




Research Article

Trajectory Privacy Protection Based on Sensitive Stay Area Replacement in Publishing

Ji Yali ^{1,2} Gui Xiaolin ^{1,2} Dai Huijun ^{1,2} An Jian,^{1,2} Zhu Hongyi,^{1,2} Peng Zhenlong,^{1,2} and Lin Xinyang^{1,2}

¹School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

²Key Laboratory of Computer Network of Shaanxi Province, Xi'an, China

Correspondence should be addressed to Gui Xiaolin; xlgui@mail.xjtu.edu.cn and Dai Huijun; dhj74@126.com

Received 14 October 2021; Revised 20 August 2022; Accepted 6 September 2022; Published 21 October 2022

Academic Editor: Eric Florentin

Copyright © 2022 Ji Yali 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.

In data opening and sharing, a trajectory privacy protection based on sensitive stay area replacement (SSAR) is proposed in order to improve the availability of protected trajectory while maintaining the same security. Firstly, the stay areas are extracted by analyzing the movement characteristics. Secondly, the sensitive stay areas are obtained according to the user-defined sensitive semantic positions. Then, the sensitive areas are constructed according to the user's privacy requirement needs to randomly select the substitution. Finally, part of the sampling positions in the sensitive area are reset. The parameter selection experiment and the comparison experiment with other schemes show that the trajectory similarity is improved by more than 35% in SSAR. That is to say, SSAR can greatly improve the availability of protected trajectory on the basis of ensuring the safety degree.

1. Introduction

With the popularity of the mobile devices with positioning capabilities such as smart phones, vehicle-borne GPS, smart wearing devices, location-based services (LBS) are widely pervasive in social, commercial, and other fields. Not only has great convenience of life been brought, but also unprecedented high value location data (such as interest points and trajectories) have been generated by LBS applications. The opening and sharing of large location data promote the construction process of a smart city. By analyzing the trajectory of mobile users, the situation of people flow between regions can be acquired, which can be used in urban planning and traffic planning, the hot spots and active time periods of visitors can be deduced, which can be used for locating various large shopping malls, and the traffic flow can be calculated which can be used to plan the public transport facilities reasonably. However, the original trajectory data is closely related to some sensitive information such as users' home address, hobbies, behavior patterns, and so on. If the data are not protected when publishing and sharing, it is easy to obtain users' privacy through mining

and analysis. Therefore, it is very important to the privacy protection of trajectory data before opening and sharing [1–5]. Locations contained in the trajectory of mobile users do not exist independently, which are related in time and space. Thus, the trajectory privacy protection cannot be processed independently on each position, and the integrity of the trajectory should be taken into account. In addition, both scientific research and commercial applications need to mine and analyze available information from the protected trajectory, which means that in publishing, trajectory privacy protection for mobile users should not only prevent malicious attackers from speculating user-sensitive information but also ensure the integrity and availability of the protected trajectory.

Up to now, the related research on trajectory data opening and sharing has been extensively carried out at home and abroad. Tian et al. [6] proposed a personalized trajectory generalization algorithm for trajectories with different privacy preferences. Han et al. [7] classified the trajectories containing sensitive location points according to the correlation of location points on the trajectory and then used a random response mechanism to select a reasonable

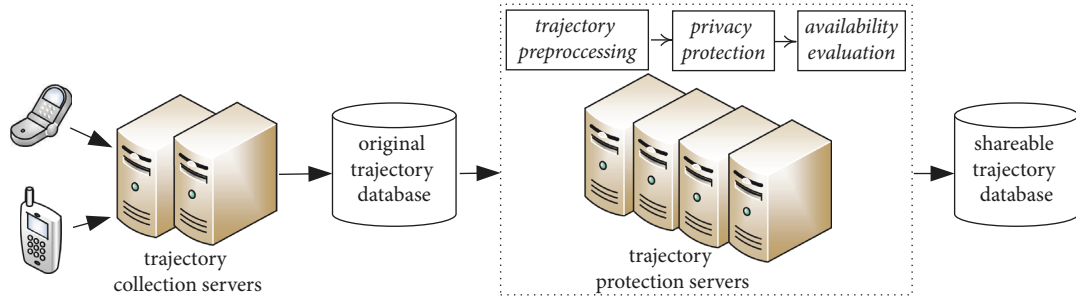


FIGURE 1: Trajectory privacy protection system structure.

candidate set to replace the sensitive points in the trajectory. Wang et al. [8] proposed the interpolation trajectory-anonymous privacy protection algorithm with temporal and spatial granularity constraints, which is on the basis of a hierarchical model and interpolation-based modified Hausdorff distance on an adjacent segment. Chen et al. [9] proposed a new differential privacy scheme based on the Recurrent Neural Network for Dynamic trajectory privacy Protection (RNN-DP). It introduced a recurrent neural network model to handle the real-time data effectively instead of the full data, and design a prejudgment mechanism to increase the availability of differential privacy technology. Domingo-Ferrer et al. [10] proposed metrics to quantify the data confidentiality and utility achieved by SDC methods based on the permutation model and distinguish two privacy notions. In these schemes, they do not take the semantic location or trajectory into account. So, malicious attackers can easily obtain users' privacy by using map background knowledge. To combat semantic attacks, some scholars have also proposed some trajectory privacy protection schemes that take the semantic attributes into account. Tu et al. [11] proposed to prevent user trajectories from being reidentified and semantically attacked by trajectory merging. Although this scheme can avoid semantically attacking, the time cost of the trajectory merging is large. Hu and Yang [12] proposed a personalized trajectory privacy protection method based on location semantic perception to achieve the personalized goal of privacy protection parameter setting and policy selection. Tan et al. [13] proposed semantic trajectory anonymizing based on the k -anonymity model, in which, there are sensitive areas that contain $k-1$ POI (Points of Interest) points that are similar to the sensitive points, and trajectory ambiguity is executed based on the motion modes, road network topologies and road weights in the sensitive area. Ye et al. [14] proposed a novel approach to conceal the actually visited sensitive place. The trajectories are not simply considered as a sequence of the coordinates in Euclidean space, but they combine the semantics-aware information with the background knowledge of the underlying map for the location points. With the emergence of new attack methods such as in [15] which is based on semantic trajectory patterns, many trajectory privacy protection methods will be proposed.

In this paper, a trajectory privacy protection method based on sensitive stay area replacement (SSAR) is proposed in view of the privacy requirements of trajectories for mobile

users in publishing, which fully considers the semantic properties of trajectories and achieves privacy protection while ensuring the integrity and availability of trajectory by replacing sensitive stay areas and resetting sampling.

2. Problem Statement

The structure of the trajectory privacy protection system in publishing is shown in Figure 1. The trajectory collection server collects trajectory data and stores them in the trajectory database. The trajectory protection server protects the collected trajectories and generates a shareable trajectory database, which includes three modules of trajectory preprocessing, privacy protection, and availability evaluation. The SSAR involved the three modules.

The existing trajectory privacy protection in publishing only regards trajectory as a sequence of locations with temporal attributes in European space, which only considers the temporal and spatial attributes of the trajectory, but ignores the location semantic corresponding to each sampling, which is the semantic of trajectory.

Define 1. Stay Area $SA(tb, te, Tra, sem)$. The physical meaning of a stay area is the area where the user frequently visit and stay for a long time. In user trajectory $Tra(M, (L_1(x_1, y_1), t_1), (L_2(x_2, y_2), t_2), \dots, (L_n(x_n, y_n), t_n)))$, M denotes the user, $(L_j(x_j, y_j), t_j)$ denotes the sampling time t_j and the corresponding sampling location $L_j(x_j, y_j)$, t_b represents the start time of the stay, t_e represents the end time of stay, which are, respectively, corresponding to a sampling time while $b < e$, and sem denotes the semantic position.

Users will have a large number of stay areas in their daily life, such as staying at home from 00:00 a.m. to 06:00 a.m. and staying at the company from 09:00 a.m. to 12:00 a.m. Although the samplings on trajectory have corresponding semantic attributes, malicious attackers with map background knowledge pay more attention to stay areas, because it is easy to infer users' privacy such as their work address, hobbies, religious beliefs, and physical condition by mining their stay areas. As shown in Figure 2, it is clear when and how long the user stays after analyzing the movement characteristics of the original trajectory, obtaining the stay areas, and mapping into the map. The trajectory privacy protection method which ignores semantic attributes cannot guarantee the security when encountering the attackers with map background knowledge who can analyze the sensitive

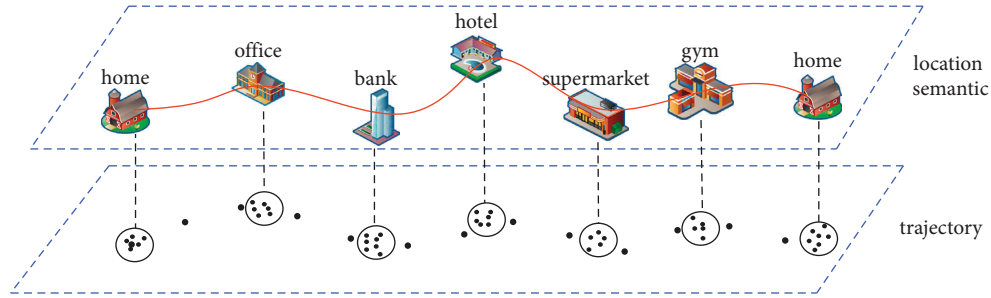


FIGURE 2: Trajectory and location semantic.

information in the user's trajectory. There are already some trajectory privacy preservation methods which have taken semantic attributes into account, while the way to deal with the sensitive location on the trajectory is still unreasonable, some shortcomings remaining in security and performance.

Different users may have different sensitivities to the same semantic location in real life. For example, patients and doctors have different sensitivities to hospitals. Patients may not want to expose their physical health while doctors generally do not mind exposing their workplace. Therefore, the user's personalized privacy requirement should not be neglected in trajectory privacy protection when the trajectory is protected. If the same standard is adopted for all users, it may lead to privacy disclosure of inadequate trajectory protection for some users and data loss of excessive trajectory protection for some other users. Protecting the sensitive stay areas does not deal with all samplings on the trajectory, which can not only hide the sensitive information of users, but also reduce the damage to the original trajectory, and achieve a better balance between privacy protection and data availability.

The process of trajectory privacy protection in this paper is shown in Figure 3. The little dots are the location points in the plane of the original trajectory and sharing trajectory. In the plane of the stay area, a big dot is the location of a stay area, a little dot is a location point. Firstly, the user's movement characteristics are analyzed according to the original trajectory, and the set of user's stay areas within a day is extracted by the multidimensional clustering of three attributes including time, longitude, and latitude, and the corresponding actual locations of the stay areas are acquired and their semantics are marked by map inverse analysis. Then, the set of sensitive stay areas (i.e., banks and hotels in Figure 3) is obtained according to the user-defined sensitive semantic locations, and different POI satisfying privacy requirements are found combining with the user's movement direction for the rational plans of privacy region for each sensitive stay area. The semantics and distance characteristics of POI in the privacy region are analyzed to select the smallest rectangle containing these POIs as the sensitive area, and a substitution stay area (i.e., restaurant and barbershop in Figure 3) is chosen randomly in the sensitive area. Finally, in order to prevent the sudden change of locations caused by the replacement of stay areas and reduce the trajectory change, only part of the samplings in the sensitive area are reset, and the number of samplings in the sensitive area is ensured to be

consistent with the original trajectory, forming the final trajectory data that can be shared. The proposed scheme aims at maintaining a good balance between trajectory security and availability through replacing the sensitive stay areas with the stay areas of different semantics around them while hiding the user-sensitive information by resetting a small number of samplings. The scheme adopts personalized privacy requirement, in which users can customize their own sensitive semantic location set and privacy protection degree, and ensures the integrity and availability of the trajectory while ensuring the privacy and security of that.

3. Sensitive stay Area Replacement

3.1. Stay Area Extraction. Stay area extraction includes sampling locations clustering and semantic parsing. Sampling locations clustering processes the positions of trajectory from three dimensions which are time, distance, and quantity. Please refer to the personal interest area extraction in [16] for specific. In the semantic parsing, the minimum covering circle $CR(OA(x_O, y_O), R)$ of all sampling positions $Set_{L_{b,e}} = \{L_b, L_b + 1, \dots, L_e - 1, L_e\}$ in the stay area $SA(tb, te, Tra, sem)$ is calculated firstly. Then, the Baidu Map Web Service API is called and the reverse address coding service is used to obtain semantic attributes. The location of the stay area is represented by the centre $OSA(x_O, y_O)$ of the minimum covering circle as shown in Figure 4. The pseudocode of stay area extraction is shown in Algorithm 1.

where $MinCovering()$ is to get the minimum covering circle.

3.2. Privacy Region Construction

Define 2. Sensitive stay area. If the semantic location of stay area $SA(tb, te, Tra, sem)$ is an element in the user-defined sensitive semantic location set $SEMP$, the stay area is considered to be a sensitive stay area. For example, $SEMP = \{\text{hotel, bank, hospital}\}$. $SEMP$ divides all stay areas of trajectory into two parts: the sensitive stay area set and the nonsensitive stay area set. Protecting the sensitive stay area set can not only avoid malicious attackers inferring user's sensitive information but also reduce the processing cost and improve the integrity and the availability of trajectory.

A privacy region is a ring that expands the minimum covering circle of sensitive stay areas and meets the privacy

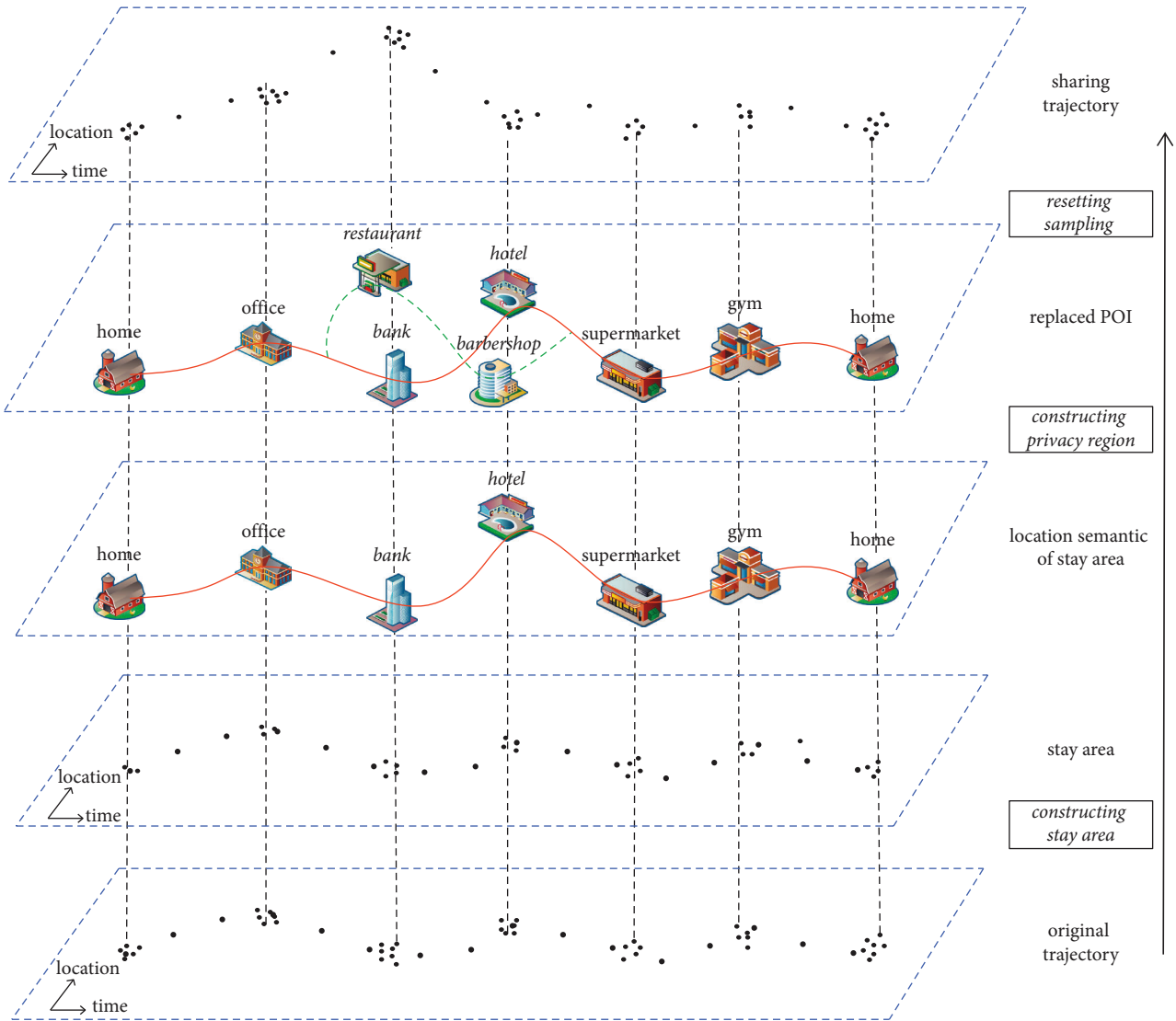


FIGURE 3: The process of SSAR.

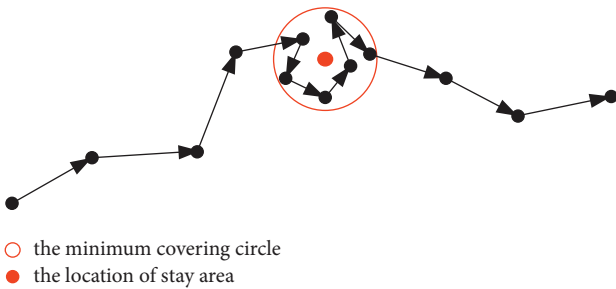


FIGURE 4: The example of minimum covering circle.

requirement l . Privacy requirement l indicate that there is at least l POI in the privacy region which is different from the semantics of the sensitive stay area, reflecting the location and semantic diversity of POIs in the privacy region as shown in Figure 5 when $l = 7$. The pseudocode of privacy region construction is shown in Algorithm 2.

3.3. Sampling Positions Resetting. In order to maximize the consistency of the trajectory shapes, the original trajectory is modified as little as possible. The sensitive region is constructed as the reset range of the sampling positions. When resetting, the movement speed on the original trajectory should be fully considered, and the number of sampling positions in the sensitive area should keep unchanged.

Define 3. Sensitive region SR. The sensitive region is the smallest rectangular region that contains all the POIs in the privacy region and the location of the sensitive stay area SA.

When a POI is randomly selected in the sensitive area, its location $L_{rp}(x_{rp}, y_{rp})$ is used to replace the location of stay area $OSA(x_o, y_o)$. After the replacement, the sampling position will not be able to reach the substitution position within the sampling interval and the position will suddenly change, which is easy for the attacker to infer that the trajectory has been modified, so the sampling position on the

```

Input:
Tra( $M, (L_1(x_1, y_1), t_1), (L_2(x_2, y_2), t_2), \dots, (L_n(x_n, y_n), t_n)$ )
Output: SetSA( $tb, te, Tra, sem$ )
(1) clustering
Tra( $M, (L_1(x_1, y_1), t_1), (L_2(x_2, y_2), t_2), \dots, (L_n(x_n, y_n), t_n)$ )
(2) for each Set $L_{b,e} = \{L_b, L_b + 1, \dots, L_e - 1, L_e\}$ 
(3)   CR $_{SA}(O(x_O, y_O), R) =$ 
      MinCovering (Set $L_{b,e} = \{L_b, L_b + 1, \dots, L_e - 1, L_e\}$ )
(4)   call API
(5)   returnsem
(6) end for
(7) returnSetSA
    
```

ALGORITHM 1: Stay Area Extraction.

```

Input: CR( $O_{SA}(x_O, y_O), R$ ), privacy requirements  $l$ 
Output: privacy region PR =  $\{O_{SA}(x_O, y_O), R\}$ 
(1) PR =  $\{O_{SA}(x_O, y_O), R\}$ 
(2) scount = |PR|
(3) count = 0
(4)  $r = R$ 
(5) do
(6)    $r++$ 
(7)   PR =  $\{O_{SA}(x_O, y_O), r\}$ 
(8)   count = |PR| - scount
(9) whilecount <  $l$ 
(10) returnPR =  $\{O_{SA}(x_O, y_O), r, l\}$ 
    
```

ALGORITHM 2: Privacy Region Construction PRB(CR($O_{SA}(x_O, y_O), R$), l).

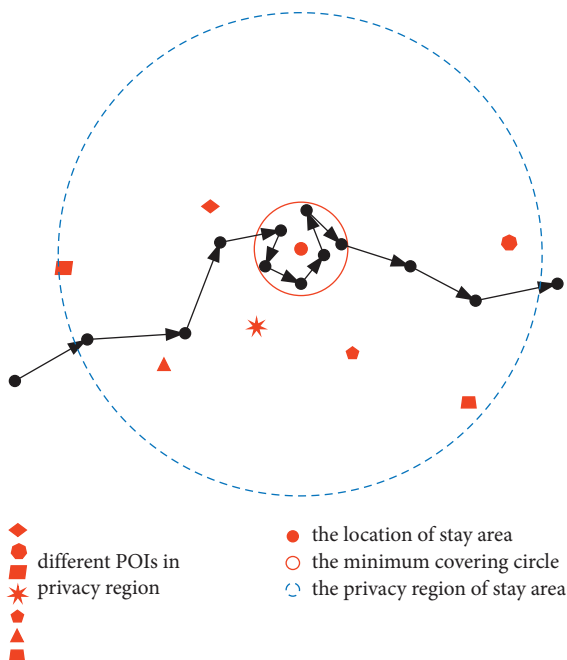


FIGURE 5: The example of privacy region.

trajectory needs to be reset. |SA| positions are randomly generated within the radius R and the centre of circle $L_{rp}(x_{rp}, y_{rp})$ to form a substitution stay area SA' (tb, te, Tra', sem') of the sensitive stay area SA according to the velocity range $[v_{min}, v_{max}]$ in the sensitive region. Assuming that the first sampling position in the sensitive region is A and the last is B, find C on the original trajectory from A to SA and make the difference between the distance from C to OSA(x_O, y_O) and the distance from C to $L_{rp}(x_{rp}, y_{rp})$ the smallest. Similarly, find D on the original trajectory from SA to B. The new sampling positions are determined according to the velocity range and sampling time while ensuring that the number of sampling positions is equal to the original trajectory. By finding two sampling positions C and D, it is not necessary to reset the sampling position of the entire sensitive area in the process of resetting the sampling position, which not only reduces the time cost of resetting but also improves the integrity of the trajectory. After resetting the samplings, the privacy of the trajectory is protected. The sharable trajectory can be published for mining and analysis. An example of sampling positions resetting is shown in Figure 6. The pseudocodes of sampling position resetting are shown as Algorithm 3.

```

Input:  $L_{rp}(x_{rp}, y_{rp}), R, [v_{min}, v_{max}],$ 
 $\{L_A, \dots, L_b - 1, L_b, L_b + 1, \dots, L_e - 1, L_e, L_e + 1, \dots, L_B\}$ 
Output:  $\{L_A, \dots, L_C, L_C' + 1, \dots, L_b', \dots, L_e', \dots, L_D' - 1, L_D, \dots, L_B\}$ 
(1) find  $C$  from  $A$  to  $SA$  with  $\min(|\text{Dis}(L_C, O_{SA}(x_o, y_o)) - \text{Dis}(L_C, L_{rp}(x_{rp}, y_{rp})|)$ 
(2) find  $D$  from  $SA$  to  $B$  with  $\min(|\text{Dis}(L_D, O_{SA}(x_o, y_o)) - \text{Dis}(L_D, L_{rp}(x_{rp}, y_{rp})|)$ 
(3)  $\Delta v = v_{max} - v_{min}$ 
(4) num1 = the number of moving sample points in  $C$  to  $SA$ 
(5) num2 = the number of moving sample points in  $SA$  to  $D$ 
(6) for  $j = 1$ : num1
(7)    $dis_j = (v_{min} + \text{random}[0, \Delta v]) * t * j$ 
(8)   find new position with  $\text{Dis}(C, L_{C+j}') = dis_j$ 
(9) end for
(10) for  $k = 1$ : num2
(11)   $dis_k = (v_{min} + \text{random}[0, \Delta v]) * t * (\text{num2} + 1 - k)$ 
(12)  find new position with  $\text{Dis}(D, L_{e+k}') = dis_k$ 
(13) end for
(14) randomly get  $e - b + 1$  positions in circle  $(L_{rp}(x_{rp}, y_{rp}), R)$  to  $\{L_b', L_b' + 1, \dots, L_e' - 1, L_e'\}$ 

```

ALGORITHM 3: Sampling position setting.

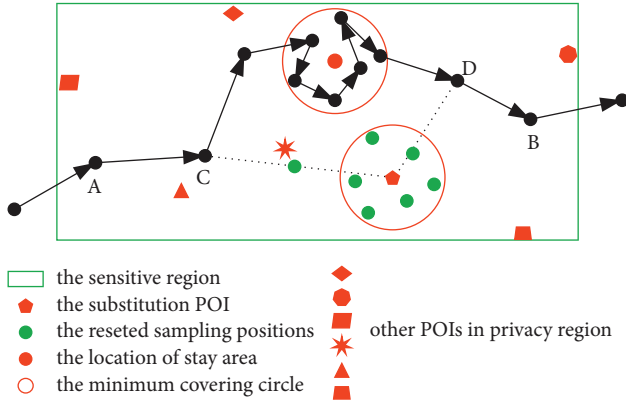


FIGURE 6: The example of sampling positions resetting.

4. Experiments and Analysis

4.1. Evaluation Indexes. Assuming that the attacker has the following background knowledge. Combined with some auxiliary information, the attacker can determine the user true identity by some means such as reidentification. Besides, the attacker has the map background knowledge to know the actual geographical location corresponding to any sampling location and can determine whether the two sampling positions are reachable by calling the map interface. If the probability that a sensitive stay area is recognized is P_{SSA} , then

$$P_{SSA} = \frac{1}{l}. \quad (1)$$

Define 4. Average recognition rate P_{ave} . For the database with $|\text{Traj}|$ trajectories, if each trajectory contains $|\text{SSAP}|$ sensitive stay areas, the average probability that all users' sensitive stay areas are recognized in the whole trajectory database is as follows:

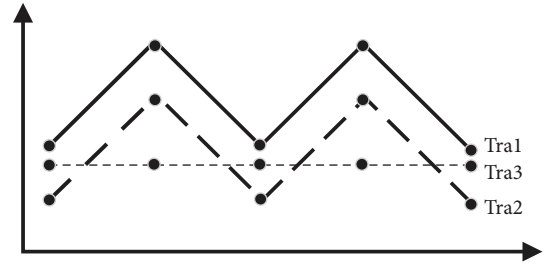
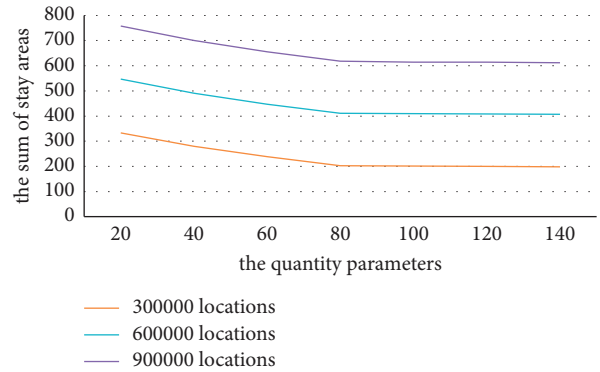


FIGURE 7: The example of the trajectory shape.

FIGURE 8: The relationship of the quantity parameters σ_n and the sum of stay areas.

$$P_{ave} = \frac{\sum_1^{|\text{Traj}|} \left(\sum_1^{|\text{SSAP}|} 1/l / |\text{SSAP}| \right)}{|\text{Traj}|}. \quad (2)$$

The average recognition rate P_{ave} reflects the average probability that sensitive information on all trajectories in the trajectory database are leaked. The smaller the average recognition rate is, the higher the privacy protection intensity is.

The similarity between protected trajectory and original trajectory is an important index to measure the availability of

trajectory. The measurement of similarity can not only be conducted in time and space, but also in the trajectory shape. Just as Figure 7 shows, there are five samplings in each of the trajectories Tra1, Tra2, and Tra3, if the similarity between trajectories is measured only by the distance between samplings, the Tra2, and the Tra3 are similar to the Tra1, while it is obvious that Tra2 is more similar to Tra1.

$Tra_i(M_i, (L_1(x_1^i, y_1^i), t_1^i), \dots, (L_n(x_n^i, y_n^i), t_n^i)))$ and $Tra_j(M_j, (L_1(x_1^j, y_1^j), t_1^j), \dots, (L_n(x_n^j, y_n^j), t_n^j)))$ are two trajectories.

The trajectory location distance, trajectory shape distance and trajectory distance between them are defined as following.

$$D_{sha}(Tra_i, Tra_j) =$$

$$\sum_1^n \sqrt{(x_{s+1}^i - x_s^i/t_{s+1}^i - t_s^i - x_{s+1}^j - x_s^j/t_{s+1}^j - t_s^j)^2 + (y_{s+1}^i - y_s^i/t_{s+1}^i - t_s^i - y_{s+1}^j - y_s^j/t_{s+1}^j - t_s^j)^2}. \quad (4)$$

Define 7. Trajectory distance. Trajectory distance can be computed with Equation.

$$D(Tra_i, Tra_j) = \alpha D_{loc}(Tra_i, Tra_j) + (1 - \alpha) D_{sha}(Tra_i, Tra_j), \quad (5)$$

where $\alpha \in [0, 1]$ is the weight. The smaller the trajectory distance is, the smaller the deviation between two trajectories is, and the higher the similarity is.

4.2. Experiment Results. In experiments, CPU is Intel (R) Core (TM) i7-6700 CPU @ 3.40 GHz 3.41 GHz, memory is 20 GB, and the operating system is Windows 10 Professional Edition. The data set comes from the GeoLife project of Microsoft Asia Research Institute.

4.2.1. Parameters Selection. The stay area extraction involves three parameters that are time parameter σ_t , distance parameter σ_d , and quantity parameter σ_n . The time and distance reflect the speed of users, it is about 3 km/h in real life, so the time parameter is $\sigma_t = 5$ min and distance parameter is $\sigma_d = 500$ m in this paper. The quantity parameter essentially reflects the length of stay time, it is very important. 300,000, 600,000, and 900,000 sampling locations are used to determine quantity parameter. The sampling time intervals is 15 s in Figure 8. The three data sets show similar trends, and the total number of stay areas decreases rapidly when the quantity parameter changes from 20 to 80. After 80, the decline rate tends to be stable. It is because the initial stay time is relatively short, many areas such as making a telephone call or waiting at a bus stop for a few minutes are regarded as the stay areas. Then, with the increase of stay time, many insufficient stay areas are screened out, only the longer and more meaningful ones remained. So, the quantity parameters is $\sigma_n = 80$ in next experiments.

Define 5. Trajectory location distance. Trajectory location distance can be computed with Equation(3).

$$D_{loc}(Tra_i, Tra_j) = \frac{\sum_1^n \sqrt{(x_s^i - x_s^j)^2 + (y_s^i - y_s^j)^2}}{n}. \quad (3)$$

Define 6. Trajectory shape distance. Trajectory location distance can be computed with Equation.

4.2.2. Practical Test. An example of practices of the second stay area is bank, and set it as a sensitive stay area. Then, a sensitive region which is the rectangle as shown in Figure 9(b) is constructed according to the privacy requirement $l = 7$, and it contains seven different POIs besides the sensitive stay area. Then, a POI is randomly selected as a substitution position, and the sampling position is reset to obtain a new trajectory after privacy protection, as shown in Figure 9(b), A sharable trajectory is obtained.

4.2.3. Privacy Requirement. The privacy requirement l not only directly affects the security of the trajectory database but also affects the size of the sensitive region and the number of resetting sampling positions, that is the availability of a protected trajectory. The GeoLife project does not collect users' personalized privacy requirements, so the value about l should be determined through experiments. 344 stay areas were extracted by analyzing 500,000 sampling positions of 93 trajectories in the experiment. Moreover, the GeoLife project did not collect the sensitive semantic location set of each user, so seven sensitive semantic locations such as hospital, hotel, bank, and bar were set as sensitive semantic location set in the experiment, and 78 sensitive stay areas were extracted based on sensitive semantic location set. The experiment results are shown in Figure 10. Figure 10(a) shows the average recognition rate and the privacy requirement l , it is the security of the trajectory, Figure 10(b) shows the trajectory distance and the privacy requirement l , it is availability of trajectory, and Figure 10(c) shows the running time and the privacy requirement l when $\sigma_t = 5$ min, $\sigma_d = 500$ m and $\sigma_n = 80$, it is the performance of SSAR. It can be seen from Figure 10 that the security is improved, and the availability and performance are declining with the increase of the privacy degree. The selection of l should ensure high security, $l \geq 12$ by analyzing Figure 10(a). To ensure the availability, $l \leq 18$ by analyzing Figures 10(b) and 10(c).



FIGURE 9: The example of practical test by SSAR. (a) The original trajectory and three stay areas(the circle). (b) The sharable trajectory, the sensitive region(the rectangular), and the replacement stay area(the circle).

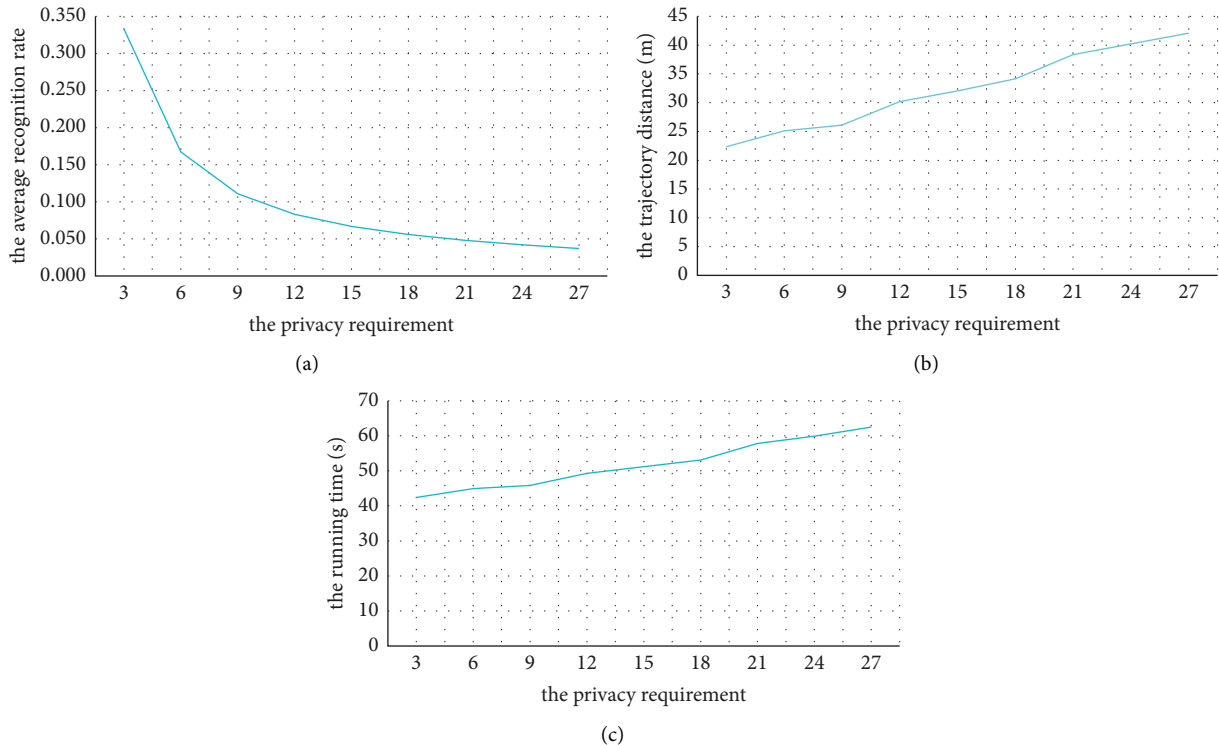


FIGURE 10: The relationship of the personalized privacy requirements l and the evaluation indexes (a) The privacy requirement l and the average recognition rate(security), (b) The privacy requirement l and the trajectory distance(availability), (c) The privacy requirement l and the running time(performance).

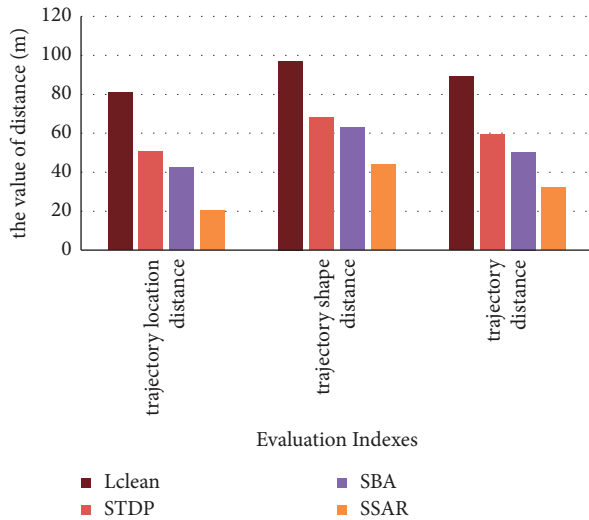


FIGURE 11: Usability comparison in Lclean, STDP, SBA, and SSAR.

4.2.4. Usability Comparison. In order to fully illustrate the higher availability in SSAR, two existing main schemes are selected for comparing in this paper. Lclean in reference [7] selects substitution points in the whole trajectory or trajectory segment to replace the user's sensitive stay areas. STDP in reference [13] selected a POI in sensitive areas containing $k - 1$ POI points which are similar to the sensitive points. SBA in reference [14] selected the same type of POI based on the classification tree of POI in the cloaking region. The availability of each scheme is measured by trajectory location distance, trajectory shape distance, and trajectory distance. In this experiment, 500,000 sampling positions are used, $\sigma_t = 5$ min, $\sigma_d = 500$ m and $\sigma_n = 80$. The average distance when $12 \leq l \leq 18$ is shown in Figure 11.

The experiment results show that the trajectory distance in SSAR is reduced by 56.802 m, 27.108 m, and 18.272 m less than in Lclean, STDP, and SBA. Respectively, the trajectory similarity is increased about 36.2% at least.

5. Conclusion

In trajectory publishing, the existing schemes cannot balance the security and availability, and ignore the semantic attributes, which brings great security risks. To solve these problems, trajectory privacy protection based on sensitive stay area replacement (SSAR) is proposed in this paper. Different semantic POIs are used to replace users' sensitive stay areas and the sampling positions are reset to hide users' sensitive information. SSAR reduces the damage to the original trajectory and achieves a good balance between security and availability. Firstly, the stay areas are extracted by analyzing the user's movement characteristics, and the sensitive stay areas are obtained by combining the user's sensitive semantic location set. Secondly, the sensitive areas are constructed according to the user's movement direction, the semantic characteristics, and the distance characteristics of the surrounding POIs, and the sensitive areas meeting privacy requirement to randomly select substitution position. Finally, part of the samplings in the sensitive area are

reset according to the user's mobile speed. In order to verify the effectiveness of SSAR, some experiments are carried out on the real trajectory set, and two main schemes are selected to compare with. The experiment results show that SSAR has better integrity and availability while ensuring the trajectory safety.

In future work, we plan to extend our method to the real-time data publishing and study the evaluation index system, improving the security and availability to trajectory privacy protection.

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program under Grant no.2018YFB1800304 and the National Key R&D Program of Shaanxi Province under Grant no.2019-GY-005.

References

- [1] X. Liu, Q. Xie, and L. Wang, "Personalized extended (α, k) -anonymity model for privacy - preserving data publishing," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 6, 2017.
- [2] K. Gu, L. Yang, Y. Li, and Y. Bo, "Efficient trajectory data privacy protection scheme based on laplace's differential privacy," *Informatica*, vol. 42, no. 3, pp. 417–438, 2018.
- [3] L. Xiangyu, C. Jinmei, X. Xiufeng, C. Zong, R. Zhu, and J. Li, "Dummy-Based Trajectory Privacy Protection Against Exposure Location Attacks," in *Proceedings of the 16th International Conference On Web Information Systems and Applications*, pp. 368–381, China, September 2019.
- [4] S. Li, H. Shen, Y. Sang, and H. Tian, "An efficient method for privacy-preserving trajectory data publishing based on data partitioning," *The Journal of Supercomputing*, vol. 76, no. 7, pp. 5276–5300, 2020.
- [5] H. W. Jiang and K. Hu, "A clustering-anonymity approach for trajectory data publishing considering both distance and direction," *Journal of Computing and Information Technology*, vol. 29, no. 1, pp. 1–12, 2022.
- [6] F. Tian, S. Zhang, L. Lu, H. Liu, and X. Gui, "A Novel Personalized Differential Privacy Mechanism for Trajectory Data Publication," in *Proceedings of 2017 International conference on Networking and Network Applications*, pp. 61–68, Kathmandu, Nepal, October 2017.
- [7] Q. Han, D. Lu, K. Zhang, X. Du, and M. Guizani, "Lclean: a plausible approach to individual trajectory data sanitization," *IEEE Access*, no. 6, pp. 30110–30116, 2018.
- [8] X. Wang, Z. Zhang, Y. Luo, and Q. Yu, "Hierarchical interpolation point anonymity for trajectory privacy protection," *Intelligent Data Analysis*, vol. 23, no. 6, pp. 1397–1419, 2019.

- [9] S. Chen, A. Fu, J. Shen, Y. Shui, W. Huaqun, and S. Huaijiang, "A new differential privacy scheme base on Recurrent Neural Network for Dynamic trajectory privacy protection," *Journal of Network and Computer Applications*, vol. 168, no. 10, 2020.
- [10] J. Domingo-Ferrer, K. Muralidhar, and M. Bras-Amoros, "General confidentiality and utility metrics for privacy-preserving data publishing based on the permutation model," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2506–2517, 2021.
- [11] Z. Tu, K. Zhao, and F. Xu, Y. Li, L. Su, and D. Jin, Beyond k-anonymity: protect your trajectory from semantic attack," *Sensing, Communication and Networking*, pp. 1–9, 2017.
- [12] Z. W. Hu and J. Yang, "Trajectory privacy protection based on location semantic perception," *International Journal of Co-operative Information Systems*, vol. 28, no. 03, Article ID 1950006, 2019.
- [13] R. Tan, Y. Tao, W. Si, and Y. Y. Zhang, "Privacy preserving semantic trajectory data publishing for mobile location-based services," *Wireless Networks*, vol. 26, no. 8, pp. 5551–5560, 2020.
- [14] A. Ye, Q. Zhang, Y. Diao, J. Zhang, H. Deng, and B. Cheng, "A semantic-based approach for privacy-preserving in trajectory publishing," *IEEE Access*, vol. 8, pp. 184965–184975, 2020.
- [15] W. Zhang, W. Yang, H. Zhang, and X. Zhenqiang, "De-anonymization attack method of mobility trajectory data based on semantic trajectory pattern," in *Proceedings of the 14th EAI International Conference on Mobile Multimedia Communications*, pp. 354–366, Guiyang, China, November 2021.
- [16] J. I. Ya-Li, G. U. I. Xiao-Lin, D. A. I. Hui-Jun, and P. E. N. G. Zhen-Long, "Constructing users' interest regions with two steps for trajectory privacy protection," *Chinese Journal of Computers*, vol. 40, no. 12, pp. 2734–2747, 2017.