Research Article

Translocation-Based Algorithm for Publishing Trajectories with Personalized Privacy Requirements

Shuai Wang, Chunyi Chen, and Guijie Zhang

1School of Computer Science and Technology, Changchun University of Science and Technology, 7089 Weixing Road, Changchun 130022, China
2Teacher Training Centre of Colleges and Universities of Jilin Province, Jilin Normal University, Siping, Jilin 136000, China
3College of Computer Science and Technology, Jilin Normal University, Siping, Jilin 136000, China

Correspondence should be addressed to Chunyi Chen; chenchunyi@hotmail.com

Received 21 December 2019; Revised 9 February 2020; Accepted 30 March 2020; Published 22 April 2020

Academic Editor: Eric Florentin

Copyright © 2020 Shuai Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Up to now, a large amount of trajectory data have been collected by trusted servers because of the wide use of location-based services. One can extract useful information via an analysis of trajectory data. However, the privacy of trajectory bodies risks being inadvertently divulged to others. Therefore, the trajectory data should be properly processed for privacy protection before being released to unknown analysts. This paper proposes a privacy protection scheme for publishing the trajectories with personalized privacy requirements based on the translocation of trajectory points. The algorithm not only enables the published trajectory points to meet the personalized privacy requirements regarding desensitization and anonymity but also preserves the positions of all trajectory points. Our algorithm trades the loss in mobility patterns for the advantage in the similarity of trajectory distance. Related experiments on trajectory data sets with personalized privacy requirements have verified the effectiveness and the efficiency of our algorithm.

1. Introduction

With the maturity of location technology and wireless communication technology, location-based service (LBS) becomes more and more popular. In LBS, servers for anonymity (SA), which are trusted, receive location service requests of users and accumulate a large amount of trajectory data. The analysis of trajectory data can reveal useful knowledge, but it also threatens the privacy of the trajectory bodies. Trajectory data should be processed to protect its privacy before being published to third party for analysis, which is called the privacy protection for publishing trajectories.

The location service requests (LSR) contain personalized privacy requirements of the body on that location, and the requirements can inform the SA of body’s sensitivity. For the trajectory data with personalized privacy requirements, the existing privacy protection methods usually resist the identity link attack of trajectories and realize the personalized anonymity of trajectories [1–3]. Due to the curse of dimensionality [4], the anonymity of trajectories will cause a lot of data distortion.

We propose a Translocation-based Personalized Privacy Protection (TPPP) for publishing the trajectories with personalized privacy requirements on trajectory points. We assume that the personalized privacy requirements set by the body of the trajectory include sensitivity and anonymous threshold of the trajectory points. This method realizes the personalized privacy protection through the translocations of the trajectory points. Figure 1 shows a trajectory with personalized privacy requirements. Each trajectory point \( p_i \) has specific sensitivity thresholds \( d_s \) and the translocations of trajectory points cannot be assigned into the sensitive area in red. The trajectory publisher defines a maximum of distortion distance and is represented as \( d_{\text{max}} \). The translocations of trajectory points cannot exceed the range of the dotted line in green. The anonymous threshold \( k_i \) of \( p_i \) is not marked in Figure 1. \( k_i \) means the probability that inferring
the position of $p_i$ through its translocation trajectory point $p_i^*$ should not be more than $1/k_i$. We design two privacy protection algorithms according to whether the adversary has mastered the privacy requirements of $p_i$. Experiments show that the two algorithms are efficient and produce less distortion of trajectory data for publishing.

The main contributions of our proposed method are as follows:

First, we propose a privacy protection framework for the trajectory data with personalized privacy protection requirements, which include desensitization and anonymity of trajectory points.

Second, we propose a translocation strategy, by which the algorithm retains all the spatial-temporal positions in the original trajectory data. No dummy trajectory points are added, and no trajectory points that cannot be anonymous are removed, so the availability of published data is improved.

Third, a modified Manhattan distance metric is proposed to measure the distance between trajectory points in the space-time region, which improves the disadvantage of Euclidean distance in measuring the reachable distance between two points.

Finally, for the trajectory data with personalized privacy protection requirements, an experimental method is designed to compare the personalized privacy protection algorithm with non-personalized privacy protection algorithm on effectiveness and efficiency. We also verify the effectiveness and efficiency of our algorithms by comparison with the baseline algorithms.

The organizational structure of this paper is as follows. The first section introduces the research status and the main contributions of our work. The second section introduces the related work of personalized privacy protection for publishing trajectories. The third section describes the system model, defines the data symbols involved in the method, and introduces the problem statement. In the fourth section, we propose a modified Manhattan distance for measuring the accessible distance on the road network. The fifth section introduces two privacy protection algorithms under different background knowledge of adversary. In the sixth section, 24 trajectory data sets with personalized privacy requirements are constructed based on two classical trajectory data, and the comparison of different algorithms is implemented by using the median method. The last section summarizes the paper.

2. Related Work

Trajectory data is typical multidimensional data, and its $k$-anonymous publishing is NP-hard [5]. There are two kinds of non-personalized privacy protection methods according to different privacy objects. First, the publishers set a unified privacy protection level for all trajectories [6–11] and then complete the privacy protection for publishing trajectories. Second, trajectory points are taken as atomic operation objects; the publisher sets a unified privacy protection level for all trajectory points [12–16] and then completes the privacy protection. Wang, et al. [17] summarized and classified the non-personalized privacy protection methods of trajectory data. Bonchi et al. [18] introduced typical methods for publishing trajectories in detail.

Due to the maturity of location technology and communication technology, LBS is developed rapidly towards personalization. Some LBS allow users to set personalized privacy requirements on every LSR, so the trajectory data with personalized privacy requirements can be collected.

The personalized privacy protection for publishing trajectories can be divided into two categories by the originator of personalized privacy requirements.

2.1. Trajectory Body Sets the Personalized Privacy Requirements. This kind of method takes the trajectory identity as the privacy protection object and resists attacks on identity links by anonymizing trajectories. The algorithm firstly clusters the trajectory and then selects the highest level of privacy requirement in the cluster as the privacy level of the cluster. This method does not improve the effective privacy level but greatly increases the data distortion. Mahdavifar et al. [1] selected the trajectory with the highest level of anonymity requirement in trajectory data as the center to cluster the trajectories. Then the cluster was expanded until its capacity is not less than the privacy requirement of the center. Finally, the trajectory data is reconstructed and published. Lan et al. [2] mapped the trajectories and their relations to a weighted graph, and the trajectories whose similarity is not less than a certain threshold are connected by the edge. This algorithm takes the privacy difference between the trajectories on two endpoints as the weight of the edge. The algorithm constructed the anonymous set to realize privacy protection by partitioning the graph. Hu et al. [3] assumed that there was an independent privacy level on different trajectory segments and selected the maximum privacy requirements of trajectory segments as the privacy requirement. Similar to non-personalized trajectory privacy protection method, the
method that takes trajectory as the anonymous object will cause serious data distortion.

2.2. Trajectory Publisher Sets the Personalized Privacy Requirements. Firstly, the algorithm sets the privacy requirements of the trajectory points according to the distribution characteristics and geographical location of the trajectory data. Then, trajectory point is taken as the atomic operation object by the algorithm, and personalized privacy protection of trajectory points is realized through perturbation [19, 20] or generalization [21–23].

The personalized privacy protection of trajectory points is usually based on the classification tree of its semantic attributes. Dai et al. [19] first annotated the semantic attributes of all sampling points on the trajectories to establish a classification tree. Then they extracted the sensitive stop points and replace them by selecting the appropriate user interest points. Finally, they published the reconstructed trajectory data. Komishani et al. [21] assumed that each trajectory point has a unique sensitive attribute value, and all the sensitive attribute values in the trajectory data form an attribute classification tree. The algorithm took the level of sensitivity in the attribute classification tree as the privacy level of the trajectory and then generalized the trajectory. It further prevented identity link attacks by removing the critical subtrajectories. Deldar and Abadi [22] created a personalized noise trajectories tree based on preallocation of privacy level and realized the trajectory differential privacy protection with personalized location privacy requirements. Yang et al. [23] first defined the information active point with the level of personalized privacy requirements. Then the active information points are clustered into anonymous areas, so that the information leakage of the information active points did not exceed the privacy threshold of the area.

Such methods abandon the anonymity of trajectories and cannot resist attacks on identity links. The algorithms can only generalize the specific location of trajectory points according to different privacy requirements and make sure the probability of inferring the original location does not exceed a certain threshold [19, 21–23]. Compared with the first category, this kind of methods that takes the trajectory point as the privacy protection object can reduce the level of privacy protection, but it reduces the data distortion significantly. Semantic attribute classification tree is the basis of generalization, and it is obtained through data analysis or subjective judgment; subjectivity is a shortcoming of such methods.

2.3. Other Relevant Studies. Gedik and Liu [24] proposed a privacy protection framework for LBS, which allowed each trajectory point to specify the minimum anonymous threshold and the maximum spatial-temporal resolution. The algorithm detects a set of neighbors for each newly added message, in which the new message and its neighbors are enclosed in each other’s maximum tolerance box. Then the anonymous set containing new information is obtained by eliminating the trajectory points that do not meet the requirement of anonymity. If the number of trajectory points is not less than the highest level of anonymity requirement in the anonymous set, the anonymity succeeds; otherwise, the new message cannot be anonymous immediately. Gao et al. [25] allowed the publisher to select the model parameters of the algorithm individually without considering the personalized privacy requirements of different trajectories.

2.4. Problems and Shortcomings. The existing personalized privacy protection methods based on trajectory anonymity cannot solve the curse of dimensionality, which is bound to cause a lot of data distortion. The generalization method based on trajectory points uses the semantic attributes of the trajectory points frequently to set its privacy level, but they are all subjectively assigned. So the semantic attributes usually lead to errors in the evaluation of privacy requirement level and privacy protection processing of trajectory points. For the trajectory data with personalized privacy requirements defined by the trajectory bodies, the existent privacy protection algorithm based on trajectory points is for LBS. The privacy protection algorithm that processes the trajectory data with personalized privacy requirements on trajectory points, which is defined by trajectory bodies, is seldom reported, and our method belongs to this type.

3. System Model and Problem Statement

3.1. System Model

3.1.1. The Generation Process of Trajectory Data. Trajectory data is typically generated in LBS, consisting of three primary objects: the user, the trusted SA, and the untrusted location service provider (LSP), as shown in Figure 2. Location service request (LSR) issued by users includes user identity, location information, service content, and privacy requirements. In order to protect trajectory privacy, users send LSR to SA instead of LSP, and SA processes the received LSRs for privacy and generates a deliverable LSRs set. This process is called privacy protection of trajectories for LBS. Then SA sends the processed LSRs to LSP, and finally LSP returns the location service answers (LSA) to users through SA.

During the process of LBS, SA stores a large amount of original trajectory data, which can be analyzed to find useful information. Before publishing the trajectory data to the third party for analysis, SA should transform them to generate releasable trajectory data and ensure that the third party cannot disclose personal privacy from it. This process is called privacy protection for publishing trajectories, which is also the purpose of our study.

3.1.2. Personalized Privacy Requirements of Users. The privacy of the trajectory is objective requirement of the trajectory body. It should not be set by the collector only according to the semantic and distribution characteristics of the trajectory on the map. Due to the difference on the trajectory bodies and the spatial-temporal location, each trajectory point may have different privacy requirement, which is also known as the personalized privacy requirement.
For example, GPS accuracy is set to a high level during navigation services by some users. For others, the accuracy of GPS varies according to spatial-temporal location: sometimes users allow accurate GPS service to improve navigation accuracy, but they need to reduce GPS accuracy in sensitive spatial-temporal location.

3.1.3. Model of Trajectory Data

Definition 1
(1) Location service request (LSR), \( \langle \text{TID}, r_{\text{no}}, \{t, x, y\}, k, \{d_t, d_x, d_y\}, \text{Co} \rangle \) [24], where TID is user identity, \( r_{\text{no}} \) is the record number of current request, \( \{t, x, y\} \) is the spatial-temporal location of service request, \( k \) is the anonymity threshold, \( \{d_t, d_x, d_y\} \) is the desensitization threshold, and \( \text{Co} \) is the content of location service request. It should be noted that \( \{d_t, d_x, d_y\} \) is the minimum disturbance distance for the trajectory point to reach desensitization requirements, while \( \{d_t, d_x, d_y\} \) in LBS refers to the maximum distortion threshold of location service [24].

The trajectory data in SA can be regarded as a collection of trajectory points; it is defined as follows.

Definition 2
(2) Trajectory. A collection of spatial-temporal sampling positions belongs to the same moving body [8], denoted by \( \text{T} = \{\text{p}_1, \text{p}_2, \ldots, \text{p}_{|\text{T}|}\} \), where \( |\text{T}| \) represents the number of points in \( \text{T} \). The trajectory point \( \text{p}_i \) is represented as a tetrad \( (\text{TID}_o, t_o, x_o, y_o) \), where \( \text{TID}_o \) denotes the body of \( \text{T} \), \( t_o \) denotes the sampling time, and \( \{x_o, y_o\} \) denotes the spatial coordinate. \( \{t_o, x_o, y_o\} \) is called a triple. Trajectory data is a collection of trajectories [26], denoted by \( \text{TD} = \{\text{T}_1, \text{T}_2, \ldots, \text{T}_{|\text{TD}|}\} \), where \( |\text{TD}| \) represents the number of trajectories. Intuitively, TD is a set of points: \( \{\text{p}_1, \text{p}_2, \ldots, \text{p}_{|\text{TD}|}\} \), where \( |\text{TD}| \) represents the number of trajectory points in \( \text{TD} \) [17].

We add private information into Definition 2 and describe the trajectory point with personalized privacy requirement by Definition 3.

Definition 3
(3) Trajectory point with personalized privacy requirement. \( \text{p} = \langle \text{TID}, \{t, x, y\}, k, d \rangle \), where \( d = \{d_t, d_x, d_y\} \) is the personalized desensitization threshold; that is to say, the spatial-temporal range with \( \{t, x, y\} \) as the center and \( \{d_t, d_x, d_y\} \) as the radius is the sensitive region of \( \text{p} \), and \( \text{p} \) must be disturbed outside this range in order to be desensitized. Sometimes we also use \( d_t \) and \( d_x \) to represent the spatial-temporal components of the desensitization threshold, respectively. The anonymity threshold is denoted by \( k \). When \( \text{p} \) is translated to \( \text{p}^* \), in order to realize \( k \)-anonymity of \( \text{p} \), there must be at least \( k \) adjacent individuals around \( \text{p}^* \).

\( k \leq 1 \) indicates that there is no requirement for privacy protection. Even if the trajectory body sets desensitization threshold \( d > 0 \), the possibility of being restored \((1/k)\) does not make sense.

\( k > 1 \) indicates that the trajectory bodies need to protect the trajectory point with the anonymous level no less than \( k \). Even if there is no desensitization requirement \( (d = 0) \), it only indicates that the translocation range of \( \text{p} \) is \([0, d_{\text{max}}]\), where \( d_{\text{max}} \) is the maximum distortion threshold proposed by the analyst. In the following statements, we only use \( d_{\text{max}} \) to express its components in the spatial dimension, while the components in the temporal dimension can be expressed by \( d_{\text{max}}/v \), where \( v \) is the trajectory speed.

Definition 4
(4) Transformed Trajectory Database. According to certain rules, the replacement of \( \text{p}_i \) with other regions is called position transformation of \( \text{p}_i \) in which \( \text{p}_i \) is a trajectory point of \( \text{T} \), \( 0 \leq i \leq |\text{T}| \). The region may be a point or an area, or empty, denoted by \( \text{p}_i^* \). The new trajectory is called transformed trajectory of \( \text{T} \), denoted by \( \text{T}^* \). The new trajectory data is called the transformed trajectory data of \( \text{TD} \), denoted by \( \text{TD}^* \). Typically, the published trajectory data is also represented by \( \text{TD}^* \) [17].

3.2. Problem Statement. Before publishing \( \text{TD} \) to the third party for analysis, the process that transforms \( \text{TD} \) into \( \text{TD}^* \) based on the personalized privacy requirement is called personalized privacy protection for publishing trajectories.

For the \( \text{TD} \) with personalized privacy requirements mentioned in the data model, the main concern of our algorithm is to ensure the personalized privacy requirements and improve the availability of published data as much as possible. We hope to design trajectory privacy protection algorithm based on the trajectory points to meet the personalized privacy requirements, which include desensitization from the sensitive area and \( k \)-anonymity of the trajectory points. But the \( k \)-anonymity of trajectories is not required in our study. At the same time, the original position of the trajectory points should be maintained to reduce the distortion of \( \text{TD}^* \) and improve its usefulness.

The background knowledge of adversary is related to \( \text{TD} \)’s privacy policy directly. We divide the assumptions of the background into two levels. The strict assumption is that the adversary has mastered some trajectory points of the target and the privacy requirement level of the target trajectory points. The second level assumes that the adversary only knows some trajectory points of the target but does not know the privacy requirements level of the target trajectory points. This is also a more realistic set of background
knowledge. The privacy requirements are more private than the location of the trajectory point. It is hard to imagine that an adversary has mastered its privacy requirements without knowing the sensitive location of the target. In the case that the adversary does not know privacy level of the goal points, the trajectory points can achieve a more efficient $k$-anonymity and desensitization. In the Solution section, we discuss the privacy protection policies in both cases.

4. Modified Manhattan Space Distance

Although the traditional Euclidean distance accurately expresses the linear distance between geographical objects, the paths are usually polygonal in the actual geographic connected region. Therefore, it is more reasonable to use the Manhattan distance to measure the distance between the trajectory points [3].

In most geographical areas, road direction depends on physical geographical conditions, and the direction of trajectory data is the reflection of road direction. Therefore, the main directions of the local roads can be estimated approximately by the statistical characteristics of the trajectories. The coordinate axis can be adjusted so that it is parallel to the two main path directions in the region, and the trajectory point coordinates can be projected onto the new coordinate axis. This coordinate axis transformation is a compromise between Manhattan distance and the road network constraint distance, which improves the authenticity of the distance between the track points and then avoids the fitting of the specific road network information and the trajectory data in the privacy protection. The transformation process is as follows:

(a) Take the Due east as the reference vector; the trajectory segment in each direction was sampled. The radian of the angle formed by the direction of the trajectory segment and the reference vector was defined as

$$\theta = \arctan \left( \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right), \quad \theta \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right].$$

(1)

It can be converted into a nonnegative angle and rounded into

$$\bar{\theta} = \text{INT} \left( \frac{1}{2} \left( \theta + \pi \right) \right) \cdot \frac{180}{\pi}, \quad \bar{\theta} \in [0, \pi].$$

(2)

(b) Take these two directions with the longest mileage as new horizontal and vertical axis of the map.

(c) Take these two vectors as new axes and transform coordinates of each point. Assume that the ideal coordinate conversion directions are as follows: the anticlockwise rotation angle of the $X$-axis is equal to $\alpha$, and the anticlockwise rotation angle of the $Y$-axis is equal to $\beta$; then the new coordinates of the original trajectory point $(x, y)$ under the new coordinate axis are

$$\begin{align*}
y' &= y - \frac{\cos \alpha}{\cos (\alpha - \beta)} x - \frac{\sin \alpha}{\cos (\alpha - \beta)} y \\
x' &= x - \frac{\cos \beta}{\cos (\alpha - \beta)} y + \frac{\sin \beta}{\cos (\alpha - \beta)} x.
\end{align*}$$

(3)

(d) Considering that the roads in most areas are almost vertically intersected, when we simply select two mutually perpendicular directions as the new coordinate axis direction, the coordinate transformation formula is simplified as follows:

$$\begin{align*}
x' &= x \cos \alpha + y \sin \alpha, \\
y' &= y \cos \alpha - x \sin \alpha.
\end{align*}$$

(4)

(e) According to the new coordinates of each trajectory point, the sum of the difference of each dimension is the modified Manhattan distance:

$$\text{Dis} = \Delta x + \Delta y + \Delta tv,$$

(5)

where $v$ is the average speed of the trajectory.

For example, we take the trajectory data of New York (see Figure 3(a)) to calculate the modified Manhattan distance, and the sum of the trajectory length at various angles is shown in Figure 3(b).

Comparing our new coordinate axis with the map of New York on Baidu (Figure 3(c)), we found that the main directions of the trajectories obtained by statistics (about 60 and 150) are basically consistent with the directions of urban traffic roads. Then the rotation angles of the $X$-axis and $Y$-axis in the trajectory coordinate system are, respectively, equal to $\alpha = 60 - 0 = 60$ and $\beta = 150 - 90 = 60$. In the transformed coordinate system, the distance from B to O in Figure 3(d) is converted from BC + CO to BF + FO, which is reasonable.

5. Solution

Intuitively, TD with personalized privacy requirements is a collection of trajectory points. As shown in Definition 3, all trajectory points are distributed in a three-dimensional space spanned by time and space dimensions. We can take trajectory points as atomic objects of trajectory privacy protection algorithms. In order to improve the usability of TD, our method gives up the $k$-anonymity of trajectories and only requires the trajectory points to meet $k$-anonymity and desensitization requirements. Desensitization threshold $d$ is a distance relationship; that is, $p^*$ should keep a certain distance from its original position $p$; anonymity threshold $k$ is a structural relation; that is, the
probability to recognize the original position of $p$ from $p^*$ is no more than $1/k$.

We show the anonymity and desensitization of trajectory points after translocation in Figure 4. We combine the two spatial dimensions into one and mark the spatial-temporal dimensions as $(x, y)$ for convenience.

In Figure 4, the dark gray area is the sensitive area of its centroid $p$, and the displacement position $p^*$ should be out of it to ensure desensitization. The light gray area is the replaceable area of $p$, and desensitization can be realized if $p$ is replaced in this area, and the distortion of $p^*$ meets the requirements of the trajectory data analyst. If $p$ is moved to the white area, the distortion will exceed the acceptable range of the data analyst. The upper limit of this distortion is usually set by SA, so it is consistent for each trajectory point. We assume that, for each trajectory point $p$, the distance between the translocated position $p^*$ and the original position $p$ cannot exceed the effective threshold $d_{\text{max}}$.

In order to ensure that the target trajectory point $p^*$ used for translocation satisfies $k$-anonymity, the number of trajectory points that can be translocated to $p^*$ should not be less than $k$. We use the dotted arrow to represent the translocated relation between the trajectory points. In Figure 4, five trajectory points can be translocated to $p^*$, so $p^*$ meets $5$-anonymity.

The most important factor determining the strategy for protecting the privacy of trajectory points is whether the adversary has known the privacy requirements $(k, d)$ of the target trajectory points. We discuss the privacy protection algorithms in two cases, respectively, in the following part.

5.1. Privacy Requirements are Known by Adversary (PRK).

We assumed that the information possessed by the adversary includes $p_i^* = \langle \text{TID}_i, \{t_i^*, x_i^*, y_i^*\}, k_i, d_i \rangle$ and TD*. The goal of the adversary is to get the original position $\{t_i, x_i, y_i\}$ of $p_i^*$.
The trajectory body requires that $p_i^* \prime$ is outside the sensitive area of $p_i$, and the probability to find $[t_i, x_i, y_i]$ is no more than $1/k_i$. The data analyst requires that the distortion of $p_i^* \prime$ should not exceed $d_{\text{max}}$. In order to find a suitable $p_i^* \prime$ for $p_i$, we design TPPP for publishing trajectories in the case of Privacy Requirements are Known (PRK).

5.1.1. Description of the Algorithm. We take two steps to find a trajectory point that meets the desensitization and anonymity requirements of $p_i$.

First, we search for the candidate trajectory point for translocation (CTPT) $p_i^* \prime$ that meets desensitization requirements of $p_i$ in $\text{TD}$. The distance between $p_i^* \prime$ and $p_i$ was required to conform to the following relation:

$$d_i \leq |p_i - p_i^* \prime| \leq d_{\text{max}}.$$  

(6)

Second, we search the trajectory point $p_j$ that can be translocated to $p_i^* \prime$ to achieve the same desensitization threshold ($d_i$) as $p_i$:

$$d_j \leq |p_j - p_i^* \prime| \leq d_{\text{max}}.$$  

(7)

Then, the points conforming to formula (7) are added to set $P_j$ and the number of elements is represented by $|P_j|$. Since $d_i$ is uniformly used as the desensitization threshold, formula (7) actually reflects the adversary’s method to invert the position range of $p_i$, which is $P_j$ under the condition of mastering $d_i$ and $p_i^* \prime$. However, it must be noted that the anonymity level of $p_i^* \prime$ is the number of $p_i$ conforming to formula (7) but not $|P_j|$. If $P_j$ satisfies formula (8), then $p_i$ can be translocated to the position of $p_i^* \prime$; the probability of deducing the original position according to $d_i$ and $p_i^* \prime$ is no more than $1/k_i$.

$$|P_j| \geq k_i.$$  

(8)

Therefore, $\text{TID}_i$ of $p_i$ can be translocated to $p_i^* \prime$ to form a new trajectory point:

$$p_i^* = \langle \text{TID}_i; \{t_i', x_i', y_i'\} \rangle.$$  

(9)

At this point, the records in $\text{TD}^*$ are

$$\langle \emptyset; \{t_i, x_i, y_i\} \rangle,$$

(10)

$$\langle \text{TID}_i; \text{TID}'_i; \{t_i', x_i', y_i'\} \rangle.$$  

(11)

Record (10) shows that the algorithm deletes not $p_i$ but its $\text{TID}_i$ and privacy requirements. There are two benefits: on one hand, the data analyst can make use of the trajectory points without $\text{TID}$ for necessary data analysis, which can improve the accuracy of data analysis, such as the density of trajectory points on some time point. On the other hand, the trajectory points without $\text{TID}$ can also improve the density of trajectory points locally, so as to improve the anonymity level of trajectory points adjacent to it.

If $|P_j| < k_i$, $p_i$ cannot satisfy $k_i$-anonymity; then we iteratively find trajectory points that satisfy formulas (7) and (8) in $\text{TD}$ and translocate $p_i$ to a new position.

For $p_i$, what should we do when there is no trajectory point that meets both formulas (7) and (8)? That is, there is no replaceable position that meets both anonymity and desensitization requirements. To address this interesting problem, we propose four options and choose the fourth one according to the same reason as Record (10).

1. Increase the system distortion threshold $d_{\text{max}}$ to expand the replaceable region of $p_i$. This will cause more data distortion;
2. Decrease the sensitive threshold $d_i$ of $p_i$ to expand the replaceable area. This will lead to a privacy crisis;
3. Decrease $p_i$’s anonymity threshold $k_i$. It would also lead to a privacy crisis;
4. Delete $\text{TID}_i$ of $p_i$.

The pseudocode of TPPP (PRK) is shown below.

5.1.2. Data Structure. In Algorithm 1, both step (3) and step (5) need to traverse the whole $\text{TD}$; it leads to a lot of redundant calculations. We use time-window structure based on timing chain [17] to simplify this operation.

$\text{TD}$ is stored as records, and we map $\text{TD}$ to a linked list. The temporal distance between the trajectory points is one-dimensional and is positively correlated with the spatial-temporal distance between the trajectory points. Therefore, by filtering in the time dimension, the scope of comparison between trajectory points can be narrowed. We first sort the trajectory data list by sampling time and formed a sequential linked list (SLL) and then limited the number of candidate trajectory points with the time-window structure. This will greatly speed up the determination of the translocated relation between the trajectory points.

The time-window structure is shown in Figure 5. In step (3) of Algorithm 1, when the operation pointer (OP) points to $p_i$, we search CTPTs of $p_i$. Only the point $p_i^* \prime$ whose sampling time is $t'$ meets the condition $d_i, t \leq |t_i - t'| \leq d_{\text{max}} - t$ can be selected as the CTPT of $p_i$. Then we estimate whether
The time window meets the desensitization threshold, and the time complexity is \(O(N_w)\). The second step is to calculate the anonymity level for each CTPT of \(p\), and the time complexity is \(O(N_w \cdot N_d)\).

Therefore, the time complexity for one trajectory point is \(O(N_w + N_w \cdot N_d) = O(N_w + N_w \cdot x\% \cdot N_w) = O(x\% \cdot N_w)\). We need to do this process for each point. Since the number of trajectory points in \(TD\) is \(N\), the total time complexity of the algorithm is \(O(N \cdot N_w \cdot (x\% \cdot N_w + 1))\).

5.1.4. Demonstration of Example. We take \(p_1\) in Figure 6 as an example to demonstrate the process of PRK algorithm, where \(k_1 = 5\).

Firstly, we search for CTPTs in the desensitization region of \(p_1\) (as shown in Figure 6). We can see that CTPTs \((p_1) = \{p_3, p_4, p_5, p_6\}\).

Then, we further examine the anonymity of CTPTs (Figure 7). Since the adversary has mastered the desensitization threshold \(d_1\) of \(p_1\), we suppose that \(p'\) can be used as the CTPT of the \(k\) trajectory points, where \(p'\) is a CTPT of \(p_1\). The \(k\) trajectory points include \(p_1\), and the distance from \(p'\) satisfies formula (7). If \(k \geq 5\), \(p'\) can be used as the replaceable position of \(p_1\). Figure 7 shows whether the four CTPTs of \(p_1\) satisfy 5-anonymity.

We can see from Figure 7 that \(p_4\) and \(p_5\) can be used as the replaceable positions of \(p_1\), so that \(p^*_1\) satisfies \(d_1\)-desensitization and 5-anonymity, and the distortion of \(p^*_1\) does not exceed \(d_{\text{max}}\).

5.1.5. The Effectiveness on Privacy Protection. In the example above, we assume that the translocation target of \(p_1\) is \(p_6\), and the published trajectory points are shown in Figure 8. The background knowledge known by the adversary includes \(k_1, d_1\), and \(d_{\text{max}}\). We analyze the effectiveness of \(p^*_1\) for privacy protection.

1) Desensitization. As shown in Figure 8, we can draw a circle with \(p_1\) as the center and \(d_1\) as the radius. Since \(p^*_1\) (\(p_5\)) is outside the sensitive region of \(p_1\), the desensitization of \(p^*_1\) for \(p_1\) can be guaranteed.

2) Anonymity. Adversary can draw two concentric circles with \(d_{\text{max}}\) and \(d_1\) as the radius, respectively, and take \(p^*_1\) as their center. There are 5 trajectory points that can be
translocated to the position of $p_1^*$ and achieve desensitization. So $p_1^*$ satisfies 5-anonymity and the probability of determining the original position of $p_1^*$ is no more than 1/5, as shown in Figure 8.

5.1.6. The Effectiveness on Preserving the Mobility Patterns of Trajectories. Temporal properties related to the trajectories, such as mobility patterns, are important features of the trajectory data. At the beginning of our study, we considered the problem of preserving the mobility patterns of the published trajectories. We could avoid time conflict by limiting the temporal distance between current trajectory point $p$ and target trajectory point $p^*$ to $[t_i - ((t_i - t_{i-1})/2), t_i + ((t_{i+1} - t_i)/2)]$ for translocation. However, this would cause a large distortion of spatial

**Figure 6:** Desensitization region and CTPTs of $p_1$. $p_2$ lies in the sensitive region of $p_1$; if $p_1$ is translocated to $p_2$, desensitization of $p_1$ cannot be achieved. $p_3$, $p_4$, $p_5$, and $p_6$ are in the desensitization region of $p_1$ and are called CTPTs. $p_7$, $p_8$, and $p_9$ exceed the valid region of $p_1$.

**Figure 7:** The anonymity of the four CTPTs of $p_1$. (a) $p_3$ can be used as the replaceable position of $p_1$, $p_5$, and $p_7$. The anonymity of $p_3$ for $d_1$ is 3. (b) $p_4$ can be used as the replaceable position of $p_1$, $p_2$, $p_6$, $p_7$, and $p_8$. The anonymity of $p_4$ for $d_1$ is 5. (c) $p_5$ can be used as the replaceable position of $p_1$, $p_2$, $p_3$, $p_4$, and $p_6$. The anonymity of $p_5$ for $d_1$ is 5. (d) $p_6$ can be used as the replaceable position of $p_1$, $p_2$, and $p_5$. The anonymity of $p_6$ for $d_1$ is 3.
information, so we gave up to preserve the temporal characteristics in the published trajectory data. Our strategy divides the information of the trajectory point into two parts ((TID), (t, x, y)) and then transfers its (TID) to the location that meets the privacy protection requirements, so there is no spatial-temporal conflict in the published trajectories. The distortion of the transformed trajectories is limited by $d_{max}$, so the overall trend of the trajectories will be preserved and the statistical characteristics of the trajectory points will be less distorted.

Intuitively, our method defines the scope of the original trajectory $T$ with the published trajectory $T^*$: a tubular region centered on $T^*$, where the wall thickness of the tube at the sampling time $t$ can be represented as $(d_{max} - d)$. The original trajectory $T$ is on the wall of the tube. This is similar to the result of the method based on trajectories clustering, which generalizes the original trajectories by a curved column with an axis $T^*$ and a radius $\delta$.

### 5.2. Privacy Requirements are Unknown by Adversary (PRU)

In the case that privacy requirements are unknown, we assume that the adversary’s background knowledge is $p_i^* = \langle TID_i, \{t_i^*, x_i^*, y_i^*\} \rangle$. The adversary’s target and the privacy requirements of trajectory bodies are consistent with those in PRK. Since the adversary does not know $(k_i, d_i)$, which are the privacy requirements of $p_i$, it is impossible to judge the position range of $p_i$ based on $d_i$ and $p_i^*$. Therefore, formula (6) in PRK can be changed as follows:

$$d_i \leq |p_j - p'_j| \leq d_{max}.$$  \hspace{1cm} (12)

It is reasonable to replace $d_i$ with desensitization threshold $d'_i$ of $p_i$. Formula (12) reflects whether $p'_j$ can be used as CTPT of $p_i$ to realize desensitization of $p_i$. Another benefit of this change is that $|p_j|$ is fixed as the number of trajectory points that can be replaced to $p'_j$ and realize desensitization, namely, the anonymity of $p'_j$. In this way, we can calculate the anonymity of all trajectory points in one time, instead of calculating the anonymity based on $d_i$ for each CTPT of $p_i$.

Comparing formula (7) with formula (12), we can see they are in the similar form. In other words, in the case of PRU, by a round of comparison such as formula (7), the judgment that whether $p'_j$ can be used as CTPT of $p_i$ can be completed, and the accumulation of anonymity of $p'_j$ finished at the same time.

In order to record the distance relation and structural relation between trajectory points on TD, we mapped the trajectory points and their position relation to a directed relation network of trajectory points (RNTP), where the nodes represent the trajectory points. Desensitization can be achieved by replacing the trajectory point at the beginning of the directed edge with the trajectory point at the end of the directed edge. The indegree of a node $p$ indicates the number of the points taking $p$ as CTPT, which also reflects the anonymity of $p$.

On the obtained RNTP, for each trajectory point $p_i$, we select a neighbor whose indegree is not less than $k_i$ as the translocation trajectory point of $p_i$. If there is no such neighbor, we delete TID of $p_i$ directly, according to the same reason as Record (10).

#### 5.2.1. The Data Structure of RNTP

In order to improve the efficiency of execution, we select the linked list as the storage structure of RNTP [17]. As shown in Figure 9, $P_{previous}$ points to the previous node of the current node $p$, $SerialNum$ is the unique Serial Number of $p$ on TD list, and TID is the identifier of the trajectory point, representing the body of the trajectory point. $(t, x, y)$, $k_i$ and $d_i$ are the spatial-temporal position, anonymous threshold, and desensitization threshold of the trajectory point, respectively. $P_{next}$ points to the next node of $p$, $NeighborNum$ is the indegree of $p$, and $P_{neighbor}$ points to the neighbors tree (NT) of $p$.

The neighbor set of current node is stored in the form of the linked list in the binary search tree. $P_{father}$ points to the parent node of $p$, $SerialNumber$ is the unique sequence number of the neighbor node, $P_{left}$ and $P_{right}$ point to the left and right child nodes of $p$, and $P_{neighbor}$ Points to the current neighbor in the main storage linked list.

#### 5.2.2. Construction of RNTP

To improve efficiency, trajectory points in TD are first stored in a SLL, and the concept of time-window is introduced, which is the same as in PRK. The difference is that PRU needs to link the trajectory points in TD into a directed relation network through the neighbor tree of each trajectory point.

In order to build RNTP, we detect CTPTs for each trajectory point in the time-window of $p_i$. If $p'_j$ satisfies $d_i \leq |p_j - p'_j| \leq d_{max}$, the directed edge from $p_j$ to $p'_j$ was established, and the indegree of $p'_j$ was increased by 1, which was also the anonymity of $p'_j$. The schematic diagram for the time-window and neighbor tree of the node $p_i$ is shown in Figure 10.

#### 5.2.3. Algorithm Based on Translocation of Trajectory Points on RNTP

In RNTP, NT$_i$ has recorded all CTPTs that satisfy desensitization. In order to make $p_i''$ satisfy $k_i$-anonymity, we only need to select a trajectory point $p'_j$ randomly in NT$_i$, whose indegree is not less than $k_i$, and translocate TID of $p_i$ into $p'_j$ to complete the personalized privacy protection of $p_i$. 
Data structure of the node in RNTP

\[ P_{\text{next}} \] SerialNum TID x, y \[ P_{\text{prev}} \] NeighborNum \[ P_{\text{children}} \] \[ k, d, x, y \]

The previous triple \quad The next triple \quad The neighbors tree

Figure 9: Data structure of the node and its NT in RNTP.

Sequential linked list of TD

\[
\begin{align*}
(TID_1, t_1, x_1, y_1) \\
(TID_2, t_2, x_2, y_2) \\
\vdots \\
(TID_{n-1}, t_{n-1}, x_{n-1}, y_{n-1}) \\
(TID_n, t_n, x_n, y_n)
\end{align*}
\]

Neighbors tree of \( p_i \)

\[
\begin{align*}
nb_1 \\
nb_2 \\
nb_3 \\
nb_4
\end{align*}
\]

Figure 10: The time-window for \( p_i \) in SLL of TD and NT of \( p_i \).

The new trajectory points formed after translocation as shown in formula (9). At this time, the record form in TD is consistent with Record (10) in the case of PRK. The pseudocode of the algorithm is shown as follows.

By calling Algorithms 2 and 3 successively, the personalized privacy protection based on translocation can be realized, and we can obtain TD*.

5.2.4. Analysis of the Efficiency. In the case of PRU, the assumptions are the same as PRK. The privacy protection algorithm is divided into two steps: constructing RNTP and the translocation of trajectory points, so we analyze the efficiency of the algorithm in two steps.

Step 1. Time complexity of Algorithm 2

Algorithm 2 estimates whether the distance between trajectory point \( p_i \) and the points in its time-window meets the desensitization requirement and joins the trajectory point \( p'_i \) that meets the requirements into the neighbor tree NT of \( p_i \) and then updates the indegree of \( p'_i \). The time complexity is \( O(N_u) \). The time complexity of the above operation for all trajectory points is \( O(N \cdot n \cdot N_u) \).

Step 2. Time complexity of Algorithm 3

The time complexity of finding CTPTs that meet \( k_i \) anonymity in the neighbor tree of \( p_i \) is \( O(N_d) \). The time complexity of the above operation for all trajectory points in RNTP is \( O(N \cdot n \cdot N_u) \).

In the case of PRU, the privacy protection of the TD can be completed by calling Algorithms 2 and 3 successively, so the total time complexity is \( O(N \cdot n \cdot N_u \cdot (x + 1)) \).

5.2.5. Demonstration of Example. We take the trajectory point in Figure 6 as the example to demonstrate the process of privacy protection algorithm in the case of PRU.

The analyst sets a consistent maximum distortion threshold \( d_{\text{max}} \) for all trajectory points, but the sensitivity threshold \( d \) for each trajectory point is independent. Therefore, each trajectory point has independent range for selecting its CTPTs. For example, in Figure 11(a), the sensitivity threshold of \( p_1 \) is small, and \( p_4 \) is not in its sensitivity range, so \( p_4 \) can be regarded as the CTPT of \( p_4 \). The sensitivity threshold of \( p_4 \) is relatively large, and \( p_3 \) is in the sensitive range of \( p_4 \), so \( p_3 \) cannot be regarded as CTPT of \( p_4 \). According to Figure 11(a), we can establish the RNTP, as shown in Figure 11(b).

We can see from Figure 11(b) that the indegree of each trajectory point is its anonymity level, and it can be statistically calculated, as shown in Table 1.

The anonymous requirements of each trajectory point are shown in Table 1. Then we can randomly replace the trajectory points to the adjacent trajectory points that meet their anonymous requirements, as shown in Figure 12.

In Figure 11(b), the anonymity threshold of \( p_4 \) is 4, but there is no trajectory point whose anonymity reaches 4 in its CTPTs, so only TID of \( p_4 \) is deleted for the same reason as Record (10). \( p_3 \) and \( p_8 \) were all translocated to the position of \( p_4 \), but no trajectory points were translocated to the position of \( p_8 \). The result is shown in Figure 12.

5.2.6. The Effectiveness of TPPP (PRU). According to the assumption of the background knowledge of the adversary in the case of PRU, the information known is only \( d_{\text{max}} \). We take \( p'_i \) in Figure 12 as an example to analyze the effectiveness of privacy protection after translocation.

(1) Desensitization. As shown in Figure 11(a), we draw a circle with \( p_i \) as the center and \( d_i \) as the radius. Since \( p'_i \) is outside the sensitive region of \( p_i \), the desensitization of \( p'_i \) related to \( p_i \) can be guaranteed.

(2) Anonymity. As shown in Figure 12, the adversary can draw a circle with \( p'_i \) as the center and \( d_{\text{max}} \) as the radius. There are three trajectory points that can be translocated to the position of \( p'_i \) and achieve desensitization. So the probability of determining the original position of \( p'_i \) is no more than 1/3.

In Figure 12, if the adversary has known the privacy requirements of \( p_i(k_i, d_i) \), when \( d_i > \text{Dis}(p'_i, p'_c) \) and \( d_i > \text{Dis}(p'_c, p'_b) \), the original location of \( p_i \) can be uniquely determined as \( p'_i \). Therefore, the published trajectory data generated by PRU algorithm is not safe in the case of PRK.

6. Simulation and Experimentation

6.1. The Purpose of Experiment and Baseline Algorithms.

We proposed the TPPP algorithm and analyzed its effectiveness in the section above. Here we compare the typical
trajectory privacy protection algorithms and TPPP in terms of data distortion, mobility patterns, and execution efficiency on the TDs which set personalized privacy requirements.

Considering the personalization and granularity of privacy protection objects, we select three typical trajectory privacy protection algorithms as baseline algorithms.

**Algorithm 2: Map TD to a relation network of trajectory points.**

1. Input: TD
2. Map TD to a sequential linked list (SLL)
3. For i = 1 to |TD|p
4. For each p’ in the time-window of pi
5. If di ≤ |pi − p’| ≤ dmax then
6. create a node in NTi and point its Pneighbor to p’
7. Pr.NeighborNum++
8. Endif
9. Next p’
10. Next i
11. Output SLL as the RNTP

**Algorithm 3: Algorithm based on translocation of trajectory points on RNTP.**

1. Input: RNTP
2. For i = 1 to |TD|p
3. Select a p’ in the NTi of pi while p’.NeighborNum ≥ di
4. If exist p’ then add TIDi to p’
5. Delete TIDi from pi
6. Next i
7. Output RNTP as the TD*

**Figure 11:** The distance relations and the structure relations of the trajectory points. (a) Personalized desensitization requirements of trajectory points. (b) Relation network of trajectory points.

Considering the nonpersonalized trajectory privacy protection algorithm based on trajectory anonymity, we chose the classical NWA algorithm [7]. Considering the personalized privacy protection algorithm based on trajectory anonymity, we selected the algorithm in [1] and represented it with CPPP. Considering the nonpersonalized privacy
6.2. The Method of Experiment. We design two groups of experiments. The first group of experiments compared classical algorithms with our algorithm. The second group of experiments compared the two proposed algorithms on TDSs with different characteristics on personalized privacy requirements. Then the data distortion and mobility patterns of the processed TDSs and time efficiency of the algorithm are evaluated.

6.2.1. Data Distortion. The information loss in TDS is represented by formula (12).

\[ |TD|_p \] represents the number of trajectory points in TDS, \( p_1^* \) is the transition position of \( p_i \), and \( Dis \) represents the Hamiltonian distance between the two positions.

If \( p_i \) is deleted by privacy protection algorithm, then the distortion is defined as a certain value \( \Omega \) generally. For the suppressed trajectory point, the following method is adopted to calculate its Dist: we assumed that, on the microscopic trajectory segment \( T \), the trajectory body performs uniform linear motion. According to the time position of \( p_i \) on the segment \( p_{i-1}^* \) and \( p_{i+1}^* \) in \( T \), we can determine a position \( p_i^* \) on the line from \( p_{i-1}^* \) to \( p_{i+1}^* \) and take the distance between \( p_i \) and \( p_i^* \) as the distortion caused by deleted \( p_i \).

We illustrate this approach in Figure 13. Suppose that there are three continuous trajectory points: \( p_1, p_2, \) and \( p_3 \). When \( p_2 \) is suppressed by the privacy protection operation, only \( p_1 \) and \( p_3 \) will be present in TDS. When calculating Dis (\( p_2, p_3 \)), we take \( x = |p_1, p_2|/|p_2, p_3| \) as proportional parameters and take the point \( p_2^* \) on the line \( p_1^*, p_3^* \), such that \((|p_1, p_2|/|p_2, p_3|) = x \). Then we treat Dist (\( p_2, p_2^* \)) as the distortion of \( p_2 \) in TDS.

6.2.2. Mobility Patterns. In order to capture the difference in mobility patterns before and after the transformation of trajectory data, we refer to the method of [19], including trajectory direction deviation (TDD) and trajectory distance utility (TDU).

TDU is obtained by the following formula:

\[
TDU(T, T^*) = \sum_{i=1}^{T} \text{Dis}(p_i,p_i^*). \tag{13}
\]

TDU represents distance similarity between trajectories. The experimental results on the trajectory data are the same as those of Infloss, and this conclusion can also be derived mathematically. We differ from [19] only in that we do not normalize the results, but the advantage of this is that it is easier to compare with the formulas in Infloss.

TDD is obtained by the following formula:

\[
TDD(T, T^*) = \frac{1}{|T|} \sum_{i=1}^{T} \cos \theta_i, \tag{14}
\]

where \( \theta_i \) is the angle between the \( i \)th trajectory segment on \( T \) and that on \( T^* \). TDD reflects the direction change and temporal characteristics of the trajectory segment.

6.2.3. Two Instructions. For the trajectory data whose personalized privacy requirements are set on trajectory points by the trajectory bodies, there are few widely used privacy protection methods for publishing trajectories. Therefore, the baseline algorithms we chose cannot directly process the

<table>
<thead>
<tr>
<th>Trajectory points</th>
<th>Indegree/Anonymity</th>
<th>Anonymous requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 )</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>( p_4 )</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>( p_5 )</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>( p_6 )</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( p_7 )</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>( p_8 )</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( p_9 )</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 1: The indegree and the anonymous requirements of each trajectory point.
TD with personalized privacy requirements. In order to effectively compare and evaluate various algorithms, we make the following two statements encountered in the experiment.

(1) Personalized Privacy Requirements Change the Experimental Method of Privacy Protection Algorithm Fundamentally

Traditionally, TD is generated by the trajectory bodies without personalized privacy requirements, while privacy requirements are set by the data publisher. Therefore, we compare the performance of different algorithms with same privacy requirements. We compare the applicable scope of the algorithm by changing TD’s overall privacy requirements.

However, in TD with personalized privacy requirements, both TD and privacy requirements are generated by the trajectory bodies. The personalized privacy protection algorithm should dispose different privacy requirements in TD at the same time. So the experiment cannot compare the applicable scope of the algorithms by adjusting the privacy requirement from the perspective of TD publisher. To create different comparison scenarios for the algorithms, TDs with different personalized privacy requirements must be provided. This is why we constructed 24 TDs with personalized privacy requirements in the following section.

(2) How to Apply Baseline Algorithm is an Important Problem for TD with Personalized Privacy Requirements on Trajectory Points?

For the personalized privacy protection method for trajectories, we take the median of privacy requirements of all trajectory points in the current trajectory as the privacy requirements of trajectory, where the anonymous threshold of the trajectory $T_i$ is represented by $k_{T_i}$.

For the nonpersonalized privacy protection method for trajectories, we take the median of all privacy requirements of trajectory in TD as the privacy requirements of TD, and the anonymous threshold of TD is represented by $k_{TD}$.

For the nonpersonalized privacy protection method for trajectory points, we take the median of privacy requirements of all trajectory points in TD as the privacy requirements of TD, expressed by $k_p$.

In these above three baseline algorithms, the distortion threshold $d_{max}$ is the same, and it is set by the trajectory publisher according to the requirements of data accuracy by the trajectory analyst. In different algorithms, $d_{max}$ represents cluster radius or maximum distortion radius. $d_{max}$ is also the only parameter in PRK and PRU, because the privacy parameters used in these two algorithms are set by the trajectory bodies in TD.

The following example illustrates the estimation process of the above parameters. Suppose that a TD contains three trajectories whose length is nine; the anonymous parameters of each trajectory points are shown in Figure 14.

We can obtain the following results through the anonymous parameters of the trajectory point. The median of all the anonymous parameters in TD is $k_p = 3$. The anonymous parameter of the three trajectories is 3, 3, and 2, respectively. The anonymous parameter of all trajectories in TD is $k_{TD} = 3$.

Inevitably, our choice of the median in some privacy risks for TD. However, in the experiment, compared with choosing the maximum privacy requirement as the privacy protection parameter of the algorithms, this choice is relatively fair for different algorithms.

6.3. The Data for Experiment. Although there are some LBS that can set personalized privacy requirements, the availability of such data is limited. The data in the experiment is usually constructed manually. We select two examples of typical trajectory data as the basic data for constructing trajectory data with personalized privacy requirements, which are the real-life TD Trucks and the synthetic TD Oldenburg. For a detailed introduction of these two TDs, please refer to [17]. We only refer to the relevant Figure 15 and Table 2 in [17] to provide a brief description.

In Table 2, we describe the characteristics of the two TDs: $|TD|$ is the number of trajectories, Area is the spatial span covered by TD, and $T_{sapan}$ is the time span; $AS$ is the average speed, and Density is the spatial-temporal density of trajectory points, which is the quotient of the trajectory point’s number and spatial-temporal span.

There are two methods to construct experimental data in the existing personalized trajectory privacy protection problems. The first method simulates the trajectory body and randomly adds personalized anonymous requirements to the trajectory data and forms TDs [1–3, 24, 27–29]. The second method is according to the distribution of trajectory data and uses the reverse compilation method of digital map to set personalized privacy requirements for trajectory points. This setting is a subjective estimate based on the data itself [19, 21–23].

We use the first method to simulate the real trajectory data by adding personalized privacy requirements into TDs. According to different density of trajectory data, we set different personalized anonymous threshold. In the sparse Trucks, $k_T \in [1, 10]$. In the dense Oldenburg, $k_O \in [1, 100]$. The range of desensitization threshold is uniformly restricted to $d \in [0, 1000]$. The maximum distortion threshold $d_{max}$ is uniformly set as 8000.

Generally, the level of privacy requirements is inversely proportional to their share of TD. That is to say, there are more trajectory records with low privacy requirements, while there are fewer trajectory records with high privacy.
requirements. Similar to [27, 30], Zipf distribution is used to generate personalized privacy requirements that meet the following probability density function:

\[ P(r) = \frac{C}{r^\alpha}, \]

where \( r \) represents the level of privacy requirements; the larger \( r \) is, the stronger privacy requirement is. \( P(r) \) represents the proportion of trajectory points in the whole trajectory data, in which the intensity of privacy requirements of these trajectory points is equal to \( r \). \( \alpha \) represents the distribution density of the trajectory points with different privacy requirements, and the density of the distribution increases with \( \alpha \). \( C \) is an empirical constant; the function is to ensure that the sum of proportions of the trajectory points with different privacy requirements amounts to 1. In Table 3 we list \( C \) and the median of generated data with different \( \alpha \) and different ranges of \( r \).

Figure 16 shows the ratio of different sample values when \( r \in [1, 10] \) and \( \alpha \) is 1, 0, and –1, respectively.

In the first group of experiments, the baseline algorithms NWA and CPPP do not need to desensitize trajectories. In order to compare the baseline algorithms with our algorithm, we constructed six TDs based on the two original TDs, which only contain personalized anonymous requirements but not personalized desensitization requirements.

Obviously, when \( \alpha \) is 0 or –1, the personalized privacy requirements generated for the trajectory points do not match the facts. However, it is beneficial to test the performance of algorithms on TDs with different characteristics of privacy requirements. The parameters for constructing these six TDs are shown in Table 4.

In the second group of experiments, in order to detect our proposed algorithm PRK and PRU on multiple TDs, we set different distribution parameters \( \alpha \) for the privacy parameters \( (k, d) \) on different original TDs and constructed eighteen TDs with personalized privacy requirements. The list of parameters for constructing \( \text{TD}_{7-15} \) from Truck is shown in Table 5, in which \( k_T \in [1, 10] \) and \( d_T \in [0, 1000] \). The

**Table 2: Characteristic parameters of the two TDs**

<table>
<thead>
<tr>
<th>TD</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p) Area (km²)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
<th>([\text{TD}]_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trucks</td>
<td>112203</td>
<td>273</td>
<td>2403.30</td>
<td>79208</td>
<td>91.04</td>
<td>5.89 \times 10^{-4}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oldenburg</td>
<td>4780954</td>
<td>100000</td>
<td>634.49</td>
<td>4350</td>
<td>93.21</td>
<td>1.73 \times 10^{-4}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 14:** Estimation process of privacy protection parameters in different algorithms.

**Figure 15:** Two examples of typical trajectory data as the basic data for constructing trajectory data with personalized privacy requirements. (a) The illustration of Trucks. (b) The illustration of Oldenburg.

![Figure 14](image-url)

![Figure 15](image-url)
The sample proportion

<table>
<thead>
<tr>
<th>ID</th>
<th>Original data and the range of k</th>
<th>Distribution parameter of k</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD_1</td>
<td>Trucks, ( k_T \in [1, 10] )</td>
<td>( \alpha = -1, C = 1.81E - 02 )</td>
</tr>
<tr>
<td>TD_2</td>
<td>Trucks, ( k_T \in [1, 10] )</td>
<td>( \alpha = 0, C = 1.00E - 01 )</td>
</tr>
<tr>
<td>TD_3</td>
<td>Trucks, ( k_T \in [1, 10] )</td>
<td>( \alpha = -1, C = 3.41E - 01 )</td>
</tr>
<tr>
<td>TD_4</td>
<td>Oldenburg, ( k_O \in [1, 100] )</td>
<td>( \alpha = -1, C = 1.98E - 04 )</td>
</tr>
<tr>
<td>TD_5</td>
<td>Oldenburg, ( k_O \in [1, 100] )</td>
<td>( \alpha = 1, C = 1.00E - 02 )</td>
</tr>
<tr>
<td>TD_6</td>
<td>Oldenburg, ( k_O \in [1, 100] )</td>
<td>( \alpha = -1, C = 1.92E - 01 )</td>
</tr>
</tbody>
</table>

Table 4: Generation parameters of the data in the first group of experiments.
parameters for constructing $TD_{16-24}$ from Oldenburg are the same as those in Table 5, in which $k_O \in [1, 100]$ and $d_O \in [0, 1000]$.

6.4. Results and Analysis of Experiment

6.4.1. Data Distortion. In the first group of experiments, we execute three baseline algorithms and our algorithms on constructed $TD_{1-6}$.

The $TD^*$'s generated by the privacy protection algorithm are compared with the constructed $TD$s, and the information loss under different algorithms and constructed $TD$s is obtained by formula (12) and shown in Figure 17.

In Figure 17, $\alpha$ is the distribution parameter of personalized anonymous requirements, and the distribution status of samples is shown in Figure 16. $\log(\text{Infloss})$ is a logarithmic form of information loss. This expression has two advantages: first, it can clearly distinguish the magnitude difference in data distortion between the trajectory-based method and the method based on trajectory points; second, it can reflect the slight difference between similar methods.

To the $TD$s with personalized privacy requirements, the data distortion generated by the personalized algorithm is of the same magnitude as that generated by the non-personalized algorithm (NWA versus CPPP and IPKN versus TPPP), while these two kinds of algorithms have the same granularity (trajectory or trajectory point). However, we must recognize that nonpersonalization algorithms (NWA and IPKN) select the median of personalization privacy parameters as privacy requirements. This also means that 50% of the trajectory points fail to meet personalized privacy requirements.

When $TD$s with personalized privacy requirements are processed by the nonpersonalized algorithm, if the maximum privacy requirement is regarded as the algorithm parameter in order to meet the privacy requirements of all trajectory points, the obtained $TD^*$'s will cause serious data distortion.

$\text{Infloss}$ increases as $\alpha$ decreases. This is because the distribution parameter $\alpha$ is negatively correlated with the distribution of privacy requirements. As $\alpha$ decreases, the distribution of sensitivity threshold $d$ and anonymous threshold $k$ gradually tends to larger values, so the distortion increases. The following illustration of the experimental results also shows a similar phenomenon.

In addition, the $TD^*$ generated by privacy protection algorithm based on trajectory anonymity will cause serious data distortion. However, the distortion of $TD^*$ generated by privacy protection algorithm based on anonymity and desensitization of trajectory points is two orders of magnitude less.

In the second group of experiments, we execute our algorithms on constructed $TD_{7-24}$.

Since both algorithms are based on translocation of the original trajectory points, the distortions of the $TD^*$ generated by them are similar. The difference is mainly in efficiency and adaptability. A small amount of variation in data distortion results from the uneven distribution of the original data. The data distortion on different data is shown in Figure 18.

6.4.2. Result of TDD. In the test of mobility patterns, we refer to TDU and TDD. The result of TDU is the same as $\text{Infloss}$, so it is not necessary to draw again. Only the results of TDD, which represent the directional similarity between trajectories, are shown here.

(1) TDD of the $TD^*$'s in the First Group of Experiments. We calculated the TDD with different algorithms and parameters on the two sets of experimental data generated from different original data; Figure 19 shows how the TDD keeps trajectories mobility patterns.

As can be seen from Figure 19, since the replaced trajectory point is found on the same time slice, the published trajectories generated by the trajectory clustering algorithms (NWA and CPPP) maintained the mobility patterns.
properly. However, the privacy protection algorithm based on trajectory points exchange leads to a large mobility patterns deviation. On the experimental data based on different original data, the performances of the algorithms are similar, and the effect of mobility patterns maintenance is shown in the following order: NWA > CPPP > TPPP > IPKN.

(2) TDD of the TD* in the Second Group of Experiments. To compare PRK with PRU, we tested them on experimental trajectory data with different distribution of privacy requirement parameters \((k, d)\) and calculated the TDD of the published trajectory data.

As can be seen from Figure 20, there is no significant difference between the performances of TDD of the two algorithms on different data set, but they become worse as the distribution parameter \(ak\) decreases, while there is no significant rule or difference on TDD of PRK and PRU within each trajectory data.

From the above experiments, we can see that our algorithm is not outstanding in mobility patterns retention but performs better in the similarity of trajectory distance and information loss. It means that we gain an advantage in the distance similarity of trajectory at the cost of the loss of mobility patterns.

6.4.3. Efficiency of Our Algorithms. In the data distortion experiment, we select different TDs, algorithms, and parameters. We also measure the execution time of those experiments.

(1) The Time Efficiency of the Algorithms in the First Group of Experiments. It can be seen from Figure 21(a) that the
The execution time of IPKN is the largest, which is caused by the iteration that constructs the \( k \)-core subnet. However, the execution time for NWA and CPPP is relatively less because the time for calculating the trajectory relation network of TD is not included. PRK algorithm needs to calculate the CTPTs for each trajectory point based on its personalized privacy requirements, so the time consumption is more than PRU. PRU can construct RNTD at one time, so it saves a lot of time.

As shown in Figure 21, due to the different data size in different TDs, the differences in execution time are obvious. The efficiency of the algorithm is basically the same under different privacy requirements on the same TDs. This sublinear relation between the efficiency and the data scale is beneficial to the analysis and processing of large-scale TDs.

As shown in Figure 22, the efficiency of both algorithms is not sensitive to privacy distribution but is only related to data size. PRU is more efficient because it can calculate CTPTs of all trajectory points at one time, and its premise that the adversary does not grasp the personalized privacy requirements of the target is also consistent with reality. Therefore, PRU algorithm is an ideal privacy protection algorithm by the granularity of trajectory points, and it also provides users with the function of setting personalized privacy requirements.
7. Conclusion

We propose a Translocation-based Personalized Privacy Protection (TPPP) method for publishing trajectories. Our algorithm abandons the clustering method that takes trajectories as the privacy protection objects and avoids the serious data distortion caused by the curse of dimensionality. TPPP achieves personalized anonymity and desensitization of trajectory points while retaining all original trajectory points. According to whether the privacy requirements are known, the algorithm can be divided into PRK and PRU. The two algorithms prove to be feasible by running on the trajectory data with personalized privacy requirements, respectively, when compared with the baseline algorithm. We also propose a modified Manhattan distance metric to measure the distance between trajectory points in the space-time region, which improves Euclidean distance in measuring the reachable distance between two points.

Although the algorithm in this paper improves the availability of published trajectory data, it gives up trajectory anonymity and reduces the level of privacy protection. This trade-off between the availability and privacy of published data does limit admittedly the applicability of the proposed algorithm.

In this article, we assume that the trajectory data is published statically. In fact, as SA receive LSR continually, the trajectory data is updated constantly. The dynamic release of trajectory data is inevitable. How to ensure that the combination of trajectory data released in succession does not divulge privacy and how to improve the efficiency of dynamic publishing of trajectory data are problems to be investigated in the future.

Data Availability

All data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors gratefully thank Dr. Yu Xin who put forward some opinions about the algorithm design, http://chorochronos.datastories.org that released the experimental data, and Nanni M. who provided download links on the experimental data. This work was supported by the National Science Foundation of China (NSFC) under Grant no. 61602133, Development Program of Science and Technology of Jilin Province of China under Grant no. 20180519012JH, Scientific Items of Jilin Provincial Department of Education under Grant no. JJKH20191025KJ, and Science and Technology Development Plan Project of Siping (nos. 2016058 and 2017098).

References


