional methods, the advantages of RPAR are obvious, especially when the users are sparse.

After dividing the anonymous regions [26], if $k \bmod n \neq 0$, then the rest of the users are put into one of the anonymous regions. RPAR scheme repartitions the rest of the users into their nearest subanonymity regions in accordance with the nearest neighbor principle.

When the value of $k \bmod n$ is relatively small, no matter how to divide the remaining users, the number of users in each anonymous region will not make a big difference, so the query precision degree of results returned also will not make a big difference.

While when the value of $k \bmod n$ is larger, the method in [26] will lead to one anonymous region far larger than the others, the query precision has great difference. In our RPAR, after $k \bmod n$ users are repartitioned according to the nearest neighbor principle, the users in each subanonymity region are much the same, so the result precision degree of the final queries will not be too different.

In this paper, the condition that the total area of subanonymity region is greater than A_{min} is also considered in subanonymity region partition. The value of A_{min} is defined by the user, which generally is the area of the k -anonymity regions. As shown in Figure 3, if some of the area of subanonymity regions is too large or even close to that of the K -anonymity regions, the partition algorithm will be returned back to the k -anonymity algorithm with fewer users. Therefore, when the total area of the subanonymity regions is greater than A_{min} , the biggest subanonymity region is partitioned until the total area of the subanonymity regions is less than A_{min} .

We use the central locations of the subanonymity regions instead of subanonymity regions to issue query requests to the LBS servers. As fake positions, the central locations well hide the true users' locations. Considering the condition that some central positions would coincide with users' actual locations, due to the scheme that all users in the subanonymity regions use fake positions, even if the query results are intercepted by adversaries, identification probability of actual users is at least $1/k_i$.

5. Simulations and Performance Analysis

In this section, we implement RASA and analyze its performance.

5.1. Simulation Environment. As for the experimental environment, Windows 7 with 64-bit operating system with 4G RAM and Intel (R) Pentium (R) CPU G2030@3.00 GHz processor is adopted. Thomas Brinkhoff's mobile object generator [46] is used to generate the spatiotemporal data as the resulting dataset. The privacy preserving parameters of the experiment are determined according to the users' requirements.

5.2. Simulations and Analysis. Assume that the mobile users are deployed in a rectangular region with an area of $5000m * 5000m$. Since the effect of our RPSA scheme is obvious for the scenario with sparse users, the simulation results are generated in this scenario with our RPSA scheme.

In the following, we will compare our RPAR algorithm with the traditional k -anonymity, FCR [26], and DSCR [47] algorithms. FCR directly generates the total area of the subanonymity regions during the partition while the remaining $k \bmod n$ users are put in the same subanonymity region. However, DSCR firstly generates contiguous anonymous

FIGURE 6: Area comparison of anonymous regions.

regions; then continuous anonymous regions are divided into several subanonymity regions which may show a large number of repetitive regions.

We analyze the performance of RPAR by comparing the anonymous area, anonymous area percentage, and anonymous time of these methods with the same anonymous parameters. We assume that the value of the parameter k in the simulation is an integer in the range from 6 to 15. We also set the number of subanonymity regions n as 3. The comparison diagrams are illustrated in Figures 6, 7, and 8.

As shown in Figure 6, the area in the anonymous regions produced by our RPAR is not much different from that in FCR, whereas the area is greatly reduced compared with traditional k -anonymity and DSCR. While $k \bmod n \neq 0$ in FCR, FCR puts the rest of the $k \bmod n$ users into a subanonymity region; thus this area is oversized, causing the user query precision of this anonymous region and other subanonymity regions to vary greatly. Fortunately, our RPAR effectively solves the above problem. Note that, with the k value increasing, the area increases more slowly and the reduced area is increasingly large in RPAR compared with other three anonymous region partition schemes, making RPAR superior obviously.

Figure 7 is the comparison of the ratio of the effective area of anonymous region generated by DSCR, FCR, and RPAR with the traditional k -anonymity, namely,

$$\begin{aligned} per_2 &= \frac{Area_{DSCR}}{Area_{k-anonymity}}, \\ per_3 &= \frac{Area_{FCR}}{Area_{k-anonymity}}, \\ per_4 &= \frac{Area_{RPAR}}{Area_{k-anonymity}}. \end{aligned} \quad (7)$$

It can be seen from Figure 7 that the area ratio of DSCR is larger, and the area ratio of FCR is closer to that of RPAR.

FIGURE 7: Region percentage comparison of anonymous regions.

FIGURE 8: Comparison of anonymous time.

As a whole, the size of k is positively related to per_i , where $i = 2, 3, 4$. With the increase of k , the area of valid anonymous regions becomes increasingly large and stable. However, our RPAR scheme performs better than FCR as the value of $k \bmod n$ is larger gradually. We observe that RPAR can also obtain stable and higher quality service while reducing the area of anonymous regions, whereas FCR scheme is not stable.

Figure 8 compares the time of anonymous regions in the algorithms. Traditional k -anonymity scheme directly generates consecutive k -anonymity regions; DSCR firstly generates k -anonymity regions which are then partitioned into subanonymity regions, and the $k \bmod n$ tail-anonymity regions are processed; therefore, the anonymous time is longer than the traditional k -anonymity. More specifically, anonymous time is relatively short when $k \bmod n = 0$ while

the anonymous time is relatively long when $k \bmod n \neq 0$. FCR directly generates subanonymity regions, which results in the anonymous time being shorter. Note that our RPAR needs to deal with the tail anonymous user set in the process of directly generating subanonymity regions; therefore the time may be longer than both traditional k -anonymity and FCR, but less than DSCR. Fortunately, the additional average time is less than 0.1 ms, which is quite small, and it will not make a big difference to the algorithm performance.

6. Conclusions and Future Work

Aiming at large communication overhead, large range, and inaccuracy of query results for traditional anonymous schemes, this paper proposed an anonymous region repartition algorithm by studying the users' location privacy preservation. The anonymous region is divided into several subregions, the users' real locations are replaced by the central location, and a repartition is carried out to solve the remaining users' region set after the anonymous region segmentation. Finally, the algorithm is analyzed and the privacy degree is evaluated, and the simulation experiment is carried out. Experimental results show that the proposed scheme has some advantages in privacy anonymity.

In the future, we will research the location privacy preserving in the scenario of dense region. Furthermore, we will research the location privacy preserving under distributed environment.

Data Availability

Thomas Brinkhoff's mobile object generator is used to generate the spatiotemporal data as the resulting dataset in our paper, and the website is <http://iapg.jade-hs.de/personen/brinkhoff/generator/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by National Key R&D Programs Project of China under Grant 2017YFC0804406, NSF of China under Grant 61672321, 61771289, and 61373027, Training Program of the Major Research Plan of NSF of China under Grant 91746104, Project of Shandong Province Higher Educational Science and Technology Program under Grant J15LN19, and Open Project of Tongji University Embedded System and Service Computing of Ministry of Education of China under Grant ESSCKF 2015-02.

References

- [1] T. Song, R. Li, B. Mei, J. Yu, X. Xing, and X. Cheng, "A Privacy Preserving Communication Protocol for IoT Applications in Smart Homes," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1844–1852, 2017.
- [2] R. Li, T. Song, N. Capurso, J. Yu, J. Couture, and X. Cheng, "IoT Applications on Secure Smart Shopping System," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1945–1954, 2017.
- [3] T. Song, N. Capurso, X. Cheng, J. Yu, B. Chen, and W. Zhao, "Enhancing GPS with lane-level navigation to facilitate highway driving," *IEEE Transactions on Vehicular Technology*, vol. 99, 2017.
- [4] C. Hu, H. Li, Y. Huo, T. Xiang, and X. Liao, "Secure and Efficient Data Communication Protocol for Wireless Body Area Networks," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 2, pp. 94–107, 2016.
- [5] C. Hu, N. Zhang, H. Li, X. Cheng, and X. Liao, "Body area network security: a fuzzy attribute-based signcryption scheme," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 37–46, 2013.
- [6] C. Hu, X. Cheng, F. Zhang, D. Wu, X. Liao, and D. Chen, "OPFKA: secure and efficient ordered-physiological-feature-based key agreement for wireless body area networks," in *Proceedings of the INFOCOM, 2013 Proceedings IEEE*, pp. 2322–2330, 2013.
- [7] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, 2016.
- [8] C. Hu, W. Li, X. Cheng, J. Yu, S. Wang, and R. Bie, "A Secure and Verifiable Access Control Scheme for Big Data Storage in Clouds," *IEEE Transactions on Big Data*, pp. 1–1, 2017.
- [9] X. Zheng, Z. Cai, J. Yu, C. Wang, and Y. Li, "Follow but no track: privacy preserved profile publishing in cyber-physical social systems," *IEEE Internet of Things Journal*, 2017.
- [10] A. Alrawais, A. Alhothaily, J. Yu, C. Hu, and X. Cheng, "SecureGuard: A Certificate Validation System in Public Key Infrastructure," *IEEE Transactions on Vehicular Technology*, 2018.
- [11] X. Zheng, Z. Cai, G. Luo, L. Tian, and X. Bai, "Privacy-preserved community discovery in online social networks," *Future Generation Computer Systems*, 2018.
- [12] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Network*, vol. 32, no. 4, pp. 8–14, 2018.
- [13] G. Sun, Y. Xie, D. Liao, H. Yu, and V. Chang, "User-defined privacy location-sharing system in mobile online social networks," *Journal of Network and Computer Applications*, vol. 86, pp. 34–45, 2017.
- [14] X. Xiao, C. Chen, A. K. Sangaiah, G. Hu, R. Ye, and Y. Jiang, "CenLocShare: A centralized privacy-preserving location-sharing system for mobile online social networks," *Future Generation Computer Systems*, vol. 86, pp. 863–872, 2018.
- [15] H. Shen, G. Bai, M. Yang, and Z. Wang, "Protecting trajectory privacy: A user-centric analysis," *Journal of Network and Computer Applications*, vol. 82, pp. 128–139, 2017.
- [16] X. Zheng, G. Luo, and Z. Cai, "A fair mechanism for private data publication in online social networks," *IEEE Transactions on Network Science and Engineering*, 2018.
- [17] H. Chen and W. Lou, "On protecting end-to-end location privacy against local eavesdropper in wireless sensor networks," *Pervasive and Mobile Computing*, vol. 16, pp. 36–50, 2015.
- [18] Y. Liang, Z. Cai, Q. Han, and Y. Li, "Location privacy leakage through sensory data," *Security and Communication Networks*, vol. 2017, Article ID 7576307, 12 pages, 2017.

- [19] M. L. Damiani, "Privacy enhancing techniques for the protection of mobility patterns in LBS: Research issues and trends," in *Security and Communication Networks*, vol. 244, pp. 223–239, Springer Netherlands, 2012, European Data Protection: Coming of Age.
- [20] K. Xing, C. Hu, J. Yu, X. Cheng, and F. Zhang, "Mutual privacy preserving k -means clustering in social participatory sensing," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2066–2076, 2017.
- [21] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, and Y. Li, "Cost-efficient strategies for restraining rumor spreading in mobile social networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2789–2800, 2017.
- [22] X. Liang, K. Zhang, X. Shen, and X. Lin, "Security and privacy in mobile social networks: challenges and solutions," *IEEE Wireless Communications Magazine*, vol. 21, no. 1, pp. 33–41, 2014.
- [23] A. Sadilek, H. Kautz, and J. P. Bigham, "Finding your friends and following them to where you are," in *Proceedings of the 5th ACM International Conference on Web Search and Data Mining, WSDM 2012*, pp. 723–732, USA, February 2012.
- [24] 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.
- [25] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, pp. 31–42, ACM, San Francisco, Calif, USA, 2003.
- [26] T. C. Li and W. T. Zhu, "Protecting user anonymity in location-based services with fragmented cloaking region," in *Proceedings of the IEEE Conf. on Computer Science and Automation Engineering (CSAE)*, vol. 3, pp. 227–231, Zhangjiajie, China, 2012.
- [27] B. Bamba, L. Liu, P. Pesti et al., "Supporting anonymous location queries in mobile environments with privacy grid," in *Proceedings of the 17th ACM Inter. Conf. on World Wide Web (WWW)*, pp. 237–246, Beijing, China, 2008.
- [28] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in *Proceedings of the Inter. Conf. on Pervasive Computing (PC)*, pp. 152–170, Munich, Germany, 2005.
- [29] W. Ni and X. Chen, "Research on privacy preference in privacy location," *Journal of Software*, vol. 27, no. 7, pp. 125–130, 2016.
- [30] H. Xu, J. Xu, Y. Gong et al., "Location privacy region generation algorithm based on spatial confusion location privacy protection," *Journal of South China Univ. of Technol.: Natural Science Edition*, vol. 42, no. 1, pp. 97–103, 2014.
- [31] A. Ye, S. Lin, J. Ma et al., "An active diffusion location privacy protection method," *Acta Electronica Sinica*, vol. 43, no. 7, pp. 1362–1368, 2015.
- [32] R. Kato, M. Iwata, T. Hara et al., "A dummy-based anonymization method based on user trajectory with pauses," in *Proceedings of the 20th ACM Inter. Conf. on Advances in Geographic Information Systems (AGIS)*, pp. 249–258, California, USA, 2012.
- [33] E. Yigitoglu, M. L. Damiani, O. Abul, and C. Silvestri, "Privacy-preserving sharing of sensitive semantic locations under road-network constraints," in *Proceedings of the 13th IEEE Inter. Conf. on Mobile Data Management (MDM)*, pp. 186–195, 2012.
- [34] X. Pan, J. L. Xu, and X. F. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 8, pp. 1506–1519, 2012.
- [35] D. Chen, P. Zhang, and C. Hu, "Private and precise range search for location based services," in *Proceedings of the IEEE Inter. Conf. on Commun. (ICC)*, pp. 7347–7352, China, 2015.
- [36] L. Li, R. Lu, and C. Huang, "EPLQ: Efficient Privacy-Preserving Location-Based Query over Outsourced Encrypted Data," *IEEE Internet of Things Journal*, vol. 3, no. 2, pp. 206–218, 2016.
- [37] I.-T. Lien, Y.-H. Lin, J.-R. Shieh, and J.-L. Wu, "A novel privacy preserving location-based service protocol with secret circular shift for k -NN search," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 863–873, 2013.
- [38] M. Li, Z. Qin, and C. Wang, "Sensitive semantics-aware personality cloaking on road-network environment," *International Journal of Security and Its Applications*, vol. 8, no. 1, pp. 133–146, 2014.
- [39] C. Hu, A. Alhothaily, A. Alrawais, X. Cheng, C. Sturtivant, and H. Liu, "A secure and verifiable outsourcing scheme for matrix inverse computation," in *Proceedings of the IEEE Conference on Computer Communications (CC) INFOCOM*, 2017.
- [40] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, 2018.
- [41] L. Ni, Y. Yuan, X. Wang, M. Zhang, and J. Zhang, "A Location Privacy Preserving Scheme Based on Repartitioning Anonymous Region in Mobile Social Network," *Procedia Computer Science*, vol. 129, pp. 368–371, 2018.
- [42] L. Zhang, Z. Cai, and X. Wang, "Fakemask: A novel privacy preserving approach for smartphones," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 335–348, 2016.
- [43] R.-H. Hwang, Y.-L. Hsueh, J.-J. Wu, and F.-H. Huang, "Social-Hide: A generic distributed framework for location privacy protection," *Journal of Network and Computer Applications*, vol. 76, pp. 87–100, 2016.
- [44] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k -anonymity location privacy," in *Proceedings of the 32nd IEEE Int. Confer. on Comput. Commun. INFOCOM, Italy*, pp. 2994–3002, 2013.
- [45] R. Yarovsky, F. Bonchi, L. V. Lakshmanan et al., "Anonymizing moving objects: How to hide a mob in a crowd?" in *Proceedings of the ACM 12th Inter. Conf. on Extending Database Technology (EDT): Advances in Database Technology*, pp. 72–83, Saint Petersburg, Russia, 2009.
- [46] "Thomas Brinkhoff Network-based Generator of Moving Objects," <http://iapg.jade-hs.de/personen/brinkhoff/generator/>.
- [47] Z. Zhao, L. Li, F. Zhang et al., "Location privacy protection method based on dispersive anonymous region," *Journal of Shandong Univ. (Natural Science)*, vol. 48, no. 7, pp. 56–61, 2013.