Hindawi
Mobile Information Systems
Volume 2020, Article ID 3429256, 12 pages
https://doi.org/10.1155/2020/3429256

Research Article

A Privacy-Preserving Trajectory Publication Method Based on Secure Start-Points and End-Points

Yannian Zhao,[1,2] Yonglong Luo ,[1,2] Qingying Yu,[1,2] and Zhaoyan Hu[1,2]

[1]*School of Computer and Information, Anhui Normal University, Wuhu, Anhui, China*
[2]*Anhui Provincial Key Laboratory of Network and Information Security, Wuhu, Anhui, China*

Correspondence should be addressed to Yonglong Luo; ylluo@ustc.edu.cn

Received 19 September 2019; Revised 21 January 2020; Accepted 1 February 2020; Published 21 April 2020

Academic Editor: Nicola Bicocchi

By judging whether the start-point and end-point of a trajectory conform to the user's behavioral habits, an attacker who possesses background knowledge can breach the anonymous trajectory. Traditional trajectory privacy preservation schemes often generate an anonymous set of trajectories without considering the security of the trajectory start- and end-points. To address this problem, this paper proposes a privacy-preserving trajectory publication method based on generating secure start- and end-points. First, a candidate set containing a secure start-point and end-point is generated according to the user's habits. Second, $k-1$ anonymous trajectories are generated bidirectionally according to that secure candidate set. Finally, accessibility corrections are made for each anonymous trajectory. This method integrates features such as local geographic reachability and trajectory similarity when generating an anonymized set of trajectories. This provides users with privacy preservation at the k-anonymity level, without relying on the trusted third parties and with low algorithm complexity. Compared with existing methods such as trajectory rotation and unidirectional generation, theoretical analysis and experimental results on the datasets of real trajectories show that the anonymous trajectories generated by the proposed method can ensure the security of trajectory privacy while maintaining a higher trajectory similarity.

1. Introduction

With the development of wireless communication and positioning technology, users can publish their trajectories to obtain convenient point-of-interest information [1]. For example, smartphone applications can prompt the user with catering and accommodation suggestions by analyzing their historical trajectory information. Similarly, a municipal department can collect trajectories of taxi routes for the purpose of road network construction [2].

While enjoying the convenience offered by trajectory release, users also face the danger of privacy leakage [3]. The trajectory information published by the user often contains vital spatiotemporal information, which the malicious attacker can obtain by analyzing the trajectory posted by the user. For example, when a user frequently accesses sensitive locations, the attacker can analyze the privacy information relating to the user's interests, health conditions, and the like through information acquired on the sensitive location. Users concerned with privacy breaches may refuse to publish their trajectory information, which has become a key issue that constrains the development of location-based service (LBS). Many users think that deleting the quasi-identifiers (such as name, ID number, and phone number) in their trajectory information can protect their privacy. But it turns out that this assumption is wrong. Sweeney et al. [4] conducted experiments using 1990 US census data and found that in combination a few features can uniquely identify certain individuals in the population. The experimental results showed that 87% of the US population can be uniquely identified by gender, date of birth, and postal code. An attacker can also obtain other spatial and temporal attributes from the trajectory data with the purpose of identifying the user's private information. Advances in location-based services urgently require new privacy programs to protect and support the user.

The academic community has proposed a significant number of methods to protect trajectory privacy, including the location-based query method that incorporates a secure and trusted third party [5] and the privacy-preserving trajectory publication method to protect privacy through generalization processing and point suppression [6]. It should be noted that the above methods merely involve semantic privacy protection for stay points. Most privacy protection methods based on generalization do not take into account the security of the trajectory start- and end-points after generalization. In fact, start- and end-points can contain more information (i.e., user's place of employment) due to new developments in mobile technology. Therefore, it is necessary to adopt different trajectory privacy protection strategies for different motion modes. More importantly, increasements in the semantic information contained in trajectories have posed greater challenges for user privacy protection.

The main contributions of this paper are as follows:

(1) To generate an algorithm for a secure start- and end-points candidate set. According to the user's habits and the characteristics of the trajectory start- and end-points, a secure candidate points set is selected, and the start- and end-points of the dummy trajectory are selected.

(2) To propose a method of fitting trajectories, which is to fit two historical trajectories in different directions into one dummy trajectory. The dummy trajectory obtained by fitting can maintain an effective difference with the original trajectory.

(3) To evaluate the algorithm by different degrees between the trajectories, the trajectory data availability, and the trajectory leakage probability. The theoretical analysis and experiments show that the proposed method ensures the high availability of trajectory data under the premise of effectively protecting the privacy of users.

The content of the paper is arranged as follows: Section 2 introduces the related work of trajectory privacy protection along with related concepts and definitions used in the article; Section 3 introduces the personalized trajectory privacy protection plan proposed in the study; Section 4 includes the relevant experimental data and analysis of results; the fifth section is the summary and the next work introduction.

2. Related Work

At present, privacy protection methods based on trajectory release are typically divided into three categories: suppression method, generalization method, and dummy data method. Among them, the generalization method with k-anonymity at its core is the most widely used. The k-anonymity technique was first proposed by Sweeney et al. [7] in the proceeding of IEEE Security and Privacy in 1998. For the problem of attribute link attack, the technique proposes to hide the user's sensitive information in a set, so that the generalized user information will not be recognized by the attacker. In view of the fact that ordinary k-anonymity methods cannot resist background knowledge attacks and homogeneous attacks, Machanavajjhal et al. [8] proposed an l-diversity model that defines a diversified standard, which satisfies any anonymous set (minimum bounding rectangle).

Since the generation of k-anonymity technology originates from the privacy protection problem of relational databases, the quasi-identifiers and sensitive attributes in the database are easy to define. But when k-anonymity is applied to the high-dimensional data field of trajectory, these attributes become difficult to define [9]. For the first time, Gruteser and Grunwald [10] applied the k-anonymity technique to LBS and proposed the concept of location anonymizing. With the development of LBS, many trajectory privacy protection methods based on k-anonymity have emerged. For example, Abul et al. [11] proposed a method of mapping the trajectory of an object's motion in a three-dimensional cylinder. The uncertainty of the trajectory motion of the cylinder and another cylinder with the same uncertainty interval meant that the two trajectories cannot be distinguished from each other. Recently, Zhang et al. [12] proposed a particle swarm optimization anonymization algorithm for attribute generalization. A Venot diagram is established based on the entropy value to divide user's locations to achieve the generalization of sensitive locations.

The generalization method is less effective in providing suitable protection for user privacy in the face of single trajectory data with sensitive attributes. Sui et al. [13] analyzed the risk of trajectory data in a campus with a single trajectory diversity and provided quantitative proofs of the low diversity risk in the data. Their experiments demonstrate that, in the case of low user privacy diversity, personalized privacy protection methods are needed to compensate for the risk of privacy leakage caused by the lack of diversity.

Common personalized trajectory privacy methods often incorporate k-anonymity techniques to form a model with multiple methods to meet the specific needs of users. For example, by performing geometric rotation and translation of the real trajectory to obtain a new dummy trajectory and a set of k-anonymity [14], the shape similarity of the anonymous concentrated trajectory is highly uniform, so that the attacker cannot mine the real trajectory through the human walking model. Li et al. [15] proposed an improved trajectory rotation scheme, which makes the dummy trajectory obtained by rotation more closely resemble reality, enhancing the degree of privacy protection of anonymous collections. At the same time, the original trajectory is used as a template to generate dummy trajectories through geometric transformation. The method of trajectory through geometric transformation [16] can make the pseudo-trajectory pass the same position as the real trajectory, while maintaining a safe distance so as not to be easily recognized by the attacker. The trajectory generated by the geometric method often does not conform to human behavior. Xiao et al. [17] proposed a method based on graph partitioning, which draws on the similarity of temporal overlap, constructing a graph model of the trajectory, and implementing k-anonymity by k-node segmentation. In addition,

Komishani et al. [18] proposed a personalized privacy protection method—personalized privacy in trajectory data (PPTD)—combining the generalization and suppression methods. Based on the combination of sensitive attribute generalization and local suppression of trajectories, the optimal choice between data availability and data privacy security is selected. In particular, people's real activity space is covered by the network structure composed of roads; the anonymous locations obtained according to Euclidean space may be located in some areas that are unreachable or easily recognized by attackers. So it cannot achieve the purpose of effectively generalizing the user's real locations. Many mix-zone-based methods were proposed to protect trajectory privacy in the road network environment [19, 20].

Differential privacy is a new privacy-preserving technique based on data distortion, which protects the sensitive data and keeps the statistical properties of data by adding random noise [21–24]. Combining differential privacy with other technologies is a hot topic in recent research. For example, Wang et al. [25] proposed a differential privacy-preserving method based on information entropy to generalize sensitive areas. However, as a kind of high-dimensional data, trajectory has complex semantic information. Therefore, it is very difficult to select appropriate noise for trajectory data.

In summary, some research results have been achieved in the use of personalized privacy protection, but most methods ignore the risk of privacy leakage of attackers based on the trajectory data. As shown in Figure 1, two dummy trajectories are generated based on user's dataset. Trajectory A is generated from a school zone (Anhui Normal University); trajectory B is generated from a housing estate and only guarantees the shape similarity to trajectory A. Although both trajectories have similar directions and shapes, when the attacker knows that the user identity is a student, the start-point of trajectory B is easily determined.

Addressing this problem, we first generate a candidate set of secure start- and end-points according to user habits and propose a dummy trajectory generation method that fits the trajectory from these secure points, effectively defending against attacks based on background knowledge.

3. Preliminaries

3.1. Trajectory Data Model

Definition 1. Original trajectory: refers to a path that the user has passed at a certain time. Specifically, original trajectory is a set of ordered locations:

$$T = \{(t_1, x_1, y_1), \ldots, (t_n, x_n, y_n)\}, \tag{1}$$

where (t_i, x_i, y_i) is the i-th position of T and t_i is the timestamp when the user is in position (x_i, y_i). n represents the number of position points in the trajectory T $(1 \le i \le n)$. All trajectories of a user are represented by a set U.

Definition 2. Dummy trajectory: refers to a path obtained by camouflaging the user's original trajectory using an anonymous algorithm. The dummy trajectory is expressed in the following formula when the path is generated anonymously according to the original trajectory:

$$T^d = \left\{\left(t_1^d, x_1^d, y_1^d\right), \ldots, \left(t_n^d, x_n^d, y_n^d\right)\right\}, \tag{2}$$

(t_i^d, x_i^d, y_i^d) is the location of user's d-th dummy trajectory T^d when the timestamp is $t_i^d \cdot (1 \le d \le k)$, and k is the number of T^d in the dummy trajectory set.

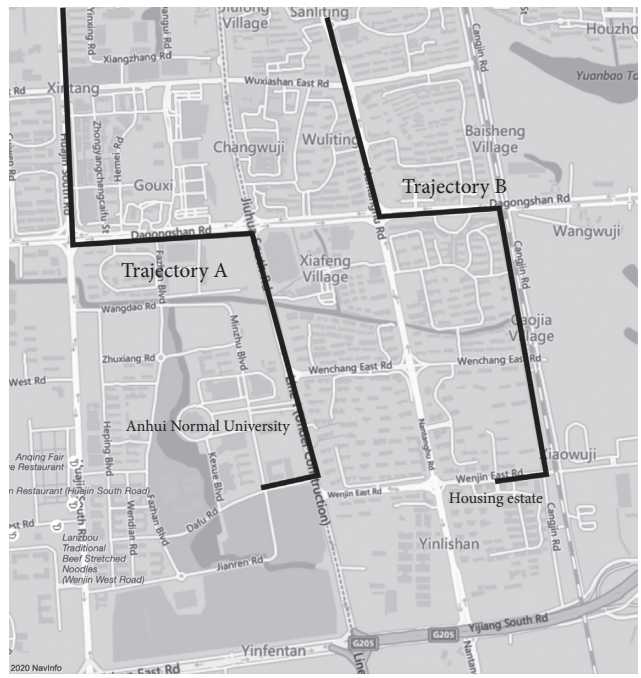

Figure 1: The start-point of the trajectory is seen by the attacker.

Definition 3. Secure start- and end-points. A candidate set containing secure start- and end-points is generated according to the user's habits:

$$\mathrm{ST} = \left\{\left(x_1^s, y_1^s\right), \ldots, \left(x_j^s, y_j^s\right)\right\}, \tag{3}$$

$$\mathrm{ED} = \left\{\left(x_1^e, y_1^e\right), \ldots, \left(x_j^e, y_j^e\right)\right\}, \tag{4}$$

where ST stands for a secure start-points candidate set and ED stands for a secure end-points candidate set, j represents the number of points in the candidate set, and (x_i^s, y_i^s) or (x_i^e, y_i^e) represents the i-th point of the secure start-(end-) points candidate set.

3.2. Related Concepts

Definition 4. Trajectory difference degree: defined as the degree of differences between the generated dummy trajectory and the original trajectory and calculated by a degree function. The difference degree is an important standard for judging the merits and demerits of dummy trajectory generation algorithm. Suppose that A_1, A_2, A_3 contains three continuous position points in the dummy trajectory; the trajectory difference degree [16] is expressed as

$$\sigma = \frac{1}{(k-1)m} \sum_{j=1}^{k-1} \sum_{j=1}^{m} \left| \theta_{\text{real}}^{i} - \theta_{j}^{i} \right|, \tag{5}$$

where k is the number of dummy trajectories and θ represents the direction angle of the i-th position in the trajectory. An example of θ creation can be seen in Figure 2, m is the number of direction angles in a trajectory, and θ creation [16] is expressed as follows:

$$\theta = \arccos\left(\frac{\overrightarrow{A_1A_2} \cdot \overrightarrow{A_2A_3}}{\left|\overrightarrow{A_1A_2}\right| \left|\overrightarrow{A_2A_3}\right|} \right). \tag{6}$$

Definition 5. Trajectory leakage probability: refers to the percentage of dummy trajectories recognized by an attacker.

Figure 3 shows the location security based on query probability. The area (location map) is divided into a grid of $n \times n$ cells. Different shades of cells represent different query probabilities. We use the method proposed in [26] to divide the area. Step (1): Divide the user activity area into 10×10 blocks. Step (2): If the number of positions in a block exceeds the threshold (depends on the number of points in the dataset), the block needs to be further divided. Step (3): Repeat Step (2) until the number of points in each block is less than the threshold. Suppose the query probability of a region where the i-th position point of T^d is located with q_i^d; then the query probability corresponding to the original trajectory T position point is q_i. If $|q_i^d - q_i| \le \Delta$, we can determine that the location is safe (Δ is a positive number which is very close to 0). Suppose T^d has a total of δ unsafe location points; then the trajectory leakage probability [26] is defined as

$$\rho = \frac{1}{\left(k - \sum_{1}^{k} (\delta/n)\right)}, \tag{7}$$

where k is the number of trajectories in an anonymous set and n is the number of position points in the trajectory.

Definition 6. Utility loss: it is composed of two aspects: (1) the loss or redundancy of location points due to the trajectory splicing. It is calculated as

$$w = \frac{\left| T\_w_{\text{dummy}} - T\_w \right|}{T\_w}, \tag{8}$$

where T_w_{dummy} is the number of location points in the dummy trajectory and T_w is the number of location points in the real trajectory.

(2) The information loss caused by generalization. It is usually measured by the standard calculation method proposed in [27]:

$$\text{LI} = \frac{\left[\sum_{i=1}^{n} \sum_{i=1}^{f} (1 - 1/\text{area}(\text{ST\_or\_ED}))\right]}{(n \times f)}, \tag{9}$$

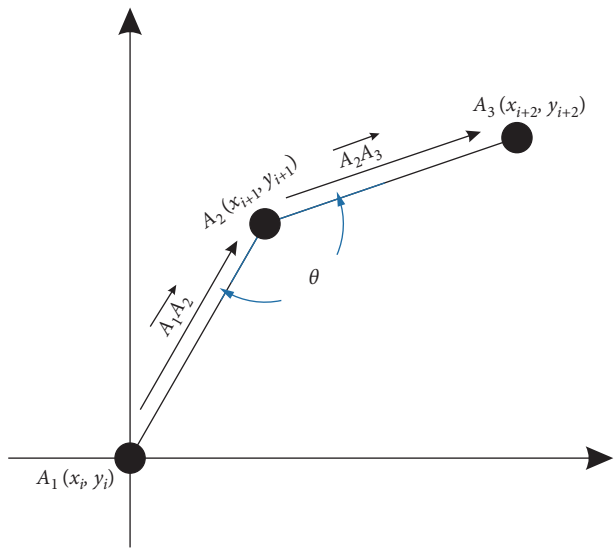

Figure 2: Trajectory angle calculation.

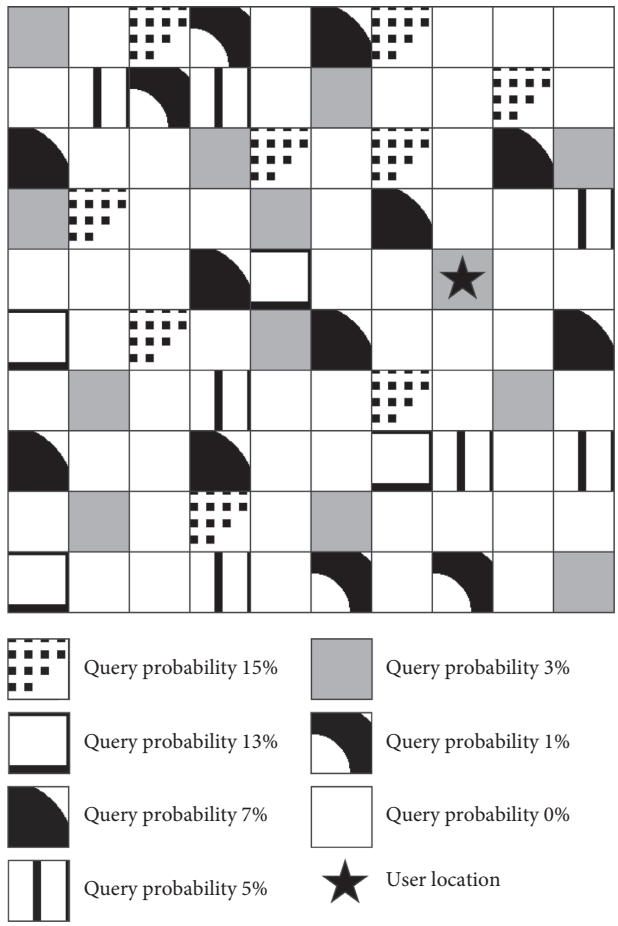

Figure 3: Cell grids to determine whether the location is safe based on the location query probability.

where n is the number of trajectories in user activity area, f is the number of semantic locations, and area(ST_ED) is the number of points in ST or ED.

4. Trajectory Generation Algorithm

Because an attacker may recognize the start- and end-points of the trajectory and find the original trajectory, we propose a dummy trajectory generation method that includes the following steps: (1) generate secure start- and end-points according to the user's dataset; (2) generate dummy trajectories according to the set developed in the previous step; (3) perform a reachability correction on generated dummy trajectories. The three steps are, respectively, implemented by the subsequent Algorithms 1–3. A schematic diagram for the proposed method is shown in Figure 4, which overviews the relationship between these three algorithms.

4.1. Generation of Secure Start- and End-Points. This section presents an algorithm for generating secure start- and end-points. First, we divide the time of day into 10-minute cycles to get a set of time intervals $\mathrm{TPS} = \{\mathrm{Tps}_1, \ldots, \mathrm{Tps}_{144}\}$ (lines 1-2). Second, an active area is divided into φ blocks according to Figure 3 (line 3). Third, we sort $\mathrm{TempTPS}_i$ by the number of timestamps within each time period in an ascending order and store in the set $\mathrm{HFTps}_i\ (1 \le i \le \varphi)$ (lines 5–13). Finally, we get the set of points in the i-th block from HFTps_i (line 14).

Based on Algorithm 1, a set of secure start- and end-points for any user can be obtained. For example, t_1 is the moment when the user's original trajectory starts and $t_1 \in \mathrm{Tps}_{32}$. Search Tps_{32} in $\{\mathrm{HFTps}_1, \ldots, \mathrm{HFTps}_{144}\}$ to get all subsets containing Tps_{32}. Randomly select a $\mathrm{HFTps}_i\ (1 \le i \le \varphi)$ that meets the requirements. Output its corresponding set of position points SL_i as a secure start-points set for the dummy trajectory. Similarly, a secure end-points set can be obtained.

4.2. Bidirectional Dummy Trajectory Generation Algorithm. This section proposes a bidirectional dummy trajectory generation algorithm, which extracts historical trajectories in the dataset and filters these through the secure start- and end-points according to the overall direction of the trajectory. As shown in Figure 5, by fitting the coincident trajectory segments, it is possible to combine two historical trajectories into one dummy trajectory.

4.3. Dummy Trajectory Correction Algorithm. In the T_{dummy} trajectory generated by Algorithm 2 (shown in Figure 6), the transition probability between positions becomes a tool for the attacker to analyze the anonymous trajectory. When the transition probability between positions A and B is low, the trajectory generated by the process is fitted and will be seen by the attacker. How to prevent an attacker from inferring the user's real trajectory by analyzing historical data is the key problem to be solved in this section.

If the probability that the trajectory of the user's dataset U passes through both points A and B is greater than the position transition probability threshold V, it indicates that the T_{dummy} trajectory generated by Algorithm 2 has a good ability to resist background knowledge attacks. The threshold V is determined according to the user's demand for privacy protection. The larger the V, the better the anonymity effect. As shown in Figure 7, when the transition probability between positions A and B is lower than V, the accessibility correction of the trajectory is completed by correcting the trajectory $A \longrightarrow B$ to $C \longrightarrow D$. The reachability determination is performed on the trajectory $C \longrightarrow D$ again, and if the threshold condition is still not satisfied, the return to Algorithm 2 recreates the trajectory.

4.4. Algorithm Analysis

4.4.1. Complexity Analysis

(1) Algorithm 1 The generation of secure start- and end-points. The for-loop is used to traverse the φ block area, and the user trajectory points are divided, classified, and counted three times. Therefore, the time spent is φ, and the time complexity is $O(\varphi)$.

(2) Algorithm 2 Bidirectional dummy trajectory generation algorithm. The algorithm includes two layers of for-loops and one layer of while-loops in the trajectory fitting stage. In the worst case, the algorithm takes $(n^2 + n + n^3)$ and the time complexity is $O(n^3)$.

(3) Algorithm 3 Accessibility correction algorithm. The algorithm consists of four for-loop sequences, so the time spent is $4n$ and the time complexity is $O(n)$.

In summary, the overall time complexity of the algorithm is $O(\mathrm{n}^3)$.

4.4.2. Security Analysis. In this paper, a dummy trajectory generation algorithm based on secure start- and end-points is designed to generate $k-1$ dummy trajectories for each real trajectory in the original dataset. It has the following characteristics, which realizes the privacy protection of trajectory data release.

First, the trajectory contains a large number of sensitive semantic locations, taking into account the attacker's acquisition of the semantic location information of the map through background knowledge. When an attacker analyzes historical data, the breach can be discussed by way of three processes:

(1) Relational inference attack: the attacker can obtain the transition probability between key areas by observing the transfer of user trajectories and combining this with background knowledge such as historical trajectory data. Since the dummy trajectory is generated by splicing between historical trajectories, it conforms to the user's habits and can resist a semantic area attack.

(2) Similarity attack: the attacker can target the trajectory set according to semantic and geographical similarity. Since the dummy trajectory is generated by historical trajectory fitting, the requirements of semantic similarity are satisfied. When the dummy

Input: user's trajectory dataset U. number of blocks φ;
Output: a set of secure points SL.
(1) Divide the user activity area into $\{\text{block}_1, \text{block}_2, \ldots, \text{block}_\varphi\}$;
(2) TPS $\longleftarrow \{\text{Tps}_1, \ldots, \text{Tps}_{144}\}$; /*144 time periods*/
(3) Get all points from U and partition them into different blocks;
(4) **for** $i \longleftarrow 1$ to φ **do**
(5) ts $\longleftarrow$ timestamps of all the points in block_i;
(6) **if** (!isempty (ts))
(7) Partition timestamps in ts into different time period of TPS and obtain the set TempTPS_i;
(8) Sort TempTPS_i by the number of timestamps within each time period in ascending order;
(9) $\text{HFTps}_i \longleftarrow$ Top five time periods in TempTPS_i;
(10) **end if**
(11) **if** (isempty (ts))
(12) Mark block_i as an unreachable area;
(13) continue;
(14) **end if**
(15) $\text{SL}_i \longleftarrow$ all the locations at the timestamps within HFTps_i;
(16) **end for**
(17) **return** SL;

Algorithm 1: The generation of secure start- and end-points.

trajectory is generated, the historical trajectory segment is selected by referring to the direction and shape of the real trajectory, so the spliced trajectory satisfies the requirements of geographical similarity.

(3) Stay-point attack: the start- and end-points of the trajectory can reflect key information such as the user's travel destination and are highly vulnerable to attack. Since the start- and end-points of the dummy trajectory are selected from the user's historical trajectory, they can resist the attacker's stay-point attack.

Secondly, the dummy trajectory is generated by referring to a segment of the real trajectory, and the similarity to the real trajectory that generates direction and shape is screened before fitting. Therefore, the dummy trajectory has a higher degree of similarity to the real trajectory. The attacker cannot recognize the real trajectory from the trajectory similarity among the set published by the user. In summary, the probability that any real trajectory is identified is that the k-anonymity privacy requirement is reached.

5. Experimental Evaluation

5.1. Experimental Environment and Dataset. The experimental environment of this evaluation is AMD Ryzen7 1700X Eight-Core Processor@ 3.4 GHz, 32 GB memory, the algorithm is implemented by MATLAB 2016b, and the program runs under Windows 10.

We have used the Geolife GPS Trajectories dataset of Microsoft Research Asia [28] to conduct the experiments. The Geolife dataset has trajectory data belonging to 182 users from April 2007 to August 2012. The dataset contains 24 million location points. Each location point package contains information such as latitude and longitude, altitude, and absolute time, with a total distance of about 1.2 million kilometers. The experiment extracts information from the dataset, including latitude, longitude, and time factors.

5.2. Experimental Results and Analysis. The parameters involved in the following experiments include the following: (1) trajectory difference degree, which is an important index to evaluate the pros and cons of generating dummy trajectory; (2) anonymity level k, that is, the number of dummy trajectories generated at one time; and (3) trajectory leakage probability, which addresses the probability that the real trajectory is seen by the attacker.

Schemes compared with this paper include the following: (1) the efficient trajectory privacy protection scheme (Rotation) [15]; (2) the pseudotrajectory privacy protection scheme based on spatiotemporal correlation in trajectory publishing (Round) [16]; and (3) the idealized k-anonymity model (Optimal); (4) k-CS algorithm: trajectory data privacy-preserving based on semantic location protection (K-CS) [21]; (5) on the privacy offered by (k, δ) – anonymity) ((k, δ) – anonymity) [29].

In order to improve the study's reliability, the following results are the average ones from 1000 repeated experiments.

5.2.1. Utility Loss. The splicing of the trajectory can result in the loss or redundancy of location points, producing a data loss, and the data loss threshold can guarantee the availability of trajectory data. When the dataset provides insufficient location points, it will produce a trajectory with higher data loss. This section verifies the relationship between the data quality of the generated dummy trajectory

Input: user's trajectory dataset U, secure points set ST, ED, trajectory $T = \{(t_1, x_1, y_1), \ldots, (t_n, x_n, y_n)\}$, anonymous parameter k, trajectory movement direction tolerance e, utility loss rate ϖ;
Output: dummy trajectory T_{dummy}.

(1) **for** $i \longleftarrow 1$ to length (ST) **do**
(2) **for** $j \longleftarrow 1$ to length (U) **do**
(3) **if** $((x_i, y_i) \in T_j(x_1, y_1))$;
(4) $T_{\text{ST}} \longleftarrow T_j$;
(5) **end if**
(6) **end for**
(7) **end for**
(8) Get T_{ED} based on the same method as shown in lines 1–7;
(9) $l_{\text{real}} = \sum_{i=1}^{n} x_i^{\text{real}} y_i^{\text{real}} n\overline{x^{\text{average}}}\,\overline{y^{\text{average}}} / \sum_{i=1}^{n} x_i^{\text{real}^2} - n\overline{x^{\text{average}}}^2$
(10) **for** $i \longleftarrow 1$ to length (T_{ST})
(11) **if** $((||l_{\text{real}}| - |l_{\text{dummy}}||)/(l_{\text{real}} > e))$
(12) Delete T_i from T_{ST};
(13) **end if**
(14) **end for**
(15) Handle T_{ED} based on the same method as shown in lines 10–14;
(16) **while** True
(17) Randomly select T_s, T_e from T_{ST} and T_{ED};
(18) **for** $i \longleftarrow 1$ to length(T_s) do
(19) **for** $j \longleftarrow 1$ to length(T_e) do
(20) **if** $((x_i^s, y_i^s) \in \text{area}(x_j^e, y_j^e))$
(21) mark $= i$;
(22) **break**;
(23) end if
(24) **end for**
(25) **end for**
(26) **end while**
(27) **for** $i \longleftarrow 1$ to mark do
(28) $\text{T}_{\text{dummy}} \longleftarrow \text{T}_{\text{dummy}} \cup (x_i^s, y_i^s)$;
(29) **end for**
(30) **for** $i \longleftarrow$ mark to length (T_e) do
(31) $T_{\text{dummy}} \longleftarrow T_{\text{dummy}} \cup (x_i^e, y_i^e)$;
(32) **end for**
(33) **if** $(|T_{\text{dummy}} - T|/T) > \varpi$
(34) Record Utility Loss and back to line 1
(35) **end if**
(36) **return** T_{dummy}

Algorithm 2: Bidirectional dummy trajectory generation algorithm.

and the size of the user's dataset. That is, when dummy trajectories are generated, the trajectory dataset satisfies the data loss threshold π, and it is considered that the dataset can generate a dummy trajectory with data loss below π. From the Geolife dataset, 10,000, 100,000, 200,000, 400,000, 800,000, and 1 million location points were selected for the experiments, and the k-value was 30 to calculate the minimum data loss threshold that each dataset can provide. Figure 8 shows the experimental results on data loss rate and utility loss rate.

As shown in Figure 8(a), when the number of datasets is less than 10,000 points, only a dummy trajectory with a data loss within 85% can be provided due to the limited quantity of data. When the number of datasets is less than 700,000 points, a data loss rate within 36% can be guaranteed.

We select 1 million location points from Geolife dataset for experiment and contrast with Algorithms k-CS and (k, δ)-anonymity. The utility loss caused by generalization is calculated by equation (9). As shown in Figure 8(b), utility loss rate increases as the privacy parameter k increases. When k is 12, the utility loss rate of k-CS is about 40%, while that of (k, δ)-anonymity is almost 80%. The main reason is that each location of the trajectory is generalized in the (k, δ)-anonymity method, and each semantic location of the trajectory is generalized in the k-CS method. In our method, only start-points, end-points, and sensitive semantic locations are generalized. Therefore, the utility loss of our method is lower than that of the other two methods. The result shows that our method meets the requirements of data availability.

5.2.2. Influence of the Number of User Trajectories on the Execution Time of the Algorithm. This section verifies the effect of the anonymity level k on the execution time of the algorithm and randomly selects a trajectory from the user trajectory set.

Input: uncorrected trajectory $T_{\text{dummy}} = \{(t_1^d, x_1^d, y_1^d), \ldots, (t_n^d, x_n^d, y_n^d)\}$, Position transfer threshold V, user's trajectory dataset U;
Output: dummy trajectory T_{dummy} that meets the requirements

(1) $count \longleftarrow 0$;
(2) **for** $i \longleftarrow 1$ to length (U) **do**
(3) **if** $(((x_1^d, y_1^d) \in U) \&\& ((x_n^d, y_n^d) \in U))$
(4) count++;
(5) **end if**
(6) **end for**
(7) **if** $(\text{count/length}(U) < V)$
(8) **for** $i \longleftarrow 1$ to mark do
(9) $T_{\text{dummy}} \leftarrow T_{\text{dummy}} \cup (x_i^e, y_i^e)$;
(10) **end for**
(11) **for** $i \longleftarrow$ mark to length (T_s) **do**
(12) $T_{\text{dummy}} \longleftarrow T_{\text{dummy}} \cup (x_i^s, y_i^s)$;
(13) **end for**
(14) **for** $i \longleftarrow 1$ to length (U) **do**
(15) **if** $(((x_1^d, y_1^d) \in U) \&\& ((x_n^d, y_n^d) \in U))$
(16) count++;
(17) **end if**
(18) **end for**
(19) **if** $(\text{count/length}(U) < V)$
(20) Delete T_{dummy} and back to Algorithm 2
(21) **end if**
(22) **else**
(23) **return** T_{dummy}
(24) **end if**

Algorithm 3: Accessibility correction algorithm.

The experimental results are presented in Figure 9.

In this scheme, the generation of dummy trajectories is only subjected to simple geometric changes and search replacement, and no high-dimensional operation is required, so the proposed algorithm has high execution efficiency. As can be seen from Figure 9, when the number of user trajectories reaches 20, the algorithm can still be executed within 1.2 seconds, in line with the needs of the actual application.

5.2.3. Comparison of Two-Way Fitting Algorithm and One-Way Generating Algorithm. When a traditional geometric method generates a trajectory, it usually generates only one dummy trajectory at a time. The trajectory generated by the traditional one-way generation algorithm has good similarity but often cannot resist semantic attacks and cannot guarantee privacy protection for users. The two-way fitting algorithm combines two unidirectional trajectories to obtain a completely new trajectory. Because of the secondary generation, the probability of being recognized by the attacker is reduced while ensuring the degree of trajectory difference. In order to prove the advantages of the two-way dummy trajectory generation algorithm, a trajectory is randomly selected from the user trajectory set for repeated experiments. In this section, the method proposed in this paper is compared with the trajectory generation algorithm of Round [15]. Figure 10 shows the trajectory difference

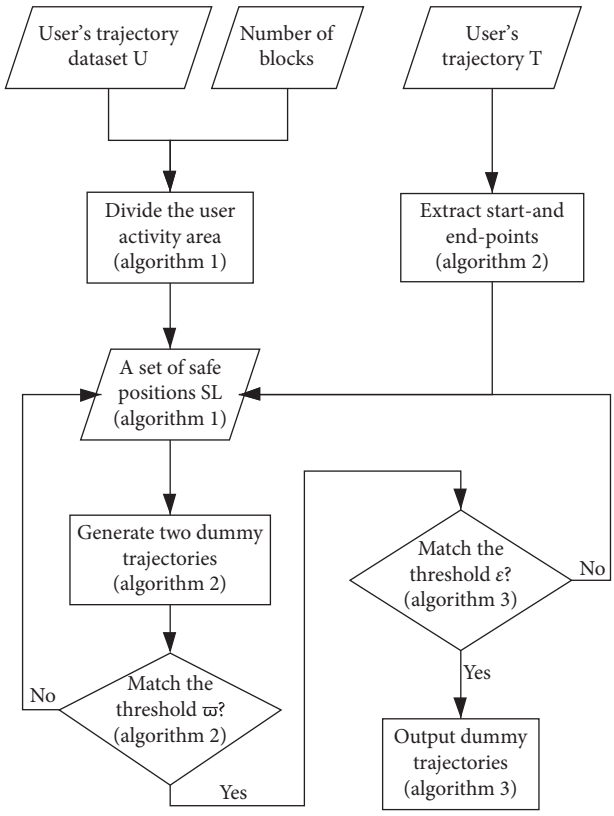

Figure 4: Schematic diagram of the proposed method.

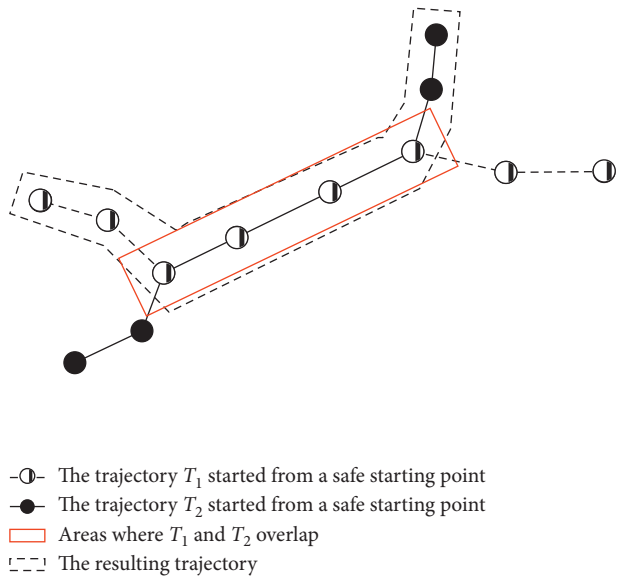

Figure 5: Bidirectional trajectory generation.

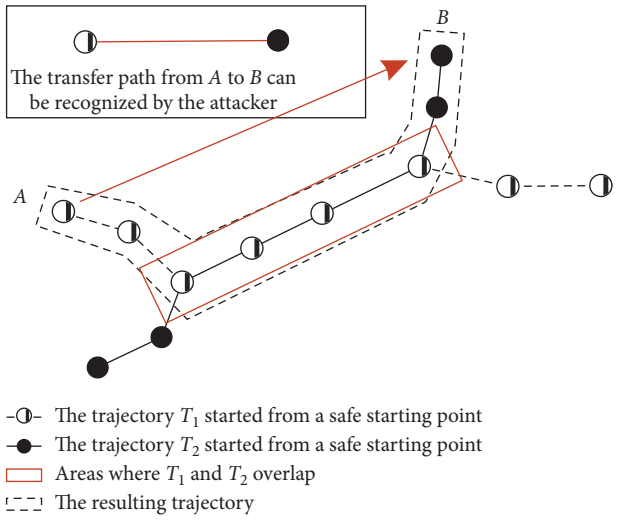

Figure 6: Trajectory reachability judgment.

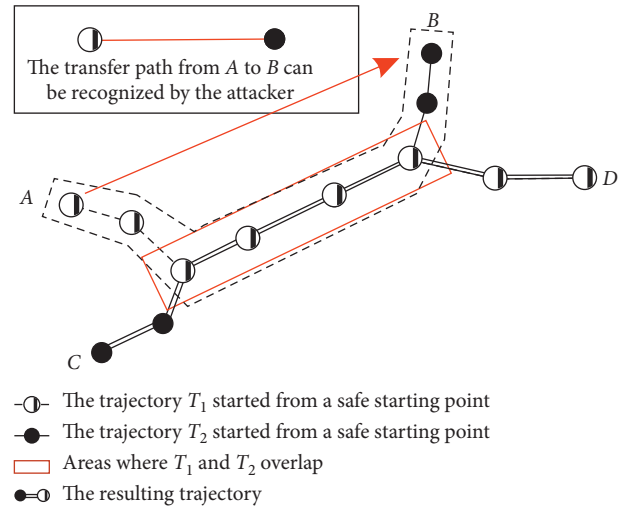

Figure 7: Trajectory accessibility correction.

comparison result with the change of k-value. The trajectory difference degree is calculated based on equation (5).

It can be seen from Figure 10 that, with the change of k-value, the difference between the set of dummy trajectory generated by the algorithm and the original trajectory is maintained at about 40%, which is slightly better than the Round [16] algorithm. The scheme can prevent an attacker from identifying a dummy trajectory according to the moving direction of the trajectory. Because the Round [16] algorithm generates dummy trajectories by means of geometric generation, it has good trajectory difference, but it has difficulties resisting semantic attacks. The proposed scheme has the advantage of the geometric generation method while resisting semantic attacks. In summary, the two-way fitting algorithm is more conducive to the protection of user trajectory privacy information.

5.2.4. Comparison of Trajectory Leakage Probability. The trajectory leakage probability reflects the degree of protection of the user's trajectory privacy. The lower the leakage probability, the higher the user's privacy protection. In order to evaluate the effect of the trajectory privacy protection method relating to start- and end-point security, this section compares the proposed algorithm with Rotation [15], Round [16], the idealized k-anonymity Optimal model, and the trajectory privacy leakage probability with the k-value. The result of the change is shown in Figure 11.

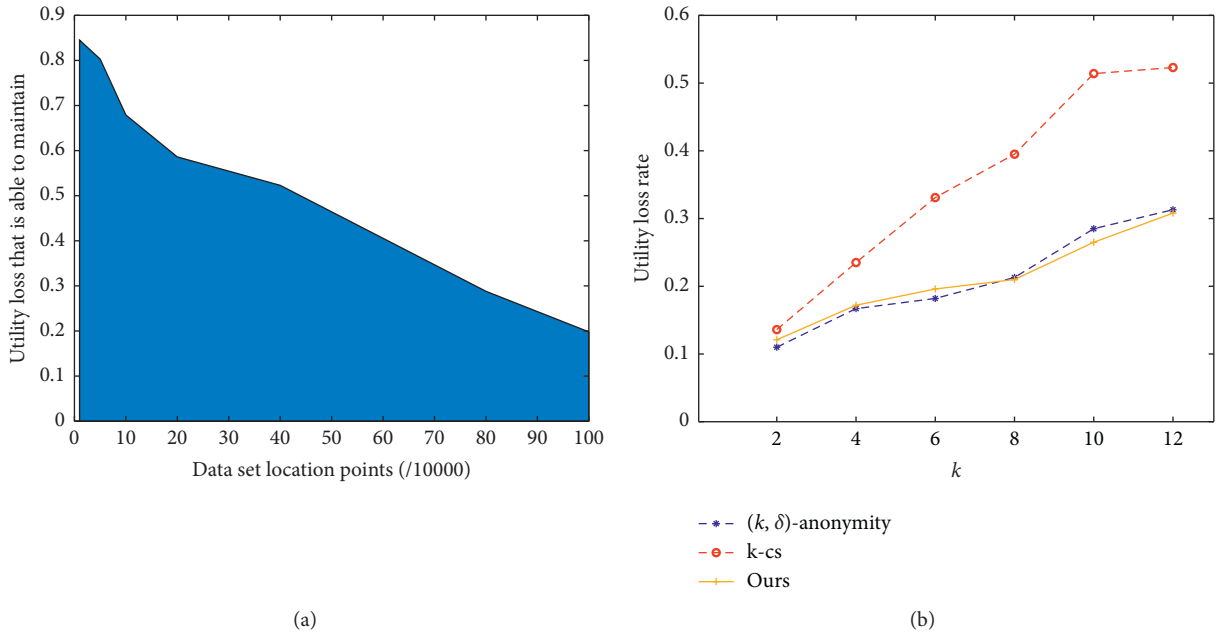

Figure 8: Loss rate. (a) Data loss rate. (b) Utility loss rate.

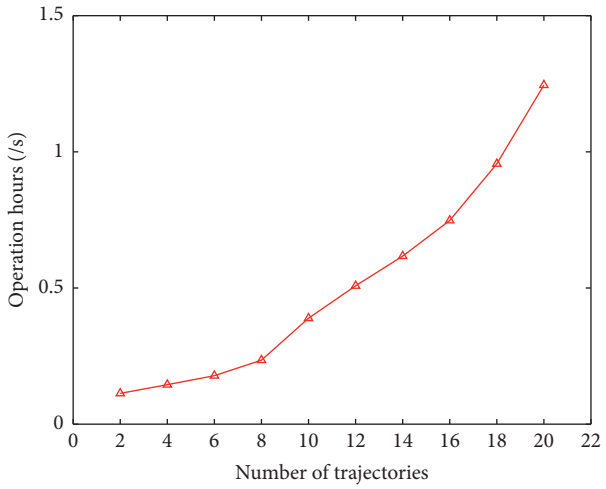

Figure 9: The effect of the number of trajectories k on the running time.

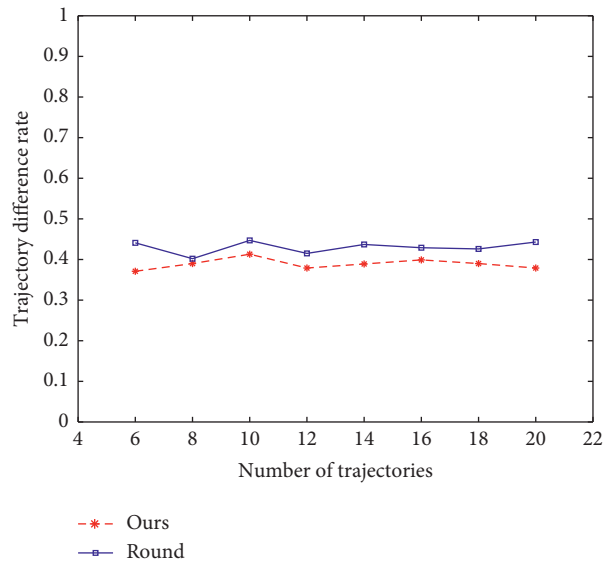

Figure 10: Single-item generation algorithm compared with the algorithm of this paper.

As shown in Figure 11(a), when the attacker only judges the background knowledge of the unreachable area, the trajectory leakages of the three schemes are basically the same. When the k-value is large enough, the trajectory leakage probability is infinitely close to the ideal state. Since the start- and end-points of the user trajectory data are easily acquired by the attacker, the attacker can find original trajectory according to the starting and ending point of the trajectory; for example, at moment t_1, the position of trajectory is block_{14}; when $t_1 \in \text{Tps}_{32}$ and $\text{Tps}_{32} \notin \text{HFTps}_{14}$, the trajectory will be recognized by attacker with background knowledge. As shown in Figure 11(b), after adding the start- and end-point determinations, the scheme of this paper can still guarantee privacy protection at the k-anonymity level. Because the Rotation algorithm generates a new dummy trajectory through rotation, the attacker's ability to identify the start- and end-points is the least developed. The Round algorithm generates a dummy trajectory based on the original trajectory but does not consider the generation of the start- and end-points. Resistance to the identification of the start- and end-points is also weak.

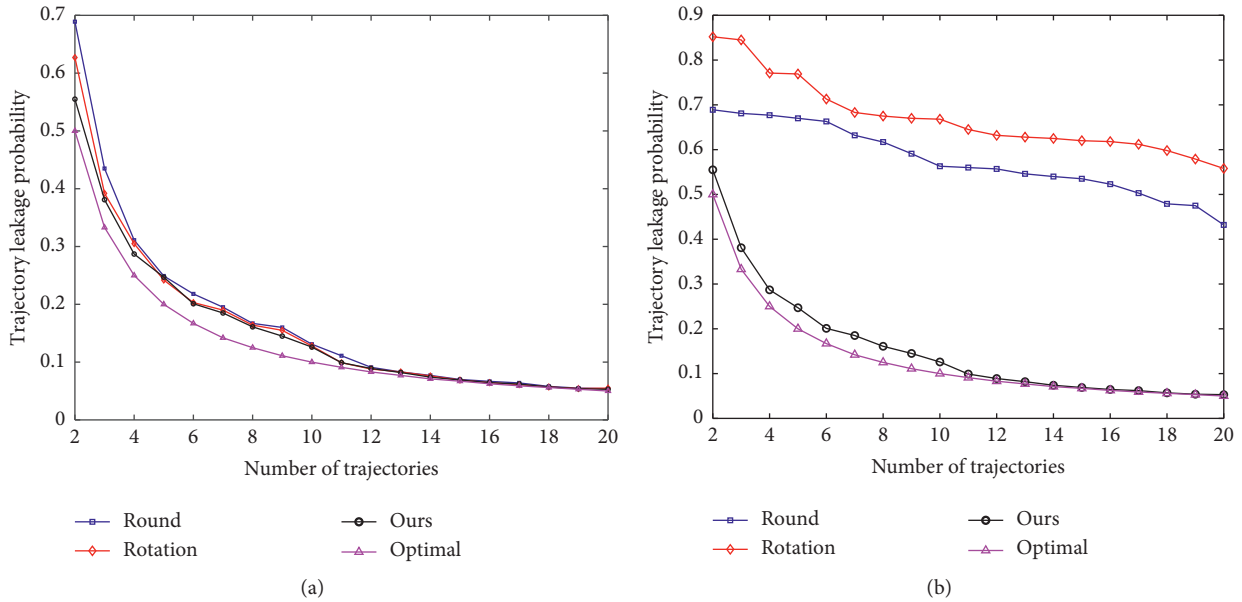

FIGURE 11: Trajectory leakage probability comparison. (a) Regardless of the safety of the start- and end-points. (b) Considering the safety of the start- and end-points.

6. Conclusion

Addressing the privacy protection problem of trajectory data release, this paper proposes a new and efficient method of protection based on security start- and end-points. The method does not rely on a trusted third party to generalize the start- and end-points based on the user's background information. We propose instead a two-way dummy trajectory generation algorithm to generate $k-1$ paths that are difficult to detect by attackers with background knowledge. Experiments show that our method reduces data loss while protecting user privacy. The dummy trajectory generation method realizes the trajectory k-anonymity and meets the needs of user privacy protection.

The generation of the secure start- and end-point candidate set depends on a large quantity of personal data belonging to the user. When sufficient data is not available, it is often difficult to achieve the desired effect. Although there are many trajectory generation algorithms [30], the dummy trajectories generated by these algorithms often do not conform to background knowledge and human habits. Therefore, future work is to start with a small number of user trajectories and design an algorithm that can generate a large amount of reliable trajectory data.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61972439, 61672039, and 61702010).

References

[1] G. Xie, H. Gao, L. Qian, B. Huang, K. Li, and J. Wang, "Vehicle trajectory prediction by integrating physics- and maneuver-based approaches using interactive multiple models," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5999–6008, 2018.

[2] W. Yang and T. H. Ai, "A method for road network updating based on vehicle trajectory big data," *Journal of Computer Research and Development*, vol. 53, no. 12, pp. 2681–2693, 2016, in Chinese.

[3] M. Li, L. Zhu, Z. Zhang, and R. Xu, "Achieving differential privacy of trajectory data publishing in participatory sensing," *Information Sciences*, vol. 400-401, pp. 1–13, 2017.

[4] L. Sweeney, "Uniqueness of simple demographics in the US. population," in *Data Privacy Lab White Paper Series LIDAP-WP4*, Carnegie Mellon University, Pittsburgh, PA, USA, 2000.

[5] T. Wang, J. Zeng, M. Z. A. Bhuiyan et al., "Trajectory privacy preservation based on a fog structure for cloud location services," *IEEE Access*, vol. 5, pp. 7692–7701, 2017.

[6] Z. Huo and X.-F. Meng, "A survey of trajectory privacy-preserving techniques," *Chinese Journal of Computers*, vol. 34, no. 10, pp. 1820–1830, 2011, in Chinese.

[7] L. Sweeney, "K-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[8] A. Machanavajjhala, J. Gehrke, D. Kifer et al., "L-diversity: privacy beyond k-anonymity," in *Proceedings of the International Conference on Data Engineering*, IEEE, San Diego, CA, USA, pp. 1–10, March 2000.

[9] S. P. A. Alewijnse, K. Buchin, M. Buchin et al., "Model-based segmentation and classification of trajectories," *Algorithmica*, vol. 80, no. 2, pp. 1–31, 2017.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, DBLP, San Francisco, CA, USA, pp. 31–42, May 2003.

[11] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: uncertainty for anonymity in moving objects databases," in *Proceedings of the International Conference on Data Engineering*, IEEE Computer Society, Cancun, Mexico, pp. 376–385, April 2008.

[12] L. Zhang, S. Yang, J. Li et al., "A particle swarm optimization clustering-based attribute generalization privacy protection scheme," *Journal of Circuits, Systems, and Computers*, vol. 27, no. 11, pp. 641–654, 2018.

[13] K. Sui, Y. Zhao, D. Liu et al., "Your trajectory privacy can be breached even if you walk in groups," in *Proceedings of the International Symposium on Quality of Service*, IEEE, Beijing, China, pp. 1–5, June 2016.

[14] T. H. You, W. C. Peng, and W. C. Lee, "Protecting Moving Trajectories with dummies," in *Proceedings of the International Conference on Mobile Data Management*, IEEE, Hong Kong, China, pp. 278–282, June2008.

[15] F. H. Li, C. Zhang, B. Niu et al., "Efficient scheme for user's trajectory privacy," *Journal on Communications*, vol. 36, no. 12, pp. 114–123, 2015, in Chinese.

[16] K. Y. Lei, X. H. Li, H. Liu et al., "Dummy trajectory privacy protection scheme for trajectory publishing based on the spatiotemporal correlation," *Journal on Communications*, vol. 37, no. 12, pp. 156–164, 2016, in Chinese.

[17] J. Xiao, L. Xu, L. Lin et al., "A privacy-preserving approach based on graph partition for uncertain trajectory publishing," *IEEE*, in *Proceedings of the International Symposium on Parallel and Distributed Computing*, pp. 285–290, Kyoto Japan, May 2016.

[18] E. G. Komishani, M. Abadi, and F. Deldar, "PPTD: preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression," *Knowledge-Based Systems*, vol. 94, pp. 43–59, 2016.

[19] Q. A. Arain, Z. Deng, I. Memon et al., "Map services based on multiple mix-zones with location privacy protection over road network," *Wireless Personal Communications*, vol. 97, no. 3, pp. 2617–2632, 2017.

[20] L. Zhang, L. He, L. Desheng et al., "An attribute generalization mix-zone without privacy leakage," *IEEE Access*, vol. 7, pp. 57088–57099, 2019.

[21] Z. Huo, H. L. Cui, and P. He, "Trajectory data privacy-preserving based on semantic location protection," *Journal of Computer Appication*, vol. 38, no. 1, pp. 182–187, 2018, in Chinese.

[22] Y. Wang, Z. Cai, X. Tong, Y. Gao, and G. Yin, "Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems," *Computer Networks*, vol. 135, no. 22, pp. 32–43, 2018.

[23] X. Liu, A. Liu, and X. Zhang, "When differential privacy meets randomized perturbation: a hybrid approach for privacy-preserving recommender system," in *Proceedings of the International Conference on Database Systems for Advanced Applications*, pp. 576–591, Suzhou, China, May 2017.

[24] T. Rong, T. Yuan, S. Wen, and Y. Y. Zhang, "Privacy preserving semantic trajectory data publishing for mobile location-based services," *Wireless Networks*, vol. 1, pp. 1–10, 2019.

[25] B. Wang, L. Zhang, and G. Y. Zhang, "A novel ε-sensitive correlation indistinguishable scheme for publishing location data," *PLoS One*, vol. 14, no. 12, Article ID e0226796, 2019.

[26] B. Niu, Q. Li, X. Zhu et al., "Achieving k-anonymity in privacy-aware location-based services," in *Proceedings of the IEEE INFOCOM*, pp. 1–10, Toronto, ON, Canada, May 2014.

[27] R. Yarovoy, F. Bonchi, L. Lakshmanan et al., "Anonymizing moving objects: how to hide a MOB in a crowd?" in *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, ACM, New York, NY, USA, pp. 72–83, March 2009.

[28] Y. Zheng, X. Xie, and W. Ma, "GeoLife: a collaborative social networking service among user, location and trajectory," *IEEE Data Engineering Bulletin*, vol. 33, no. 2, pp. 32–39, 2011.

[29] R. Trujillo-Rasua and J. Domingo-Ferrer, "On the privacy offered by (k, δ)-anonymity," *Information Systems*, vol. 38, no. 4, pp. 491–494, 2013.

[30] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630–643, 2011.