Modeling and Privacy Protection Analysis and Simulation of Mobile Internet User Activity Trajectory

Shuangyu Yang, Peng Tan, Juye Song

***Abstract*— In recent years, with the rapid development of technology, location-based services have brought many conveniences to people's lives, but also brought risks of privacy leakage. Attackers can use users' trajectory information to deduce their home address, work address and other private information, which poses significant hazards and risks to users' lives. To address this challenge, there is a personalized privacy protection scheme that marks the semantic attributes of users' trajectory points based on existing POI semantic information. The resulting set of semantic attributes is classified and abstracted to obtain a classification tree, and three types of stop points are designed to select appropriate regions of POIs for replacement to generate fake trajectories. Finally, the fake trajectories are tested for obstacle crossings, and if they cross any obstacles, POIs are reselected for replacement. This article analyzes and simulates this scheme.**

***Index Terms*— privacy protection, selection of false positions, selection of stop points, semantic attribute**

I. BACKGROUND

1.1 Research status

Privacy protection of location is a problem that all types of location-related app servers must face. The existing anonymous protection of user trajectories is achieved through K-anonymity [1]. This method confuses real positions by adding K-1 fake trajectories or positions to trajectories or positions.

Building on this, Liu et al. [2] focus on personalized services through clustering trajectory preservation algorithms, and Wen [3] and Tu [4] attempt to consider the semantics of frequently accessed locations in trajectories. However, the K-anonymity method still has insufficient consideration for spatial, temporal and semantic features, and attackers can easily remove fake positions or trajectories through background knowledge or reasoning, resulting in a decrease in protection effectiveness and exposure of user trajectory privacy.

Another common method is through perturbation. This method perturbs the user's true position by adding noise to make it difficult for adversaries to determine whether a particular trajectory record is included in the database, thereby achieving the requirement of privacy protection. Huo [5] proposed a personalized trajectory noise tree method, which is based on the assumption that the data collector is "trustworthy", but many third-party collectors are not reliable. Although adding noise can make the location data more difficult to identify, it is still possible to deduce the real position through some technical means, including differential privacy attacks, reconstruction attacks, and auxiliary information attacks.

1.2 related work

This paper aims to study, analyze, and visually simulate the effectiveness of a personalized semantic trajectory privacy preservation method. Firstly, semantic attributes were considered and personalized protection was implemented. Not all points in the user trajectory contain sensitive privacy information, therefore, this paper designated stop points for the long-time stationary points, roaming points and semantic attribute sensitive points in the trajectory, and selected appropriate POIs for replacement. After the replacement was completed, a fake trajectory was generated and further consideration was given to the impact of obstacles.

II. PRIVACY PROTECTION ALGORITHM

2.1 Algorithm process

Initially, each trajectory point was semantically labeled according to the simulated user trajectory and imported POI semantic attribute table, and the labeled semantic attribute set was recursively abstracted to obtain the classification tree. Then, three types of stop points were extracted from the trajectory, namely, long-time stationary points, roaming points, and semantic attribute sensitive points. Finally, considering factors such as the user's running speed and the sudden change in trajectory direction, an appropriate selection area was constructed from which suitable POIs were selected as fake locations to reconstruct the trajectory. The reconstructed trajectory was checked to see if it passed through obstacles, and if so, a new fake location was chosen.

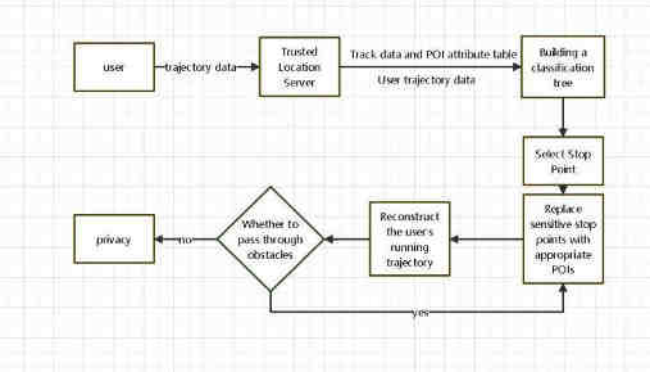


Fig.1 Algorithm flow

2.2 Data Collection and Preprocessing

To begin with, the user's current location is simulated by marking a point on the map. The user's movement trajectory is then annotated on the map in chronological order to obtain the simulated user trajectory, where each point on the trajectory

Manuscript received April 26, 2023

Shuangyu Yang, School of Software, Tiangong University, Tianjin, China

Peng Tan, School of Software, Tiangong University, Tianjin, China, 18064952003

Juye Song, School of Software, Tiangong University, Tianjin, China, 15368266421

includes latitude, longitude, and timestamp information. Next, the points on the trajectory are processed. Before doing so, the POI data for Beijing is imported. This data contains the latitude and longitude information for various entities in Beijing along with their corresponding semantic attributes, which are divided into three levels. Then, the simulated user trajectory is semantically labeled using this POI dataset. Specifically, by calculating the Euclidean distance between each trajectory point and all POIs, the semantic attribute of each trajectory point is matched with the closest POI. Finally, the semantic attribute of each trajectory point is labeled with the semantic attribute of the nearest POI. After labeling the semantic attributes, a classification tree is constructed. The semantic attributes are treated as leaf nodes of the classification tree. Based on their semantic meanings, they are divided into different categories. Leaf nodes with similar semantic meanings are abstracted into a generalized category word as a parent node. This process is repeated, and a root node is ultimately abstracted and generalized to complete the construction of the classification tree.

2.3 Extract stop point

There are three types of designated stop points, and different extraction methods are used for the different types. For the first two types of stop points, a time threshold and a distance threshold are set to simultaneously extract the stops, and then the most significant semantic attribute is selected as the sensitive semantic attribute. For the third type of stop point, the class with the highest semantic proportion was defined as the sensitive semantic attribute by looking for the class with the highest semantic proportion in the first two types of stop points. Depending on the privacy level, the corresponding number of semantic categories are mapped to different levels of the classification tree.

2.4 Selection of false positions

Considering the factors of user speed and reverse mutation of the trajectory, two asymmetric semicircles are used to construct an appropriate selection area. First, two semicircles are constructed with the stop point as the center and a third of the radius between the current and previous and next 100 points of the original trajectory. Then, two cases are separated to find suitable POIs for selection. If the previous and next points are on the same side of the current stop point, check if the POI is also on the same side and if it is within the radius, If not in the same side, only the radius needs to be satisfied. After finding all the POIs that meet the conditions, one POI is randomly selected as a fake location. In consideration of the fact that the reconstructed trajectory may pass through obstacles in actual situations, a list of randomly generated GPS coordinates was used to simulate obstacles in the reconstructed trajectory segment. The trajectory segment was then checked for intersection with the obstacles, and if there is any intersection, a new POI was selected for replacement.

III. IMPLEMENTATION OF ALGORITHMS

Compared with reference [6], the algorithm implementation in constructing POI selection area and reconstructing the trajectory was modified. In reference[6], the construction of POI area was based on two asymmetrical semicircles with a radius of half the distance between the previous and next stop points of each stop point. However, the region constructed in this way would be large when the two stop points were far apart, increasing the possibility of trajectory mutation when selecting appropriate POI as false locations in this region. The region size should be neither too big to increase the possibility of trajectory mutation nor too small to cause little change between the false and original trajectories, thus failing to achieve privacy protection. Therefore, we set the region radius as one third of the distance between each stop point and the previous/next 100th point in its original trajectory, constructing two asymmetrical semicircles. In the trajectory reconstruction part, not only stop points, but also some adjacent trajectory points near the stop points were replaced in reference [6] to reduce the impact of trajectory mutation. However, we had already considered the impact of trajectory mutation in constructing POI selection areas, so we only replaced the stop points. Regarding the impact of obstacles on trajectories, we randomly generated coordinate points to simulate obstacles, and when the reconstructed trajectory segment passed through them, we would replace the stop points accordingly. This method not only simulated the impact of obstacles on trajectories more realistically but also reduced the possibility of trajectory mutation.

IV. ARCHITECTURE

The system was based on B/S architecture, consisting of a browser-side and a server-side, where html+JavaScript was used to develop the browser-side, and Python language was utilized for deep learning and interface calls in the server-side. The front-end and back-end were connected via API interfaces, and the original trajectory points were selected and sent to the server-side to form the trajectory through calling the Baidu Map API. After confirmation, the back-end called the deep learning module to generate the protected trajectory and saved it to a file. After a period of time, the front-end page queried and called the generated trajectory through an interface and marked it on a Baidu Map by using the Baidu Map API.

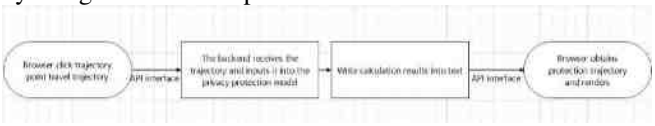


Fig.2 System architecture diagram

V. SIMULATION RESULTS

We conducted multiple experiments using the system and the results indicated that the algorithm was able to achieve privacy preservation effects by reconstructing the user trajectory while not deviating too much from the original trajectory. As shown in Fig.3, the red line represents the user's original trajectory and the blue line represents the reconstructed trajectory. It can be observed that the reconstructed trajectory replaced the sensitive points in the original trajectory without causing any major disruptions.

However, there are some limitations to this study. The simulation used data that was manually marked from a series of continuous points in the frontend to simulate the user trajectory, which may not fully represent real trajectories. In addition, we only tested the algorithm on a single input dataset and cannot determine its performance in a broader range of inputs.

In the future, research should focus on testing the algorithm's robustness and generalizability on more complex models and a wider range of inputs to determine its effectiveness.

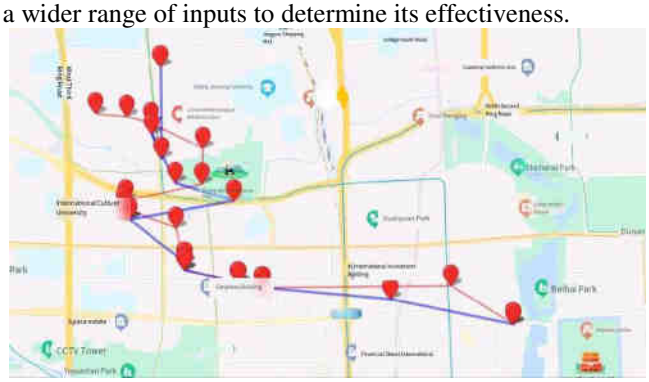


Fig.3 Result diagram

VI. CONCLUSION

The user trajectory contains a lot of sensitive information, and the leakage of real trajectory poses a serious threat to the user's privacy and security. In this paper, a personalized trajectory privacy preservation method was studied, analyzed, and simulated. The system achieved the reconstruction of the user's original trajectory. In order to achieve personalized protection for different user trajectories, the method fully considers the role of semantic attributes, designs three different types of stopping points, and selects appropriate fake locations for replacement to generate fake trajectories. Moreover, the method takes into account the possibility of trajectory direction mutation and obstacles, and adopts corresponding methods to minimize their impact. However, the method introduced in this paper is only applicable to the protection of static trajectories, and the obstacles are randomly simulated without being verified in the real space. In the future, further tests and extensions will be carried out in the real space.

REFERENCE

[1] Sweeney L. k-anonymity: A model for protecting privacy[J]. International journal of uncertainty, fuzziness and knowledge-based systems, 2002, 10(05): 557-570.

[2] Liu X, Xie Q, Wang L. Personalized extended (α, k)-anonymity model for privacy-preserving data publishing[J]. Concurrency and Computation: Practice and Experience, 2017, 29(6): e3886.

[3] Wen R, Cheng W, Huang H, et al. Privacy preserving trajectory data publishing with personalized differential privacy[C]//2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). IEEE, 2020: 313-320.

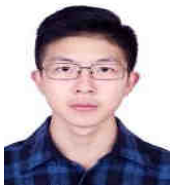
[4] Tu Z, Zhao K, Xu F, et al. Beyond k-anonymity: protect your trajectory from semantic attack[C]//2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). IEEE, 2017: 1-9.

[5] Zheng Huo, Xiaofeng Meng. A trajectory data publishing method satisfying differential privacy. [J]. Journal of Computer Research,2018,41(02):400-412.(in Chinese)

[6] Dai Y, Shao J, Wei C, et al. Personalized semantic trajectory privacy preservation through trajectory reconstruction[J]. World Wide Web, 2018, 21: 875-914.

Shuangyu Yang Female, Chongqing Province, is a junior student in the School of Software at Tiangong University, Tianjin Xiqing District Guest Water, 399 West Road, 300387, China.

Peng Tan Male, Sichuan Province, is a junior student in the School of Software at Tiangong University, Tianjin Xiqing District Guest Water, 399 West Road, 300387, China.



Juye Song Male, Yunnan Province, is a junior student in the School of Software at Tiangong University, Tianjin Xiqing District Guest Water, 399 West Road, 300387, China.