A k-Anonymous Location Privacy Protection Method of Polygon Based on Density Distribution

Yong-Bing Zhang[1,2], Qiu-Yu Zhang[1], Yan Yan[1], Yi-Long Jiang[2], and Mo-Yi Zhang[1]

(Corresponding author: Qiu-Yu Zhang)

School of Computer and Communication, Lanzhou University of Technology[1]
No. 287, Lan-Gong-Ping Road, Lanzhou 730050, China
(Email: zhangqylz@163.com)
Gansu Institute of Mechanical & Electrical Engineering[2]
No. 107, Chi-Yu Road, Tianshui, Gansu 741001, China

(Received July 26, 2019; Revised and Accepted Dec. 28, 2019; First Online Feb. 3, 2020)

Abstract

In order to solve the problem of out-off-balance caused by accuracy of location information between privacy protection security and query service quality, considering basic information comprehensively such as the environment and geographical features and so on, and adopting k-anonymous privacy protection mechanism, we present a k-anonymous location privacy protection method of polygon based on density distribution. Firstly, a k-anonymous irregular polygon region is structured in whole area. Then, according to the preset anonymous region and density threshold, the better effects of anonymous are obtained by expanding the region or adding the random dummy locations. Experimental results show that the proposed method improves the efficiency of anonymous and query accuracy. The balance between privacy protection security and query quality is achieved.

Keywords: Anonymous Region; Density Distribution; Irregular Polygon; k*-Anonymous; Location-Based Service (LBS); Location Privacy Protection*

1 Introduction

With the development of mobile location technology and wireless communication technology, more and more mobile devices in the market have GPS precise positioning function, which makes Location-Based Service (LBS) become one of the most promising services to mobile users [11]. However, when LBS services provide convenience and great benefits to the society, its problem of sensitive information leakage has attached more attentions by many people. Because users' location is shared among different Location Service Providers (LSPs), untrustworthy third parties can easily steal users' privacy via analyzing and comparing these location information [17]. For example, through capturing recent users' trace, some location information can be analyzed by adversary such as home addresses, workplaces, and health conditions, etc. Therefore, it is necessary to ensure the safety of users' location privacy.

In order to prevent the leakage of location privacy information, many different methods are proposed by experts and scholars, including fuzzy method, encryption method and strategy-based method. Because of the better reliability, the fuzzy method is the most commonly used in the field of location privacy protection, which is mainly realized by means of spatial anonymity or dummy location, and needs the help of Fully-Trusted Third Party (TTP) [22]. When there is a location service requirement, the mobile user first sends the query request to the TTP, a k-anonymous region containing the user's location is generated by the TTP and then it will be sent to the LBS server for query. In the existing methods, the anonymous region is constructed by regular geometric shapes. However, the actual terrain is not a regular geometry. Therefore, the area of invalid region is increased greatly, which not only consumes more time, but also reduces the accuracy of the query result.

In the k-anonymous location privacy protection, in order to improve query efficiency and query accuracy, a k-anonymous location privacy protection method of polygon based on density distribution is proposed. In this paper, we give full consideration to the geographical features of the current region and the density distribution. Firstly, a k-anonymous irregular polygon region is structured in whole area. Then, according to the density threshold, the location privacy protection is implemented by combining spatial anonymity and dummy location. The proposed method improves the query accuracy and the query service quality.

Our main contributions can be summarized as follows:

1) According to the characteristics of different geographic shapes, a polygon anonymous region construction method is proposed, which improves the accuracy of query result.

2) A fast polygon generation algorithm is applied to construct anonymous region, which improves the query efficiency.

3) According to neighbor users' density distribution, a location privacy protection method combining spatial anonymity and dummy location is proposed, which improves the effectiveness of location privacy protection.

The remaining part of this paper is organized as follows. Section 2 reviews related work of location privacy protection. Section 3 gives system model of this paper. Section 4 describes two algorithms and analysis. Section 5 gives the experimental results and performance analysis as compared with other related methods. Finally, we conclude our paper in Section 6.

2 Related Work

The location privacy protection methods are divided into two main categories [8] according to the system architecture, including distributed structure [9] based on Point to Point (P2P) and central server structure based on TTP [16]. In the distributed structure, location privacy protection is accomplished through collaboration between users. Chow *et al.* [4] proposed a P2P-based spatial anonymity method. In this method, the k-anonymous privacy protection based on distributed architecture is achieved by using location information of neighbors' node, but the security of the neighbors' node is ignored. The P2P-based scheme is simple and flexible, but which greatly increases various overhead of the smart phone. Furthermore, the users' locations are mobile rather than static. In centralized structure based on TTP, a method of location privacy protection based on TTP is proposed by Xie *et al.* [18]. This structure model has a good effect of privacy protection, which is currently the primary choice for location privacy protection. Li *et al.* [13] proposed a location privacy protection scheme based on efficient information cache, which reduces the number of times that the users' access to TTP. In this method, the query efficiency is improved, and the probability of information leakage is reduced, but the burden of the mobile client is increased.

In addition, Cheng *et al.* [3] put forward an independent structure model, and users' location privacy is protected according to their own abilities and knowledge. The structure of this method is simple, which is easy to merge with other structures, but it requires high performance for mobile clients. Li *et al.* [12] put forward a multi-server architecture, users can be divided into different subsets according to the security requirements, and each location server can only obtain partial subset. The concealment of location is improved in this method, but it is mainly suitable for the social network. Li *et al.* [14] put forward a location privacy protection method based on privacy information retrieval, and its location privacy protection is implemented by using retrieval and encryption. The location privacy is well protected in this method, but the overhead of communication and hardware is increased, and the query quality is reduced. With the maturity and popularity of cloud service technology, Yuji *et al.* [24] proposed a location privacy protection method based on searchable encryption. By accessing to the cloud server in the encrypted state, the security of location data and query records is guaranteed, but query efficiency and query accuracy need to be improved further.

In recent research, k-anonymous [25] is still the mainstream method of location privacy protection, which was born in the relational database, and its key attribute is dealt with using generalization and fuzzy technology. So none of the records can be distinguished from other k-1 records, and the location anonymity is realized. The method of k-anonymity location privacy protection is mainly divided into spatial region anonymity and dummy anonymity. Gruteser *et al.* [7] proposed a k-anonymity location privacy protection method, and its location privacy is protected by constructing k-anonymous region. The region must meet two conditions: 1) The area of the region reaches a certain value; 2) There are k users in the region. Due to the above two limitations, the effect of location privacy protection is improved, but all users must have the same location anonymity requirement.

Gedik *et al.* [6] put forward the location k-anonymity method to meet the user's personalized privacy requirements. The user can define the k value and anonymous level to realize personalized anonymity, but the actual effect is poor when the k value is too large. Lu *et al.* [15] have designed the k-anonymous method to add dummy locations by using circular or rectangular regions, but too many randomized locations are easily recognized by adversaries. Yin *et al.* [23] proposed an improved k-anonymity method. By setting the range of k parameters, the combination of pseudonyms and anonymity was used to improve the privacy protection effect, but the density of the anonymous region was not considered. Dewri *et al.* [5] adopts dummy location instead of user's current location to send query requests, but the accuracy of query results is low, and the extra communication cost of LBS server is increased. Jia *et al.* [10] put forward a method of combining k-anonymity and encryption technology, double protection is achieved via encrypted user and TTP, but the communication cost was relatively large.

In the research of k-anonymous region construction, Bamba *et al.* [1] put forward the method of Grid partition. In this method, there were two algorithms for Top-Down Grid Cloaking and Bottom-Up Grid, which were available for different privacy requirements. Xu *et al.* [20]

proved that the size of k-anonymous region has great influence on the accuracy of query results, which provided guidance for the research of division of anonymous regions. On this basis, Zhao *et al.* [26] proposed a method of circle anonymous region division, and Yang *et al.* [21] proposed an augmented reality rectangle partition anonymous method. These methods divided the whole area into a combination of some geometric shapes to achieve privacy protection, and further reduced the area of anonymous regions. In order to make the number of users to meet the privacy protection requirements, it is usually realized by enlarging or reconstructing the region. But in k-anonymous location privacy protection method, k-density and area parameters need to be fully considered, and the best anonymous region is generated according to the terrain characteristics.

The above methods solve the problems of LBS privacy protection from different angles, and different ways to construct anonymous regions are proposed. But in practical applications, the shapes of these anonymous regions are often influenced by the terrain such as desert, high mountains and river. Anonymous regions are not regular geometric shapes, such as circles, rectangles, etc. These methods increased the area of invalid areas, such as circles, rectangles, grids and other regular geometric shapes, and increased the computing cost of the server. Moreover, adversaries can easily analyze and identify the users'location based on invalid region and terrain features. Then, Xie *et al.* [19] proposed a k-anonymity algorithm of irregular polygon. By constructing an irregular polygonal anonymous region, the area of the invalid region is reduced. However, it takes much time to generate anonymous region of polygon, and query quality is reduced. Moreover, this method only considers the Euclidean space distance, but not the users' density distribution and the diversity of the environment.

Therefore, in the k-anonymous location privacy protection method, it is necessary to fully consider the users' density distribution and the terrain features. Based on the above analysis, a k-anonymous location privacy protection method of polygon based on density distribution is proposed. According to the density distribution of users, the method combining k-anonymity and dummy location is adopted to further improve the privacy protection effect and query quality.

3 System Model

3.1 System Structure

In LBS service, the most widely used is to query the nearest interest point. If users want to know the nearest shopping mall, hotel, gas station, hospital, etc., they need to send their current location to the LBS server. However, LBS server is not reliable, and users' location information will be leaked to third parties intentionally or unintentionally, which will lead to privacy leakage. In the TTP-based structure, when a user needs to obtain location service, who do not send their location to LBS server directly, but first send query request to TTP. The query request will be sent to LBS server after anonymous processing by TTP. The system structure model is shown in Figure 1.

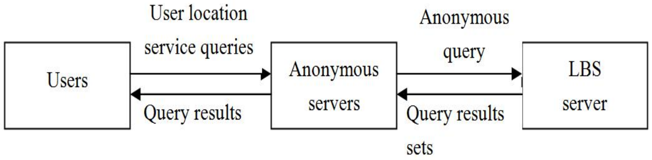


Figure 1: System structure model

In practical applications, the users hope that anonymous region can best match the actual terrain, such as street trend, bridge shape, shopping mall shape, etc., as shown in Figure 2, which can better meet users' privacy requirements and improves the accuracy of query result. In this paper, according to the geographical features of the user's location, the anonymous region of irregular polygon is constructed, as shown in Figure 3. Adopting the polygon boundary fast construction algorithm, the polygon anonymous region is generated quickly, which improves efficiency of anonymous region generation. Based on the density of users, the strategy of spatial region anonymity and dummy location is adopted, and the effectiveness of location privacy protection is further improved.

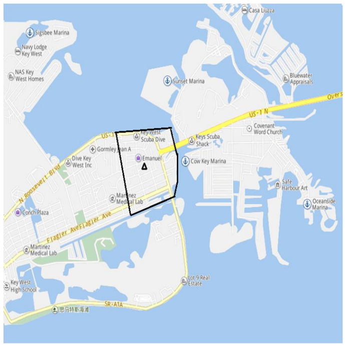


Figure 2: Effect of terrain on structuring anonymous region

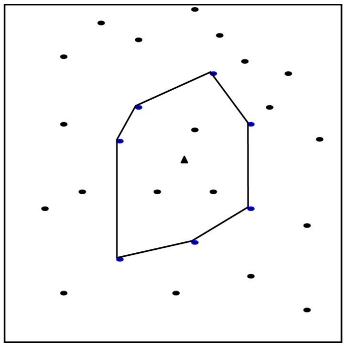


Figure 3: Polygon regional construction block diagram

3.2 Definition

Definition 1. *Let Rs represents an irregular polygonal anonymous region. Rs can be defined as* $Rs = \{U_{id}, (x_u, y_u, k\}$*, Among them,* U_{id} *represents the user's identity information; Let* (x_u, y_u) *represents the user's location coordinates:* x_u *represents the longitude of the location,* y_u *represents the latitude of the location; let k represents the anonymous parameter specified by the user.*

Definition 2. *Let* ρ *represents the density of users in the Rs region, and set threshold parameters* ρ_{max} *and* ρ_{min} *for it. Among them,* ρ_{max} *represents the maximum density, and* ρ_{min} *represents the minimum density..*

Definition 3. *Let* $S(Rs)$ *represents the area of the anonymous region. Let* S_{min} *represents the minimum area that users can accept, and* S_{max} *represents the maximum area that users can accept.*

Definition 4. *Let* $N(Rs)$ *represents the number of users in the Rs region. It can achieve the best anonymous effect when* $N(Rs) = k$*.* $N(Rs)$ *is an important parameter index, determining the degree of anonymity and the size of* $S(Rs)$ *in the system.*

4 Algorithmic Description

In this paper, a k-anonymous location privacy protection method of polygon based on density distribution is proposed. When a user queries the location, a polygon region including k locations is generated in the current area, and the k-density in the polygon region is calculated. If the k-density meets the set threshold, the polygon region with the geometric center of the polygon area as the anchor point is sent to LBS server for query. If the density is larger than the maximum threshold, the area of the polygon will be further expanded, and then the polygon region with the geometric center of the polygon area as the anchor point is sent to LBS server for query. If the density is less than the minimum threshold, a number of dummy locations are added in the polygon region, and then k locations (including dummy locations and neighbor users' locations) are sent to LBS for query.

The proposed method is realized by the following two algorithms: Algorithm 1 quickly generates an irregular polygon k anonymous region according to the coordinates of the query user and the neighbors. Algorithm 2 calculates the density of the users in the anonymous region, and adopts the corresponding anonymous strategy according to the density parameter threshold. The two algorithms are described as follows.

4.1 Algorithm 1

The principle is realized by using double-end queue: let D is a double-end queue, and all the operations of D are described in terms of that to enter the tail of queue, to go out of the tail of queue, to enter the head of queue, to go out of the head of queue.

Algorithm 1: Constructing a k-anonymous region of polygon.

Input: User's coordinates (x_u, y_u), requirement parameter k.

Output: Generate a polygonal anonymous region containing k locations.

Step 1: $n = 1$, $(x_u, y_u) = 0$.

Step 2: Set a location position near the user, take this position as the center, gradually scan the k locations, and record the coordinates of each point with (x_i, y_i).

Step 3: Select the point with minimum x-coordinate from the coordinates (x_i, y_i). If there are many points that satisfy this condition, then the point with minimum y-coordinate is selected, and the point is recorded as P_0.

Step 4: Select one direction against clockwise direction. P_x represents an arbitrary point, calculate the angle between $\overrightarrow{P_0 P_x}$ and the negative direction of y axis. Here $\overrightarrow{P_0 P_x}$ is the vector between P_0 and P_x.

Step 5: According to the angle calculated from Step 4, sort all the points from small to large, then get an ordered set $C = P_0, P_1, P_2, \cdots, P_{n-1}$.

Step 6: Remember at a certain time, the state of double-end queue D is $C = P_t, P_{t-1}, \cdots, P_0, \cdots, P_{b-1}, P_b$, traversing every point in the C:

1) If the point is P_0, then P_0 enters the tail of the queue firstly; if the point is P_1, then P_1 enters the tail of the queue; if the point is P_2, then P_2 enters the tail of the queue, and also the head of the queue.

2) Suppose that the current point P_i is traversed.

(1) If $P_{b-1}P_bP_i$ can keep the left-turn characteristics, then continue, otherwise P_b goes out of the tail of the queue; so repeat until $P_{b-m-1}P_{b-m}P_i$ can meet the left-turn characteristic, and P_i enters the tail of the queue.

(2) If $P_iP_tP_{t-1}$ can keep the left-turn characteristic, then continue, otherwise P_t goes out of the head of the queue, so repeat until $P_iP_{t-n}P_{t-n-1}$ can meet left-turn characteristic, and P_i enters the head of the queue.

Step 7: Returns the double-ended queue.

Step 8: The polygon k-anonymous region Rs is constructed.

4.2 Algorithm 2

Algorithm 2: Generating a k-anonymous result set.

Input: The k-anonymous region Rs.

Output: A k-anonymous result set.

Step 1: Set the maximum (ρ_{max}) and minimum (ρ_{min}) of the k-density.

Step 2: Take all vertices of the polygon from the double-end queue.

Step 3: Calculate the area $S(Rs)$ of the polygon region.

Step 4: Calculate S_{max} and S_{min} according to (ρ_{max}) and (ρ_{min}) .

Step 5: Judge:

1) If $S(Rs)<S_{min}$, then $k \leftarrow k+1$, execute Algorithms 1 and 2 in turn, then Step 6;

2) If $S_{min}<S(Rs)<S_{max}$, then Step 6;

3) If $S(Rs)>S_{max}$, execute Algorithms 1 and 2 in turn, then Step 7.

Step 6: Calculate the coordinate of the center location, then take the geometric center as the anchor point, and send the k-anonymous region of polygon to the LBS server for query.

Step 7: Add $[k - N(Rs)]$ dummy locations to the region randomly, and then send k locations including $N(Rs)$ users' location and $[k - N(Rs)]$ dummy locations to the LBS server for query.

4.3 Algorithm 1 Description

N location points are obtained by scanning around the query user, one of its with the minimum x-coordinate is picked. If a point with the minimum x-coordinate is not unique, a point with the minimum y-coordinate is picked. This point is defined as $P_0(x_0, y_0)$, and clockwise is selected as the default direction. The angle is calculated between $\overrightarrow{P_0P_x}$ and the negative direction of y axis, here $\overrightarrow{P_0P_x}$ is the vector between P_0 and P_x. By sorting all the points from small to large, an ordered set $C = P_0, P_1, P_2, \cdots, P_{n-1}$ is obtained. According to the following methods, all the outermost points in the set are found.

Assuming that P_i, P_j, P_k are three consecutive points on the boundary of the region (polygon vertex), its must maintain a trend of left-turn, that is $\overrightarrow{P_iP_j} \times \overrightarrow{P_jP_k} > 0$. If the three points are represented as (x_i, y_i), (x_j, y_j), (x_k, y_k), there are: $\overrightarrow{P_iP_j} = (x_j - x_i, y_j - y_i)$, $\overrightarrow{P_jP_k} = (x_k - x_j, y_k - y_j)$.

According to Algorithm 1, a k-anonymous region of polygon containing k locations is generated, as shown in Figure 4. The solid triangle symbol represents the current location of the user, and an irregular polygonal anonymous region which consists of 16 location positions is constructed.

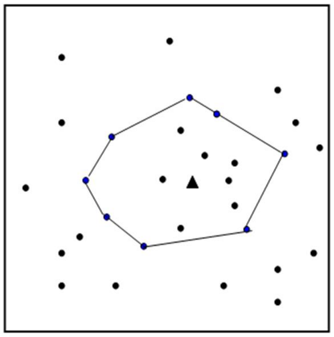


Figure 4: Block diagram of polygon region generation

4.4 Algorithm 2 Description

According to Algorithm 1, the coordinates of all vertices of convex polygon are obtained.

Suppose the n vertices on a convex polygon are ordered counterclockwise as $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $\cdots$, $P_n(x_n, y_n)$, then the area of the polygon is:

$$S_n = \frac{1}{2}\sum_{i=1}^{n-1}(x_i y_{i+1} - x_{i+1} y_i) + \frac{1}{2}(x_n y_1 - x_1 y_n) \qquad (1)$$

The area $S(Rs)$ of the polygon k-anonymous region is calculated by Equation (1). The maximum area (S_{max}) is calculated by $S_{max} = k/\rho_{min}$, and the minimum area (S_{min}) is calculated by $S_{min} = k/\rho_{max}$.

Then judge by (S_{max}) and (S_{min}):

1) If $S(Rs)<S_{min}$, the anonymous region needs to be further expanded, and then the method of spatial anonymity is used to protect location privacy.

2) If $S_{min}<S(Rs)<S_{max}$, the anonymous region meets the user's requirements, and then the method of spatial anonymity is used to protect location privacy.

3) If $S(Rs)>Smax$, the method of spatial anonymity is invalid, the location privacy protection is implemented by combining spatial anonymity and dummy locations.

4.5 Algorithm Analysis

In this paper, an irregular polygon k-anonymous region including the user's current location is quickly constructed. And then the area of the polygon anonymous region is calculated. The size of the polygon region not only affects effect of the location privacy protection, but also affects the quality of the location service. Therefore, the area threshold needs to be set, so that the size of the anonymous region is kept in a suitable range. The influence of the $S(Rs)$ on the system anonymity is as follows:

when $S(Rs) < S_{min}$, the anonymous region is too small and the range of the region is close to the exact location of the user. In this case, an adversary is very easy to inference the location of the user; when $S(Rs) > S_{max}$, the anonymous region is too large, which reduces the accuracy of query results and consumes too much resources. Therefore, in the construction of anonymous regions, S_{max} and S_{min} need to be set beforehand.

In the anonymous region, it is known from $\rho = N(Rs)/S(Rs)$ that ρ is proportional to $S(Rs)$. Therefore, the density threshold is determined, and the area threshold is determined accordingly, that is $S_{max} = k/\rho_{min}$, $S_{min} = k/\rho_{max}$. In the k-anonymous region, when the k-density is too large, it indicates that the current location is in densely populated region such as schools, hospitals, stations, churches, etc. In this case, although the k value meets the anonymity requirement, the adversary can easily obtain the user's exact location. When the k-density is too small, it indicates that the current location is in a region where few people are in that such as desert, lake, mountain, etc. In this case, the spatial anonymity method is invalid. Therefore, the area threshold of anonymous region is determined by density, and different anonymity strategies are adopted according to area threshold, which can better improve the anonymity effect.

In the anonymity processing, there are three cases according to the area threshold:

1) $S(Rs) < S_{min}$;

2) $S_{min} < S(Rs) < S_{max}$;

3) $S(Rs) > S_{max}$.

In Cases 1 and 2, spatial anonymity is used to protect location privacy. At the same time, in order to improve the accuracy of query, the central node is used as the anchor point for query. When the query result set is returned, the user can calculate the exact query result according to the distance between the current location and anchor. The central node in this algorithm is represented by $O(x_0, y_0)$, its coordinate is calculated by Equation (2). The distance between the user's current location and anchor is calculated by Equation (3), which can be used as the measure of the accuracy of query result.

$$\begin{cases} x_0 = \frac{x_1+x_2+\cdots+x_n}{n} \\ y_0 = \frac{y_1+y_2+\cdots+y_n}{n} \end{cases} \quad (2)$$

$$d = \sqrt{(x_u - x_0)^2 + (y_u - y_0)^2} \quad (3)$$

In Case 3, the method of spatial anonymity is invalid, the location privacy protection is implemented by combining spatial anonymity and dummy locations, which effectively remedies the shortcoming of spatial anonymity method. Moreover, in the selection of dummy locations, the queried neighbor users are regarded as part of the dummy locations, which further improves the indistinguishability between dummy locations and the current location.

5 Experimental Results and Analysis

In this paper, a network-based mobile node generator [2] developed by Thomas Brinkhoff, which is used to generate 1000 data nodes distributed in the whole area through a real map. The hardware environment of the experiment is as follows: 3.2 GHz Intel Core i5 processor with memory size of 4 GB. The operating system is Windows 7. The proposed algorithm is implemented by Eclipse development platform and Java programming language. Table 1 is configured for the default parameters of the experiment.

Table 1: Experimental default parameter configuration

Parameter	Value
k	[0, 100]
ρ_{min}	0.002
ρ_{max}	0.02
Number of users	[0, 1000]
Space range (km^2)	0.8×0.8

5.1 Comparison of Anonymous Time

Firstly, efficiency of the proposed method is verified by experiments. In Figure 5, we compare the anonymous region generation time with the polygon method [19] and the proposed method. As shown in Figure 5, with the increase of k, the anonymous region generation time of both methods is increasing, and its growth trends are roughly the same. As can be seen from Figure 5, the polygon partition method takes much more time than the proposed method. From the experimental result we can see that the proposed method has better efficiency.

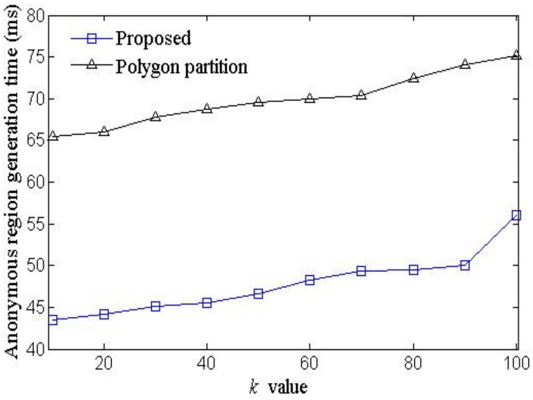


Figure 5: Average generation time of dummy

The result in Figure 5 is the best way. However, in the process of anonymous region generation, when the k value is insufficient, both methods need to repeat the algorithm several times. In addition, the proposed method needs to calculate the area and density of the anonymous region, and takes different anonymity measures according to the density threshold. When the algorithm is executed many times, the time taken for both methods is shown in Figure 6 and Figure 7.

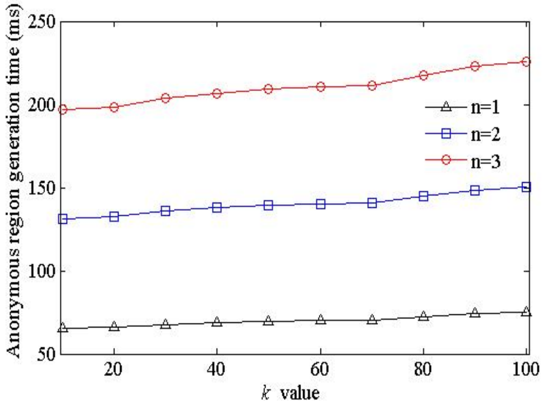


Figure 6: Average generation time of dummy

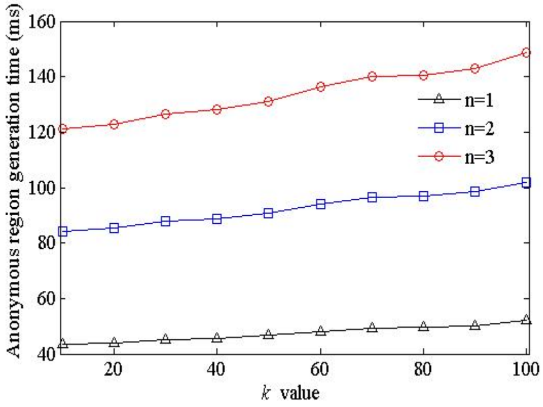


Figure 7: Average generation time of dummy

As can be seen from Figure 6 and Figure 7, when the first round of execution fails to meet the requirements, the second round and the third round will be executed. In contrast, the more the number of execution rounds, the greater the time gap between the two methods, the more obvious the efficiency advantage of the proposed.

5.2 Comparison of Anonymous Area

In the same environment, we compare the area of the anonymous with the grid partition method [1], the circular partition method [26], the rectangle partition method [21], the polygon partition method [19] and the proposed method, as shown in Figure 8.

As we can see from Figure 8, the area of five anonymous region construction methods increases with the increase of k, but the growth rates vary. This is determined by the geometric shape of the above methods. When k is the same, the area of the two polygon methods is the smallest.

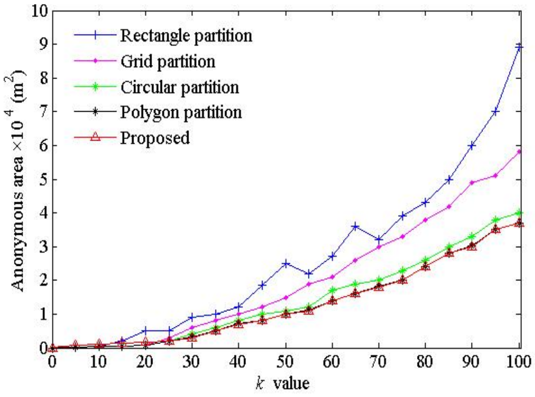
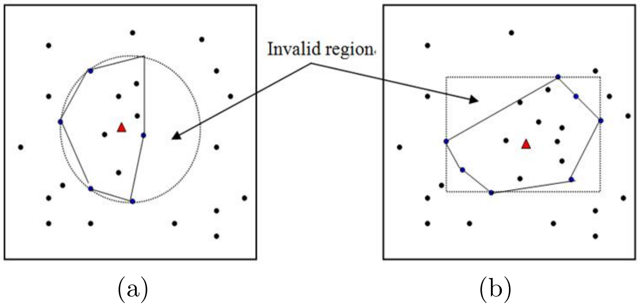


Figure 8: The area of several anonymous regional division methods

As we can see from Figure 8, when $k<30$, the area of the proposed method is slightly larger than that of the polygon partition method; When $k>30$, the area growth trend of the two methods is identical. This is because the density threshold is set in the proposed method. When $k<30$, the polygon region is expanded because it does not meet the anonymity requirement.

5.3 Analysis of Efficiency

In this paper, spatial anonymity is achieved by constructing a polygonal anonymous region. We compare the result of anonymous region construction with circular, rectangular, and polygon, as shown in Figure 9. As can be seen from Figure 9, comparing with the method of polygon construction, the methods of circular and rectangular construction enlarge the area of invalid region, its further reduce the accuracy of query result and privacy protection effect.



(a)

(b)

Figure 9: The results of anonymous region construction; (a) Invalid region of circle, (b) Invalid region of rectangle

In other methods of spatial anonymity, if the area of anonymous region is larger than the maximum area threshold, the method is invalid. In the proposed method, if the area of the polygon anonymous region is larger than the maximum area threshold, the polygon anonymous region is expanded further, and $[k - N(Rs)]$ dummy locations are added to the polygon region. As shown in Figure 10, solid dots represent the neighbor users' locations found, hollow dots represent the added dummy locations, and solid triangle represents the user's current location. K locations including users' locations and dummy locations are sent to LBS server by TTP for query. It is difficult for an adversary to distinguish the user's current location from other k-1 locations. The proposed method is effective.

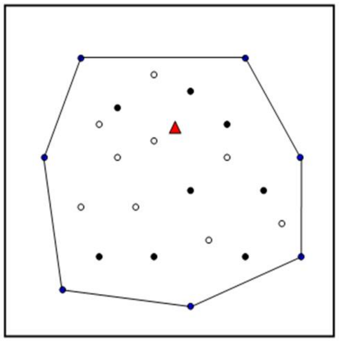


Figure 10: Combination of spatial anonymity and dummy

5.4 Comparison of Entropy

In location privacy protection method of dummy, entropy is usually used to measure effect of the location privacy protection. From the adversary's view, the anonymous set contains user's current location and k-1 dummy locations, and the probability that any location can be used as user current location is p_i. In an anonymous set, the sum of all probabilities is $\sum p_i$. Therefore, the entropy H for distinguishing the user current location in the candidate set is:

$$H = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i \tag{4}$$

In Equation (4), if all the k locations of the candidate set have the same probability, the maximum entropy will be obtained. At this time, the probability of p_i is $1/k$, and the maximum entropy H is $\log_2 k$.

In Figure 11, we compare the entropy with the proposed and other three methods. Random is the method that selects dummy locations at random. Circular dummy and grid dummy are the virtual circle and virtual grid proposed in [15].

As can be seen from Figure 11, entropy of the proposed method is larger than that of the other methods. This is because k-1 dummy locations are all added randomly besides user's current location in the method of grid dummy and circular dummy. These dummy locations are easily distinguishable from the user's current location. In the proposed method, $N(Rs)$ locations are the neighbor users, its are indistinguishable from the user's current location. The remaining $[k - N(Rs)]$ locations are the dummy locations that is added, its are less indistinguishable from the user's current location. So the entropy of the proposed method is larger, and its effect of the privacy protection is better.

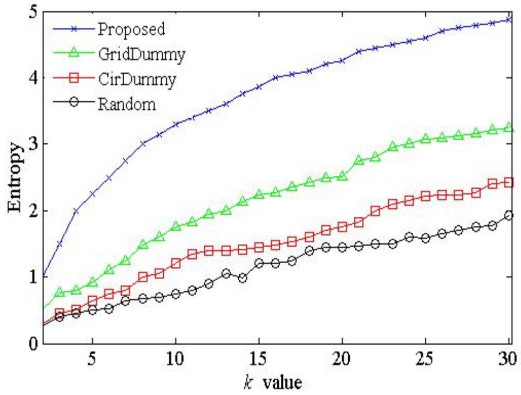


Figure 11: Entropy of the dummy locations

6 Conclusions

In recent years, the application and development of LBS are very fast, the security challenges of location privacy are becoming more and more serious. Location privacy protection has become a research hot spot in the field of information security. In the current widely used model of central server structure, aiming at the deficiency of spatial anonymity method, a k-anonymous location privacy protection method of polygon based on density is proposed. In this paper, according to the idea of k-anonymity, and adapting irregular polygon fast generation algorithm, a polygon anonymous region is constructed quickly, which improves the efficiency of anonymous region generation. At the same time, the area of polygon region is calculated through recursive method. According to the k-density distribution, an ideal and effective anonymous region is constructed. Furthermore, the privacy protection is implemented by combining spatial anonymity and dummy locations according to density parameters. And we evaluate our algorithms through a series of simulations, which show that our algorithm effectively improves the anonymity effect while taking into account query quality.

Acknowledgments

This work is supported by the National Science Foundation of China (No. 61363078, 61762059), Natural Science Foundation of Gansu (20JR5RE634), Innovation Ability Improvement Project of Colleges and Universities in Gansu (2020A-246), and Natural Science Foundation of Tianshui (2020-FZJHK-2664). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] B. Bamba, L. Liu, P. Pesti, and T. Wang, "Supporting anonymous location queries in mobile environments with privacy grid," in *Proceedings of the 17th International Conference on World Wide Web*, pp. 237–246, Jan. 2008.

[2] T. Brinkhoff, "A framework for generating network-based moving objects," *GeoInformatica*, vol. 6, no. 2, pp. 153–180, 2002.

[3] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," *Lecture Notes in Computer Science*, no. 4258, pp. 393–412, 2006.

[4] 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.

[5] R. Dewri, Y. Ray, and Y. Ray, "Query m-invariance: Preventing query disclosures in continuous location-based services," in *The Eleventh International Conference on Mobile Data Management*, pp. 95–104, May 2010.

[6] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2008.

[7] 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, May 2003.

[8] Y. Huang and Z. P. Cai, "Bourgeois a g. search locations safely and accurately: A location privacy protection algorithm with accurate service," vol. 103, pp. 146–156, 2018.

[9] 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 & Computer Applications*, no. 76, pp. 87–100, 2016.

[10] J. Jia and F. Zhang, "K-anonymity algorithm using encryption for location privacy protection," *International Journal of Multimedia & Ubiquitous Engineering*, vol. 10, no. 9, pp. 155–166, 2015.

[11] Y. Jung and J. Park, "An investigation of relationships among privacy concerns, affective responses, and coping behaviors in location-based services," *International Journal of Information Management*, no. 43, pp. 15–24, 2018.

[12] J. Li, H. Y. Yan, Z. L. Liu, X. F. Chen, X. Y. Huang, and D. S. Wong, "Location-sharing systems with enhanced privacy in mobile online social networks," *IEEE Systems Journal*, vol. 11, no. 99, pp. 1–10, 2015.

[13] L. Li, J. Hua, S. Wan, H. Zhu, and F. Li, "Achieving efficient location privacy protection based on cache," *Journal on Communications*, vol. 38, no. 6, pp. 148–157, 2017.

[14] L. Li, Z. J. Lv, X. H. Tong, and R. H. Shi, "A dynamic location privacy protection scheme based on cloud storage," *International Journal of Network Security*, vol. 21, no. 5, pp. 828–834, 2019.

[15] H. Lu, C. S. Jensen, and L. Y. Man, "Pad: Privacy-area aware, dummy-based location privacy in mobile services," in *The Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, pp. 16–27, Jan. 2008. DOI: 10.1145/1626536.1626540.

[16] T. H. Ma, J. Jia, Y. Xue, and Y. Tian, "Protection of location privacy for moving KNN queries in social networks," *Applied Soft Computing*, no. 66, pp. 525–532, 2018.

[17] T. Peng, Q. Liu, G. J. Wang, and Y. Xiang, "Multidimensional privacy preservation in location-based services," *Future Generation Computer Systems*, no. 93, pp. 312–326, 2019.

[18] M. B. Xie, Q. Qian, and S. Ni, "Clustering based k-anonymity algorithm for privacy preservation," *International Journal of Network Security*, vol. 19, no. 6, pp. 1062–1071, 2017.

[19] P. S. Xie, J. Guo, and Q. Wang, "A-anonymous polygon area construction method and algorithm based on LBS privacy protecting," *Journal of Information & Computational Science*, vol. 12, no. 15, pp. 5713–5724, 2015.

[20] J. Xu, X. Tang, H. Hu, and J. Du, "Privacy-conscious location-based queries in mobile environments," *IEEE Transactions on Parallel & Distributed Systems*, vol. 21, no. 3, pp. 313–326, 2010.

[21] Y. Yang and R. Wang, "Rectangular region k-anonymity location privacy protection based on LBS in augmented reality," *Journal of Nanjing Normal University(Natural science)*, vol. 39, no. 4, pp. 44–49, 2016.

[22] A. Y. Ye, L. Y. Cheng, J. F. Ma, and L. Xu, "Location privacy-preserving method of k-anonymous based on service similarity," *Journal of Communications*, vol. 35, no. 11, pp. 162–169, 2016.

[23] C. Yin, R. Sun, and J. Xi, "Location privacy protection based on improved k-value method in augmented reality on mobile devices," *Mobile Information Systems*, vol. 2017, no. 12, pp. 1–7, 2015.

[24] U. Yuji, M. Natsume, and Y. shota, "Private similarity searchable encryption for euclidean distance," *IEICE Transactions on Information and Systems*, vol. 100, no. 10, pp. 2319–2326, 2017.

[25] S. B. Zhang, X. Li, Z. Y. Tan, and T. Peng, "A caching and spatial k-anonymity driven privacy enhancement scheme in continuous location-based services," *Future Generation Computer Systems*, vol. 94, no. 2019, pp. 40–50, 2019.

[26] Z. Zhao, H. Hu, and F. Zhang, "A k-anonymous algorithm in location privacy protection based on circular zoning," *Journal of Beijing Jiaotong University*, vol. 37, no. 5, pp. 13–18, 2013.

Biography

Yong-bing Zhang is currently a Ph.D. student in Lanzhou University of Technology, and worked at school of Gansu Institute of Mechanical & Electrical Engineering. He received his master degree in electronic and communication engineering from Lanzhou University of Technology, Gansu, China, in 2015. His research interests include network and information security, privacy protection.

Qiu-yu Zhang Researcher/PhD supervisor, graduated from Gansu University of Technology in 1986, and then worked at school of computer and communication in Lanzhou University of Technology. He is vice dean of Gansu manufacturing information engineering research center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, multimedia communication technology.

Yan Yan associate professor. received her master degree in communication and information systems from Lanzhou University of Technology, Gansu, China, in 2005. She is currently a Ph.D. student in Lanzhou University of Technology. Her research interests include privacy protection, multimedia information security, uncertain information processing.

Yi-Long Jiang Professor, graduated from Shanghai Technology university in 1989, and then worked at school of Gansu Institute of Mechanical & Electrical Engineering. His research interests include embedded control technology, intelligent manufacturing intelligent, manufacturing technology.

Mo-yi Zhang associate professor. received her doctorate degree in manufacturing information systems from Lanzhou University of Technology, Gansu, China, in 2019. Her research interests include artificial intelligence, image processing and pattern recognition.