Efficient Trajectory Data Privacy Protection Scheme Based on Laplace's Differential Privacy

Ke Gu[†,‡,±], Lihao Yang[†], Yongzhi Liu[†], Bo Yin[†]
[†] School of Computer & Communication Engineering
Changsha University of Science & Technology, Changsha 410114, China
[‡] Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation
Changsha University of Science & Technology, Changsha 410114, China
[±] School of Information Science and Engineering, Central South University, Changsha 410083, China

Keywords: trajectory data, polygon, centroid, Laplace's differential privacy

Received: May 25, 2017

Now many applications of trajectory (location) data have facilitated people's daily life. However, publishing trajectory data may divulge individual sensitive information so as to influence people's normal life. On the other hand, if we cannot mine and share trajectory data information, trajectory data will lose its value to serve our society. Currently, because the records of trajectory data are discrete in database, some existing privacy protection schemes are difficult to protect trajectory data. In this paper, we propose a trajectory data privacy protection scheme based on Laplace's differential privacy mechanism. In the proposed scheme, the algorithm first selects the protected points from the user's trajectory data; secondly, the algorithm builds the polygons according to the protected points and the adjacent and high frequent accessed points selected from the accessed point database, then the algorithm calculates the polygon centroids; finally, the noises are added to the polygon centroids by the Laplace's differential privacy method, and the new polygon centroids are used to replace the protected points, and then the algorithm constructs and issues the new trajectory data. The experiments show that the running time of the proposed algorithms is fast, the privacy protection of the scheme is effective and the data usability of the scheme is higher.

Povzetek: Predlagana je metoda za učinkovito varovanje podatkov o poteh na osnovi Laplacove diferenčne privatnosti.

1 Introduction

1.1 Background

With the rapid development of computer and network, data mining and analysis plays an increasingly important role in our social life. The huge amounts of data (such as big data) can bring many application services to our society, such as trajectory (location) data, health and food data, traffic safety data, etc. Trajectory data is a kind of position information with large scale, fast changing and generally accepted characteristics, which mainly comes from vehicle networks, mobile devices, social networks and so on. Now many applications of trajectory data have facilitated people's daily life, thus trajectory data service is called as a kind of new mobile computing service. Currently, it is the key of developing trajectory data services that we must be able to learn and understand position information [1]. However, trajectory data is mainly collected and disseminated by mobile equipments, but many mobile devices and mobile communication technologies must integrate geographical data and individual information into trajectory data, such as individual information may contain individual privacy data, personal health status, social status and behavior habits, etc, thus mining and publishing trajectory data may divulge individual sensitive information so as to influence people's normal life [2,3,4].

Now it is the key of trajectory data privacy protection that how to protect sensitive trajectory data while providing trajectory information service on data mining. For example, if mined data is not processed and protected on fully open status, mined data may reveal user's privacy so as to affect user's normal life. Thus, it is double-edged sword that how to mine and use trajectory data. Namely we must find a compromising approach between service and protection. However, many existing privacy protection schemes cannot provide the balance of utility and protection. For example, the generalization method [5] cannot availably protect data, and the anonymous grouping method [6] is not efficient enough. Furthermore, because the records of trajectory data are discrete in database[1], some existing privacy protection schemes are difficult to protect trajectory data. Therefore, we focus on finding an efficient privacy protection scheme for trajectory data in this paper.

[1] In real world, trajectory data may not be discrete. In this paper, our focus is the combination of location data and accessed frequency, thus we consider that the records of trajectory data are discrete.

1.2 Our contributions

In this paper, we propose a trajectory data privacy protection scheme based on Laplace's differential privacy mechanism. In the proposed scheme, the algorithm first selects the protected points from the user's trajectory data; secondly, the algorithm builds the polygons according to the protected points and the adjacent and high frequent accessed points selected from the accessed point database, then the algorithm calculates the polygon centroids; finally, the noises are added to the polygon centroids by the Laplace' differential privacy method, and the new polygon centroids are used to replace the protected points, and then the algorithm constructs and issues the new trajectory data. The experiments show that the running time of the proposed algorithms is fast, the privacy protection of the scheme is effective and the data usability of the scheme is higher.

1.3 Outline

The rest of this paper is organized as follows. In Section 2, we discuss the related works about trajectory data privacy protection. In Section 3, we review the related definitions and theorems on which we employ. In Section 4, we propose an efficient trajectory data privacy protection scheme, which is based on the Laplace's differential privacy mechanism. In Section 5, we analyze and show the efficiency of the proposed scheme by the experiments. Finally, we draw our conclusions in Section 6.

2 Related work

Currently many privacy protection schemes are being widely used in many fields, such as secure communication, social network, data mining and so on. The works [5,6] first proposed the k-anonymity model to protect social network, whose anonymity protection methods mainly include generalization [7,8], compression, decomposition [9], replacement [10] and interference. Based on the works of [5,6], many other k-anonymous protection methods [11-21] were also proposed. However, the works [20,21,22] proved that some anonymous protection methods cannot protect sensitive data very well. Additionally, Cristofaro et al. [23] proposed a privacy-encrypted protection scheme. Although their scheme can ensure data security, data utility is decreased. Current location data privacy protection methods [1,24] are mainly classified to three categories: the heuristic privacy-measure methods, the probability-based privacy inference methods and the privacy information retrieval's methods. The heuristic privacy-measure methods [25,26,27,28] are mainly to provide the privacy protection measure for some no-high required users, such as k-anonymity [25], t-closing [26], m-invariability [27] and l-diversity [28]. Also, although the information retrieval's privacy protection methods can achieve perfect privacy protection, there are more or less privacy information in the released data, so these methods may result in that no data can be released, and these methods have high overhead. Additionally, the probability-based privacy inference methods can protect data and achieve better data utility under certain conditions, but the effectiveness of the methods depends on original data availability. Further, the three kinds of methods are based on a unified attack model [1], which depends on certain background knowledge to protect location data. However, with the increase of background knowledge got by the attackers, these methods could not always effectively protect location data. The works [5,6,11-19] showed the shortages of the relationship-privacy protection methods. Ting et al. [29] analyzed a variety of privacy threat models and tried to optimize the effectiveness of the data obtained while preventing different types of reasoning attacks. Bugra et al. [30] proposed the first effective location-privacy preserving mechanism (LPPM) that enables a designer to find the optimal LPPM for a LBS (location-based service) given user's service quality constraints against an adversary implementing the optimal inference algorithm. Such LPPM is the one that maximizes the expected distortion (error) that the optimal adversary incurs in reconstructing the actual location of a user, while fulfilling the user's service-quality requirement. Presently, it is the key of protecting location data to provide a privacy protection method not sensitively to background knowledge. Based on the requirement, differential privacy protection technology can exactly satisfy it. Differential privacy is a kind of strong privacy protection method, which is not sensitive to background knowledge. However, because location data has the characteristics of sparsity and farrago, many differential privacy protection methods are not enough efficient. He et al. [31] proposed a synthetic system based on GPS path, which can provide strong differential privacy protection mechanism. The proposed system gets different speed trajectory by using a hierarchical reference method to isolate the original trajectory, and then protects the speed trajectory. Chatzikokolakis et al. [32] proposed a predictive differentially-private mechanism for location privacy, which can offer substantial improvements over the independently applied noise. Their works showed that correlations in the trace can be in fact exploited in terms of a prediction function that tries to guess the new location based on the previously reported locations. Additionally, their works tested the quality of the predicted location using a private test; in case of success the prediction is reported otherwise the location is sanitized with new noise. Chatzikokolakis et al. [33] also showed a formal notion of privacy that protects the user's exact location–"geo-indistinguishability", and then proposed two mechanisms to protect the privacy of user when dealing with location-based services. Also they extended their mechanisms to the case of location traces, and provided a method to limit the degradation of the privacy guarantees due to the correlation between the points. Li et al. [34] proposed a compressive mechanism for differential privacy, which is based on compressed sensing theory. Their mechanism is to consider

every data as a single individual, so it undermines the relationship of data so as to be not suitable to protect location data. Jia et al. [1] proposed a differential privacy-based transaction data publishing scheme. Their method establishes the relationship of transaction data items by a query tree and adds noises to the query tree based on the compressive mechanism and the Laplace's mechanism. However, it is difficult to measure the effectiveness of their method on privacy protection. Zhang et al. [35] proposed an accurate method for mining top-k frequent data records under differential privacy. In their scheme, the exponential mechanism is used to sample top-k frequent data records, and then the Laplace's mechanism is utilized to generate noises to distort original data. Although the effectiveness of their method may accurately be measured on privacy protection, their method neglects the relationship of transaction data items.

3 Differential privacy

Differential privacy protection can achieve privacy protection target by making data distortion, where the common approach is to add noises into queried results. The purpose of differential privacy protection is to minimize privacy leakage and to maximize data utility [36,37]. Currently differential privacy protection has two main methods [38,39]—the Laplace's mechanism and the exponential mechanism.

DWork et al. [39] proposed a protection method for the sensitivity of private data, which is based on the Laplace's mechanism. Their method distorts the sensitive data by adding the Laplace's distribution noises to the original data. Their method may be described as follows: the algorithm M is the privacy protection algorithm based on the Laplace's mechanism, the set S is the noise output set of the algorithm M, and the input parameters are the data set D, the function Q, the function sensitivity ΔQ and the privacy parameter ε, where the set S approximately subjects to the Laplace's distribution ($\frac{\Delta Q}{\varepsilon}$) and the mean (zero), as shown in the formula (1):

$$\Pr\left[M(Q,D)=S\right] \propto \exp\left(\frac{\varepsilon}{\Delta Q}\cdot \mid S-Q(D)\mid_1\right) \qquad (1)$$

Also, in their method, the probability density function of added noise subjecting to the Laplace's distribution is as the formula (2):

$$\Pr(x,\lambda)=\frac{1}{2\cdot\lambda}\cdot e^{\frac{-|x|}{\lambda}} \qquad (2)$$

where $\lambda=\frac{\Delta Q}{\varepsilon}$, namely the added noise is independent from the data set, and is only related to the function sensitivity and the privacy parameter. The main idea of their method adds the noises subjecting to the Laplace's distribution into the output result so as to distort the sensitive data to achieve data protection target. For example, in their method, let $Q(D)$ be the querying function of top-k accessing count, then the output of the algorithm M can be represented by the following formula (3):

$$M(Q,D)= Q(D)+\left(Lap_1(\tfrac{\Delta Q}{\varepsilon}),\ Lap_2(\tfrac{\Delta Q}{\varepsilon}),...,\ Lap_k(\tfrac{\Delta Q}{\varepsilon})\right) \qquad (3)$$

where $Lap_i(\frac{\Delta Q}{\varepsilon})(1\le i\le k)$ is each round of the independent noise subjecting to the Laplace's distribution, and the noise is proportional to ΔQ and inversely proportional to ε.

Definition 3.1 $\varepsilon-$**Differential Privacy**: Given two adjacent data sets D and D' where at most a data record is different between D and D' ($|D\neq D'|=1$), for any algorithm M, whose range is $Range(M)$, if the result S outputted by the algorithm M satisfies the following formula (4) on the two adjacent data sets D and D' ($S\in Range(M)$), then the algorithm M satisfies $\varepsilon-$differential privacy.

$$\Pr\left[M(D)\in S\right]\le e^{\varepsilon}\cdot\Pr\left[M(D')\in S\right] \qquad (4)$$

Pr represents the randomicity of the algorithm M on D and D', namely denotes the risk probability of privacy disclosure. ε represents the privacy protection level, where if ε is bigger, then privacy protection degree is lower; on the contrary, if ε is smaller, then privacy protection degree is higher.

Definition 3.2 Data Sensitivity[2]: Data sensitivity is divided to global sensitivity and local sensitivity, we set Q as query function, then the global sensitivity of the function Q is defined as follows:

$$\Delta Q=\max_{D,D'}\{\mid Q(D)-Q(D')\mid_1\} \qquad (5)$$

where D and D' represent the adjacent data sets, $Q(D)$ represents the output of the function Q on the data set D, ΔQ is the sensitivity and represents the maximum of the outputs' difference.

Additionally, because the ε-differential privacy protection scheme may be used many times in the different stages of processing data, the ε-differential privacy protection scheme also needs to satisfy the following theorems:

Theorem 3.1 for the same data set, the whole privacy protection process is divided to the different privacy protection algorithms $(M_1,M_2,...,M_n)$, whose privacy protection levels are ε_1, ε_2,...,ε_n, so the privacy protection level $\sum_{i=1}^{n}\varepsilon_i$ of the whole process needs to satisfy differential privacy protection.

Theorem 3.2 for the disjoint data set, the whole privacy protection process is divided to the different privacy protection algorithms $(M_1,M_2,...,M_n)$, whose privacy protection levels are ε_1, ε_2,...,ε_n, so the privacy protection level $\max\{\varepsilon_i\}$ of the whole process needs to satisfy differential privacy protection.

[2]Differential privacy protection is to add noises to protect data, if data sensitivity is small, then it can effectively protect data while a small quantity of noises are added into original data; on the contrary, if data sensitivity is big, then a lot of noises need to be added into original data.

4 Trajectory data privacy protection scheme

In the section, we propose a trajectory data privacy protection scheme, which employs the Laplace's differential privacy method to protect the user's trajectory data. In the proposed scheme, the algorithm first selects the protected points from the user's trajectory data; secondly, the algorithm builds the polygons according to the protected points and the adjacent and high frequent accessed points selected from the accessed point database, then the algorithm calculates the polygon centroids; finally, the noises are added to the polygon centroids by the Laplace's differential privacy method, and the polygon centroids are used to replace the protected points, and then the algorithm constructs and issues the new trajectory data. The procedure of the proposed scheme is described as follows:

(1) Input the trajectory data I, the related and historic point data set D[3], the radius r and the differential privacy protection parameters ε and min_count[4];

(2) Select the protected point set A from the trajectory data I, then select the point data $f \in A$ and its corresponding adjacent points from D, where the adjacent points belong to the range of a circle that f is the center of the circle and r is the corresponding radius, and the frequent accessed counts of the adjacent points are no less than min_count, finally form the point set B;

(3) Traverse the set B, and build the corresponding polygons according to the points f and its corresponding adjacent points from B, where only one point in every polygon belongs to the trajectory data I, and then calculate the corresponding polygon centroids, and form the polygon centroid set J, where $j_i(x, y) \in J$ is the polygon centroid (see Section 4.2 for more details);

(4) Use the Laplace's mechanism to add the noises $Lap(\frac{k \cdot \Delta Q}{\varepsilon})$ into the set J, where the noises are added into the polygon centroids, and then generate the set G (see Section 4.3 for more details);

(5) Use the modified polygon centroids from G to replace the correspondingly protected points $f \in A$, and then issue the new trajectory data I'.

4.1 Processing trajectory data

The section describes how to select the related data from the trajectory data I and the related and historic point data set D. The proposed algorithm selects the protected point data $f \in A$ and its adjacent points from D. Figure 1 shows the procedure of selecting the related data. In Figure 1,

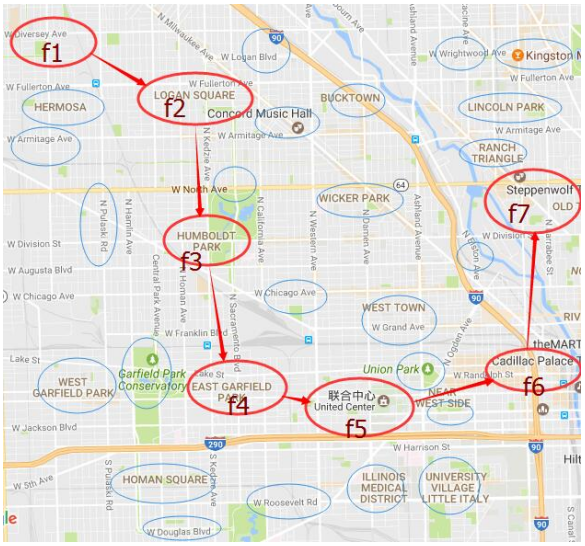

Figure 1 Processing Trajectory Data

a random trajectory of one user is shown, where the red circles and the red arrows are used to show the trajectory, and the green circles denote the accessed historic location points[5], which build the related and historic point data[6] set D. According to the Figure 1, the procedure of selecting the related data may be described as follows:

- The proposed algorithm inputs the trajectory data I of one user, the related and historic point data set D and the related privacy protection parameters r, ε and min_count;
- The algorithm selects the protected point set A from the trajectory data I;
- The proposed algorithm forms the point set B according to the point data $f_i \in A$ and its corresponding adjacent points from D, where the adjacent points belong to the range of a circle that f is the center of the circle and r is the corresponding radius, and the frequent accessed counts of the adjacent points are no less than min_count.

4.2 Building polygon model

The section describes how to build the polygon model to compute the polygon centroid. The proposed algorithm builds the polygons according to the protected points $f \in A$ and the corresponding adjacent points from D. Figure 2 shows the procedure of building polygon.

In Figure 2, the trajectory of one user is $f_1, f_2, ...f_5 \in I$, and the points $h_1, h_2, ...h_{13}$ with accessed counts come from D, where $f_2, f_4 \in A$ are the protected points.

[3]The related and historic point data include the historic location points accessed by people and the corresponding accessed counts. To the trajectory data, we may save the historic trajectory data and the related information (including accessed time and accessed count) to the database, and then the data may be classified to statistically form the set D.

[4]Our proposed scheme focuses on highly frequent accessed location data so as to distort attacker's target. So, the setting of min_count is to improve the efficiency of the proposed scheme.

[5]The adjacent point data may be related to other users.

[6]The historic duration is within one month.

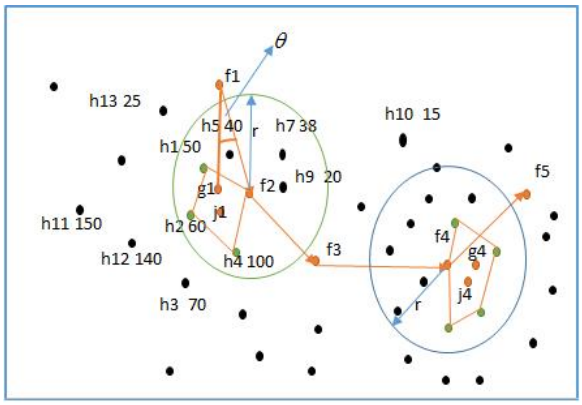

Figure 2 Building Polygon Model

In the green circle that f_2 is the center of the circle and r is the corresponding radius, the points h_1, h_2 and h_4 ($\in D$ and their accessed counts ≥ 50) and the point f_2 are used to form a polygon. Then the proposed algorithm computes the polygon centroid j_1 (noises are added to j_1 to generate a new point g_1). Similarly, the algorithm may traverse the set B to build the polygons. We need to remark that the points h_1, h_2 and h_4 is nearby the point f_2, thus the points may be used to build the polygon so as to maintain the usability of the modified trajectory, and that we set min_count is 50, thus some points whose accessed counts are less than 50 are not used to build the polygon in the green circle, such may distort the attacker's target and improve the efficiency of the proposed scheme. The procedure of building polygon model may be described as follows:

- The algorithm traverses the set B, and then selects the relevant and max-sized points to build the polygons according to the distance. For example, to a potential polygon, the algorithm selects N points as vertices from B whose coordinates are $P(x_i, y_i)$ with $i = 1, 2, 3.....N$, where one of the N points is in the original trajectory, and the other points are nearby the point;

- The algorithm computes the polygon centroids according to the vertices of the formed polygons. The formulas is described as follows:

$$j_i.x = \frac{\sum_{k=1}^{|P_i|} P_i.x_k}{n},\ j_i.y = \frac{\sum_{k=1}^{|P_i|} P_i.y_k}{n}.$$

  where $P_i(x_k, y_k)$ is the coordinate of the k_th vertices of the i_th polygon, $|P_i|$ is the vertices number of the i_th polygon, and $j_i(x, y)$ is the coordinate of the i_th polygon centroid.

- The polygon centroids are formed to the set J, where $j_i(x, y) \in J$.

4.3 Adding noises based on the Laplace's mechanism

In the section, we show how to use the Laplace's mechanism to add the noises $Lap(\frac{k \cdot \Delta Q}{\varepsilon})$[7] into the set J. The main steps of the algorithm are described as follows:

- Input the privacy protection level ε and the polygon centroid set J, and then generate the noise $Lap(\frac{k \cdot \Delta Q}{\varepsilon})$ satisfying the probability $\Pr(j(x, y), \lambda)$, where

$$\Pr(j(x, y), \lambda) = \frac{1}{2 \cdot \lambda} \cdot e^{\frac{-|j(x,y)|}{\lambda}}.$$

  In the above formula, the variant $j(x, y)$ denotes the corresponding coordinate of the polygon centroid and $\lambda = \frac{k \cdot \Delta Q}{\varepsilon}$.

- Add the noises $Lap(\frac{k \cdot \Delta Q}{\varepsilon})$ into the set J so as to disturb the polygon centroids[8]:

$$j_i.x = j_i.x \pm Lap(\frac{k \cdot \Delta Q}{\varepsilon}),$$
$$j_i.y = j_i.y \pm Lap(\frac{k \cdot \Delta Q}{\varepsilon}),$$

  where $j_i \in J$, $j_i(x, y)$ denotes the coordinate of the i_th polygon centroid, and $Lap(\frac{k \cdot \Delta Q}{\varepsilon})$ is each round of the independent noise subjecting to the probability $\Pr(j(x, y), \lambda)$. Finally, the algorithm generates the set G.

- Use the modified polygon centroids from G to replace the correspondingly protected points $f \in A$, and then issue the new trajectory data I'. For example, as the Figure 2 shown, the noise is added to j_1 to generate a new point g_1, and then g_1 is used to replace the point f_2, thus the original trajectory $f_1 \Rightarrow f_2 \Rightarrow f_3$ changes to $f_1 \Rightarrow g_1 \Rightarrow f_3$.

5 Experiment and efficiency analysis of the proposed scheme

In the section, our experiments are mainly from two aspects to evaluate the efficiency of the proposed scheme: the first one is the running time of the proposed algorithms, namely the time of extracting the available data; the second one is the effectiveness of the proposed algorithms, whose indexes include the trajectory deviation rate and the trajectory accurate rate. The test original data set comes from the simulation on the Baidu map[9], which is similar to the Gowalla

[7] ΔQ is the sensitivity of the query function Q, where we set $\Delta Q = \max\{\sqrt{(P_i.x_k - j_i.x)^2 + (P_i.y_k - j_i.y)^2}\}$ with $i = 1, 2, ...|N_P|$ and $k = 1, 2, ...|P_i|$, $|N_P|$ is the number of the polygons and $|P_i|$ is the number of the vertices of every polygon.

[8] If the formed polygon is on the left of the protected point from the trajectory data I, then the operation " $+$ " is used; otherwise, the formed polygon is on the right of the protected point from the trajectory data I, then the operation " $-$ " is used.

[9] Baidu is a network company in China. The baidu map is one of the network services provided by the company, which provides a lot of APIs for programmers to develop their applications on the map.

data set[10]. The test original data set contains user_id, accessed time, longitude and latitude and so on. The period of the test original data set is about one month. All proposed algorithms are coded by C++ and codeblocks[11]. The related parameters for the test are set as Table 1.

Table 1: Parameter Value

Parameter	Value (unit: 5 meter)
r	40,50,60,70,80,90,100,110
ε	1,2,3,4,5,6,7,8,9,10,11,12

5.1 Running time analysis

In the section, we test the running time of the proposed algorithms mainly through the time of extracting the available data, namely we test the effectiveness of computing all the polygon centroids from the available data. In the tests, when we set r=70 and ε=1,2,3,4,5,6,7,8,9,10,11,12 respectively, the time of extracting the available data is described as Table 2.

From the Table 2, we may know the time of extracting the available data is very fast, and the efficiency of computing all the polygon centroids from the available data is always increasing with the increasing of ε in a certain range.

5.2 Protection effectiveness analysis

In the section, we test the protection effectiveness of the proposed algorithms mainly through the trajectory deviation rate and the trajectory accurate rate, where the trajectory deviation rate is the angle θ formed by the modified polygon centroid and the original trajectory points, shown as Figure 3, and if the trajectory deviation rate is bigger in a certain range, then the protection effectiveness is higher; the trajectory accurate rate is used to test the protection effectiveness and usability of the noise-added data, and if the trajectory accurate rate is smaller in a certain range, then the usability is higher.

In the test, we compute the trajectory accurate rate through the following methods: 1) set the coordinate (a_i, b_i) of the polygon centroid; 2) compute the hypotenuse $c_i = \sqrt{a_i^2 + b_i^2}$; 3) compute the accurate rate $Z = \mid 1 - \frac{c_i'}{c_i} \mid$, where c_i is the original hypotenuse and c_i' is the noise-added hypotenuse. The trajectory deviation rate is bigger in a certain range, the protection effectiveness is higher; the trajectory accurate rate is smaller in a certain range, the usability is higher. So, when we set $\varepsilon = 5, 10, 15$ and $r = 40, 50, 60, 70, 80, 90, 100, 110$ respectively, Table 3,4,5 show the deviation rate and accurate rate of the trajectory data.

[10]Gowalla is a location-based social networking website where users share their locations by checking-in.

[11]The test environment is under Win10 OS, Intel i5 CPU 2.3Ghz and 8G RAM.

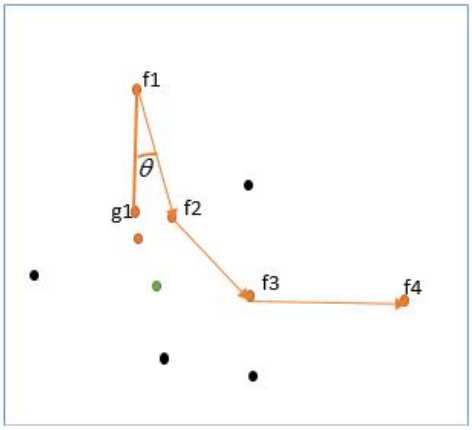

Figure 3 Trajectory Deviation Angle

From the Table 3, when $\varepsilon = 5$ and $r < 90$, we may know that the polygon centroid is not changed with the increasing of r , thus the deviation rate θ and the accurate rate Z are also not changed. Such shows that in the range of $r < 90$, the new points are not selected to build the new polygon, thus the polygon is not modified. when $r >= 90$, the new points are selected to build the new polygon, thus the polygon centroid is recomputed, thus the deviation rate θ and the accurate rate Z are changed. Such shows that the deviation rate θ could become big with the increasing of r, and the data usability becomes small. Also, from the Table 4 and the Table 5, when $\varepsilon = 10, 15$, we may get the similar results as that of the Table 3. Additionally, when we fixedly set $r = 70$ and $\varepsilon = 1, 2, 3, 4, ...15$ respectively, Table 6 shows the deviation rate and accurate rate of the trajectory data. From the Table 6, we may know that the deviation rate θ and the accurate rate Z are always increasing with the increasing of ε. That is because the constraint condition becomes small with the increasing of ε in the differential privacy mechanism. However, such also shows that the deviation rate θ becomes big so that the data usability becomes small.

6 Conclusions

Currently, because the records of trajectory data are discrete in database, some existing privacy protection schemes are difficult to protect trajectory data. In this paper, we propose a trajectory data privacy protection scheme based on Laplace's differential privacy mechanism. In the proposed scheme, the algorithm first selects the protected points from the user's trajectory data; secondly, the algorithm builds the polygons according to the protected points and the adjacent and high frequent accessed points selected from the accessed point database, then the algorithm calculates the polygon centroids; finally, the noises are added to the polygon centroids by the differential privacy method, and the new polygon centroids are used to replace the protected points, and then the algorithm constructs and issues the

Table 2: The Efficiency of Extracting Available Data

ε	1	2	3	4	5	6	7	8	9	10	11	12
Time (ms)	4	4	3	3	3	4	3	3	3	3	3	2

Table 3: Trajectory Deviation Rate And Accurate Rate ($\varepsilon = 5$)

r	c_i	c_i'	Z	θ
40	645.264	613.125	0.049807	23.2510
50	645.264	613.125	0.049807	23.2510
60	645.264	613.125	0.049807	23.2510
70	645.264	613.125	0.049807	23.2510
80	645.264	613.125	0.049807	23.2510
90	608.511	572.839	0.058621	24.7920
100	608.511	572.839	0.058621	24.7920
110	608.511	572.839	0.058621	24.7920

Table 4: Trajectory Deviation Rate And Accurate Rate ($\varepsilon = 10$)

r	c_i	c_i'	Z	θ
40	645.264	613.096	0.049852	23.2532
50	645.264	613.096	0.049852	23.2532
60	645.264	613.096	0.049852	23.2532
70	645.264	613.096	0.049852	23.2532
80	645.264	613.096	0.049852	23.2532
90	608.511	572.809	0.05867	24.7941
100	608.511	572.809	0.05867	24.7941
110	608.511	572.809	0.05867	24.7941

Table 5: Trajectory Deviation Rate And Accurate Rate ($\varepsilon = 15$)

r	c_i	c_i'	Z	θ
40	645.264	612.964	0.050057	23.2584
50	645.264	612.964	0.050057	23.2584
60	645.264	612.964	0.050057	23.2584
70	645.264	612.964	0.050057	23.2584
80	645.264	612.964	0.050057	23.2584
90	608.511	572.665	0.058908	24.7996
100	608.511	572.665	0.058908	24.7996
110	608.511	572.665	0.058908	24.7996

Table 6: Trajectory Deviation Rate And Accurate Rate ($r = 70$)

ε	c_i	c_i'	Z	θ
1	645.264	613.126	0.049806	23.25090
2	645.264	613.126	0.049806	23.25090
3	645.264	613.126	0.049806	23.25090
4	645.264	613.126	0.049806	23.25090
5	645.264	613.125	0.049807	23.2510
6	645.264	613.125	0.049807	23.2510
7	645.264	613.122	0.049812	23.2514
8	645.264	613.117	0.049819	23.2518
9	645.264	613.109	0.049833	23.2524
10	645.264	613.096	0.049852	23.2532
11	645.264	613.079	0.049879	23.2541
12	645.264	613.057	0.049913	23.2551
13	645.264	613.030	0.049954	23.2562
14	645.264	612.999	0.050003	23.2573
15	645.264	612.964	0.050057	23.2584

new trajectory data. The experiments show that the running time of the proposed algorithms is fast, the privacy protection of the scheme is effective and the data usability of the scheme is higher.

Acknowledgement

This study was funded by the National Natural Science Foundation of China (No.61402055,No.61504013), the Natural Science Foundation of Hunan Province (No.2016JJ3012).

References

[1] Ouyang Jia, Yin Jian, Liu Shaopeng, Liu Yuba. An Effective Differential Privacy Transaction Data Publication Strategy. Journal of Computer Research& development, 2014, 51(10):2195-2205. https://doi.org/10.7544/issn1000-1239.2014.20130824

[2] Loki. Available at: http://loki.com/.

[3] FireEagle. Available at: http://info.yahoo.com/privacy/us/yahoo/fireeagle/.

[4] Google latitude. Available at: http://www.google.com/latitude/apps/badge.

[5] Samarati P, Sweeney L. Generalizing data to provide anonymity when disclosing information. Proc. of the 7th ACM SIGACTSIGMOD-SIGART Symp.

on Principles of Database Systems, 1998, 188-202. https://doi.org/10.1145/275487.275508

[6] Samarati P. Protecting Respondents Identities in Microdata Release. IEEE Transactions on Knowledge and Data Engineering, 13(6):1010-1027. https://doi.org/10.1109/69.971193

[7] Fung BC, Wang K, Yu PS. Anonymizing classification data for privacy preservation. IEEE Trans on Knowledge and Data Engineering(TKDE), 2007,19(5):711-725. https://doi.org/10.1109/tkde.2007.1015

[8] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2012, 10(5):571-588. https://doi.org/10.1142/S021848850200165X

[9] Xiao X, Tao Y. Anatomy: Simple and effective privacy preservation. Proc of VLDB 2006. New York: ACM, 2006, 139-150.

[10] Zhang Q, Koudas N, Srivastava D, et al. Aggregate query answering on anonymized tables. Proc of the 23rd Int Conf on Data Engineering(ICDE). Piscataway, NJ: IEEE, 2007, 116-125. https://doi.org/10.1109/icde.2007.367857

[11] Wong RCW, Li J, Fu AWC, Wang K. (a, k)-Anonymity: An enhanced k-anonymity model for privacy-preserving data publishing. In: Proc.of the ACM 12th SIGKDD Int'l Conf on Knowledge Discovery and Data Mining, 2006, 754-759. https://doi.org/10.1145/1150402.1150499

[12] LeFevre K, DeWitt DJ, Ramakrishnan R. Mondrian multidimensional k-anonymity. Proc.of the 22nd Int'l Conf. on Data Engineering, 2006, 6(3): 25-35. https://doi.org/10.1109/ICDE.2006.101

[13] Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. L-Diversity: Privacy beyond k-anonymity. Proc.of the 22nd IEEE Int'l Conf. on Data Engineering, 2006. https://doi.org/10.1109/ICDE.2006.1

[14] Xiao X, Tao Y. Personalized privacy preservation. Proc.of the 2006 ACM SIGMOD Int'l Conf. on Management of Data, 2006, 229-240. https://doi.org/10.1145/1142473.1142500

[15] Xu J, Wang W, Pei J, Wang X, Shi B, Fu AWC. Utility-Based anonymization using local recoding. Proc.of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, 2006, 785-790. https://doi.org/10.1145/1150402.1150504

[16] Li N, Li T, Venkatasubramanian S. t-closeness: Privacy beyond k-anonymity and l-diversity. Proc.of the 23rd IEEE Int'l Conf. on Data Engineering, 2007, 106-115. https://doi.org/10.1109/icde.2007.367856

[17] Wong RCW, Fu AWC, Wang K, Pei J. Minimality attack in privacy preserving data publishing. Proc.of the 33nd Int'l Conf. on Very Large Databases, 2007, 543-554.

[18] Tao Y, Xiao X, Li J, Zhang D. On anti-corruption privacy preserving publication. Proc.of the 24th Int'l Conf. on Data Engineering, 2008, 725-734. https://doi.org/10.1109/ICDE.2008.4497481

[19] Backstrom L, Dwork C, Kleinberg J. Wherefore are thouR3579X?: Anonymized social networks, hiddern patterns and structural steganography. Proc.of the 16th Int'l Conf. on World Wide Web, 2007, 181-190. https://doi.org/10.1145/1242572.1242598

[20] Zheleva E, Getoor L. Preserving the privacy of sensitive relationships in graph data. Proc.of the 1st ACM SIGKDD Workshop on Privacy, Security, and Trust in KDD, 2007, 153-171. https://doi.org/10.1007/978-3-540-78478-4_9

[21] Korolova A, Motwani R, Nabar SU, Xu Y. Link privacy in social networks. Proc. of the 24th Int'l Conf. on Data Engineering, 2008, 1355-1357. https://doi.org/10.1109/icde.2008.4497554

[22] Cristofaro E, Soriente C, Tsudik G, et al. Hummingbird: Privacy at the time of twitter. IEEE Symposium on Security and Privacy -S&P, 2012, 285-299. https://doi.org/10.1109/sp.2012.26

[23] Beresford AR, Rice A, Skehin N, Sohan R. MockDroid: Trading privacy for application functionality on smartphones. Proc. of the 12th Workshop on Mobile Computing Systems and Applications, ACM Press, 2011, 49-54. https://doi.org/10.1145/2184489.2184500

[24] Huo Z, Meng XF. A survey of trajectory privacy-preserving techniques. Chinese Journal of Computers, 2011, 34(10):1820-1830. https://doi.org/10.3724/sp.j.1016.2011.01820

[25] Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with privacy grid. Proc. of the 17th Int'l Conf. on World Wide Web. New York: ACM Press, 2008, 237-246. https://doi.org/10.1145/1367497.1367531

[26] Liu L. From data privacy to location privacy: Models and algorithms. Proc. of the 33rd Int'l Conf. on Very Large Data Bases. New York: ACM Press, 2007, 1429-1430.

[27] Liu F, Hua KA, Cai Y. Query l-diversity in location-based services. Proc. of the 10th Int'l Conf. on Mobile Data Management. Taipei, 2009, 436-442. https://doi.org/10.1109/mdm.2009.72

[28] Ting W, Ling L. From Data Privacy to Location Privacy. Machine Learning in Cyber Trust, Berlin: Springer, 14 March 2009, pp.217-246. https://doi.org/10.1007/978-0-387-88735-7_9

[29] Bugra G, Ling L. Protecting Location Privacy. IEEE Transactions on Mobile Computing, 2008, 7(1):1-18. https://doi.org/10.1109/TMC.2007.1062

[30] He X, Cormode G, Machanavajjhala A, Procopiuc CM, Srivastava D. Differentially Private Trajectory Synthesis Using Hierarchical Reference Systems. VLDB Journal, 2015, 8(11):1154-1165. https://doi.org/10.14778/2809974.2809978

[31] Chatzikokolakis K, Palamidessi C, Stronati M. A Predictive Differentially-Private Mechanism for Location Privacy. Proc. of the 14th International Symposium on Privacy Enhancing Technologies, Berlin: Springer, 2014, LNCS 8555, pp.21-41. https://doi.org/10.1007/978-3-319-08506-7_2

[32] Chatzikokolakis K, Palamidessi C, Stronati M. Geo-indistinguishability: A Principled Approach to Location Privacy. ICDCIT 2015, Berlin: Springer, 2015, LNCS 8956, pp.49-72. https://doi.org/10.1007/978-3-319-14977-6_4

[33] Li YD, Zhang Z, Winslett M, et al. Compressive mechanism: Utilizing sparse representation in differential privacy. Proc. of the 10th Annual ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2011, pp.177-182. https://doi.org/10.1145/2046556.2046581

[34] Zhang XJ, Wang M, and Meng XF. An Accurate Method for Mining top-k Frequent Pattern Under Differential Privacy. Journal of Computer Research and Development, 2014, 51(1):104-114. https://doi.org/10.7544/issn1000-1239.2014.20130685

[35] Dwork C. The promise of differential privacy: A tutorial on algorithmic techniques. Proc. of the Foundations of Computer Science (FOCS). Piscataway, NJ: IEEE, 2011, pp.1-2. https://doi.org/10.1109/focs.2011.88

[36] Dwork C. A firm foundation for private data analysis. Communications of the ACM, 2011, 54(1):86-95. https://doi.org/10.1145/1866739.1866758

[37] Mcsherry F, Talwar K. Mechanism design via differential privacy. Proc. of the 48th Annual IEEE Symp. on Foundations of Computer Science (FOCS), Piscataway, NJ: IEEE, 2007, pp.94-103. https://doi.org/10.1109/focs.2007.4389483

[38] DWork C, McSherry F, Smith A. Calibrating noise to sensitivity in private data analysis. Proc. of the 3th Theory of Cryptography Conf (TCC06), Berlin: Springer, 2006, pp.265-284. https://doi.org/10.1007/11681878_14