Journal of Computers Vol. 36 No. 5, October 2025, pp. 21-37
https://doi.org/10.63367/199115992025103605002

Personalized Location Privacy Protection Method against Background Knowledge Attacks

Xuejun Zhang*, Jiaxin Luo, Chengze Li, Bin Zhang, and Yixing Yang

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, 730070, China
xuejunzhang@lzjtu.edu.cn

Received 10 September 2024; Revised 9 December 2024; Accepted 25 January 2025

Abstract. With the widespread use of mobile communication devices, Location-Based Services (LBS) provides diversified services to users. However, attackers can potentially infer users' private information by analyzing location trajectories during the use of LBS. To address this issue, this paper proposes a personalized location privacy protection method to counter background knowledge attacks. The proposed method first evaluates the sensitivity of location points using user-defined sensitivity levels and historical query data, and then selects an optimal route, minimizing both privacy leakage and travel costs from multiple routes recommended by the navigation system. Based on the total sensitivity and travel distance, a personalized privacy budget is assigned to the user. To protect location privacy, Laplace noise, adhering to the differential privacy mechanism, is added to the requested location points. Security analysis and simulation results demonstrate that the proposed method effectively provides personalized location privacy protection while maintaining the quality of service.

Keywords: location-based services, differential privacy, personalized location privacy protection, background knowledge inference attack, location sensitivity, trajectory privacy

1 Introduction

As a new network service model, location-based services (LBS) meet users' increasing material demands and play a significant role in promoting the sustainable development of the Internet [1]. However, when using LBS, users must provide both location information and query attributes. The contextual information associated with these data may expose users' privacy [2]. Currently, users' privacy concerns significantly hinder the healthy development of the LBS industry. Therefore, it is crucial to study location privacy protection methods capable of ensuring both privacy and service quality.

To protect location privacy, scholars have conducted extensive research. Among these, K-anonymity is one of the most commonly employed methods. Privacy protection methods based on K-anonymity or its extensions typically involve data generalization, in which each user record is replaced by multiple false data entries. This kind of approaches makes it challenging for attackers to differentiate real data from fake data, thereby protecting users' private information. Distortion-based techniques, in contrast, intentionally modify real information in LBS queries to prevent attackers from directly accessing private data. The virtual trajectory approach uses a distributed architecture to generate multiple virtual trajectories based on users' actual trajectories, thereby minimizing the risk of privacy leakage.

However, all the aforementioned privacy-preserving techniques have inherent limitations and struggle to resist background knowledge attacks, which can lead to user privacy breaches through trajectory data mining and analysis. To address this issue, the concept of differential privacy (DP) was introduced as a solution for protecting user privacy in trajectory data [3]. Differential privacy is supported by a rigorous mathematical theory that ensures that accessing a trajectory dataset protected by DP yields indistinguishable results, regardless of the background knowledge possessed by adversaries. This prevents attackers from utilizing their background knowledge to infer sensitive personal information. However, traditional differential privacy approaches allocate the privacy budget evenly across all data points, leading to insufficient protection for sensitive locations and hindering effective mitigation of privacy leakage risks during continuous queries.

To address these issues, we propose a novel personalized location privacy protection approach, which defends

* Corresponding Author

against background knowledge attacks and mitigates the risk of privacy leakage in continuous queries. It first determines the sensitivity of sensitive location points by integrating user preferences and historical query data. A fuzzy mathematical model is employed to resolve the inconsistency in users' sensitivity boundary judgments. Next, it selects a route with minimal privacy leakage and low driving cost from the multiple paths recommended by the navigation system. Finally, our method allocates a personalized privacy budget for the user based on the sensitivity degree and the distance to sensitive location points. Laplace noise, consistent with the differential privacy mechanism, is then employed to protect location privacy.

The main contributions of this paper are as follows:

1) We propose a sensitivity determination algorithm that combines the user's subjective sensitivity degree with the objective historical query probability, accurately reflecting the user's privacy preferences for different sensitive locations and enabling personalized privacy protection.

2) We propose an optimal path selection algorithm that considers the complex characteristics of the road network environment to recommend a path with minimal privacy leakage and low driving cost. This algorithm balances privacy protection, travel cost reduction, and service quality improvement.

3) We propose a personalized location privacy protection algorithm, which allocates the privacy budget by considering the distance between the requested location point and sensitive locations, as well as the sensitivity of the data. Consequently, it ensures that different location points are granted varying levels of privacy protection. This approach offers users tailored protection for their location privacy, thereby circumventing a uniform noise distribution across all location points.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 outlines the systematic framework and details our method, including the sensitivity determination algorithm, the optimal path selection algorithm, and the personalized location privacy protection algorithm. Section 4 validates the method through experiments and compares it with existing approaches. Section 5 concludes our main contributions and an outlooks for future research directions.

2 Related Work

To address the problem of user trajectory privacy leakage, researchers have proposed various privacy-preserving methods to effectively protect user location information, which include generalization, anonymization, and traditional encryption techniques. Li et al. [4] proposed a location privacy protection system based on online virtual trajectory generation, called AnotherMe. It simulates the movement patterns and points of interest of real users to create virtual user profiles and trajectories. Zhang et al. [5] proposed PriSC, a spatial crowdsourcing framework for decentralized location privacy preservation, which leverages blockchain to replace traditional centralized crowdsourcing servers and employs homomorphic encryption to ensure the confidentiality of task location policies, preventing unauthorized access and data leakage. Wang et al. [6] introduced a method to characterize semantic distance by calculating the number of hops between nodes in a location semantic tree, measuring semantic differences accordingly. However, these methods are often vulnerable to privacy attacks by adversaries with background knowledge. They fail to effectively safeguard location information in successive queries and lack robustness in privacy protection.

To address these challenges, differential privacy (DP) has been introduced as a solution for protecting user trajectory data. Zhang et al. [7] applied DP mechanisms to recommendation systems, improving recommendation accuracy based on multidimensional spatial indexing and n-gram trees. Sun et al. [8] proposed a semantically constrained local DP (SLDP) model that reduces noise by segmenting trajectories into overlapping sub-trajectories. They also utilized the positional trajectory synthesis (PLTS) framework to generate semantically relevant synthetic positional trajectory datasets. Chen et al. [9] proposed an optimal privacy budget allocation algorithm for preserving user privacy in differential privacy based on public transportation trajectory data distribution. The algorithm uses a reachable graph to store trajectory data and a query probability model to quantify the probability of a trajectory location pair being queried, thus optimizing the privacy budget for each prefix tree node.

However, traditional DP mechanisms provide the same level of privacy protection for all requested locations, which may result in under/over-protection of certain locations, a limitation that means they cannot provide personalized location privacy protection based on users' privacy preferences. To better meet users' personalized needs for location privacy, researchers have begun to explore personalized differential privacy protection methods that dynamically adjust the privacy protection level according to users' actual needs and scenarios. Li et al. [10] proposed an approach for continuous location privacy protection based on DP, dividing sensitive sections according to the road network's topological relationship and adding noise to location points through a location tree. However, it only divides privacy protection into levels and cannot meet users' personalized privacy needs. Xu et al. [11] used DP mechanisms to provide different levels of privacy protection for different request locations, meeting individualized privacy needs. However, they fail to protect privacy-insensitive locations, which may expose the user's trajectory when there are many insensitive locations. Wang et al. [12] proposed a DP-based trajectory anonymity protection method using the Frechet distance similarity measure for personalized trajectory segmentation. This method accurately captures features between trajectories and stores them in an extensible personalized trajectory graph to meet users' privacy needs for trajectory anonymization. However, the method primarily focuses on trajectory segmentation and may not be suitable for all LBS. Yuan et al. [13] proposed a mechanism called SCTP for protecting semantic associative privacy of user trajectory data in DP based on Internet of Things (IoT) applications. SCTP utilizes Hidden Markov Model (HMM) for trajectory prediction and assigns DP levels to different locations based on semantic frequencies. However, its model is complex and requires certain computational resources. Wu et al. [14] proposed a trajectory correlation privacy protection mechanism (TCPP) based on DP for protecting the correlation between multiple user trajectories, investigated the correlation privacy problem between multiple user trajectories and proposed a personalized privacy budget allocation strategy but their privacy budget allocation strategy is only based on the user's time and distance and does not take into account other personalized factors.

To address the above problems, this paper proposes a personalized location privacy protection method against background knowledge attacks. The method determines the sensitivity of the location based on the user's preference for sensitive locations and the historical query probability, and provides DP protection for different sensitive location points, followed by selecting a route with low travel cost according to the optimal path selection algorithm. This ensures service quality and provides personalized location privacy protection, effectively reducing the risk of privacy leakage in continuous queries. The method consists of three steps: (1) a sensitivity determination algorithm is introduced to evaluate the sensitivity of different locations based on the user's historical data and preferences. This solves the problem of inconsistent user sensitivity thresholds; (2) a low-cost and privacy-protecting route is recommended, taking into account the complex characteristics of the road network environment such as travel distance, congestion, and road class; (3) a personalized privacy budget is assigned to the user by the determined sensitivity and sensitivity distance, and a Laplace noise conforming to the DP mechanism is added to the requested location point to protect the location privacy.

3 Method

3.1 System Architecture

Road network users generally refer to individuals who own smart devices or vehicles equipped with smart control platforms, which possess substantial computing power and adhere to traffic regulations. Therefore, standalone architectures are widely used LBS in protection methods, such as those described in references [10, 15]. The system architecture proposed in this paper consists of four key entities: the positioning system, road network users, Wi-Fi access points (APs), and the LBS server.

As illustrated in Fig. 1, the process for providing personalized location privacy protection services to users is as follows:

1) Road network users retrieve their location information from the positioning system.

2) Users obtain semantic location data and historical query probabilities for their current range from the Wi-Fi access point.

3) Based on the location semantics obtained in step 2, users identify sensitive locations and determine their sensitivity using historical query probabilities. Next, they select a driving route that best meets their preferences from multiple options recommended by the navigation system. The selection considers four factors: driving distance, road grade, traffic congestion, and total sensitive distance. Finally, personalized privacy budgets are assigned to the requested location points according to their sensitivity and proximity to sensitive locations. To protect location privacy, Laplace noise is added, ensuring compliance with the differential privacy mechanism.

4) Users send location data requests, enhanced with differential privacy protection, to the LBS server.

5) The LBS server processes the location data submitted by users and returns the corresponding query results.

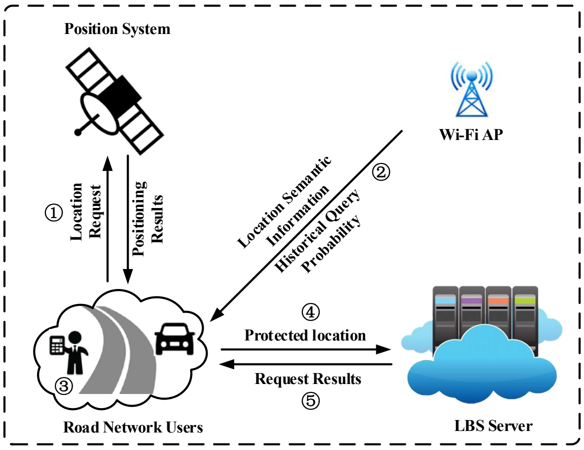


Fig. 1. System architecture

Problem Definition

Definition 1 (Query probability). The area around the user is divided into grids, each grid corresponds to a location cell, and the query probability qp_i for each location cell is calculated as follows:

$$qp_i = \frac{num(i)}{\sum_{i=1}^{n^2} num(i)}. \tag{1}$$

where $num(i)$ denotes the number of query counts for the location cell corresponding to the grid and $\sum_{i=1}^{n^2} num(i)$ denotes the sum of query counts for all location cells.

Definition 2 ($\varepsilon - DP$). There is a random algorithm M, $\mathfrak{R}$ is a set of all outputs of algorithm M, and $|D \oplus D'| = |(D \cup D') - (D \cap D')| = 1$ is satisfied for any two adjacent data sets D and D', where $S \subset \mathfrak{R}$. If random algorithm M meets

$$\Pr[M(D) \in S] \le e^{\varepsilon \times |D \oplus D'|} \times \Pr[M(D') \in S]. \tag{2}$$

Then, random algorithm M can provide $\varepsilon - DP$, where $\oplus$ denotes the set of symmetric differences of the two sets, ε represents the privacy budget, $\Pr[\cdot]$ represents the probability of the random algorithm M calculating the results of two adjacent datasets D and D'.

Definition 3 (Laplace mechanism). Function $f: D \rightarrow R^d$, its sensitivity is $\Delta_1 f$. Formula (3) provides $\varepsilon - DP$:

$$f'(D) = f(D) + g_{Laplace}(\Delta_1 f / \varepsilon). \tag{3}$$

where $g_{Laplace}(\Delta_1 f/\varepsilon)$ represents a random noise that follows a Laplace distribution, with a scale parameter of $\Delta_1 f/\varepsilon$.

3.2 Sensitivity Determination Algorithm

This section proposes a sensitivity determination algorithm to address the varying levels of sensitivity that different users may assign to the same sensitive location. When users access the service for the first time, they select sensitive location semantics based on the semantic information retrieved from the Wi-Fi access point (AP), as shown in formula (4):

$$SEN_u = \{(stype_1, deg_1), (stype_2, deg_2), ..., (stype_k, deg_k)\} . \quad (4)$$

where $stype_i$ represents the ith sensitive location semantic type, deg_i represents the sensitivity degree that the user sets for $stype_i$, with a value range of [0, 1].

Although users set the sensitivity level deg_i based on their own perception, different users have different subjective awareness. To avoid significant deviations in the sensitivity level deg_i inputted by different users, this section uses fuzzy mathematical models to address the issue of inconsistent judgment on sensitivity boundaries by users. SL is used to evaluate the sensitivity level of deg_i , which is generally specified by experts in the field of privacy protection. This article sets three sensitivity levels based on literature [16]: l = 0.7 (low), m = 0.5 (middle) and h = 0.2 (high), where l, m and h represent low sensitivity, middle sensitivity, and high sensitivity, respectively, with $SL \in \{l, m, h\}$.

$$A_1(x) = \begin{cases} 1 & 0 \le x \le l \\ \dfrac{m-x}{m-l} & l < x < m \\ 0 & x \ge m \end{cases} , \quad (5)$$

$$A_2(x) = \begin{cases} 0, & x \le l \\ \dfrac{x-l}{m-l}, & l < x < m \\ \dfrac{h-x}{h-m}, & m < x < h \\ 0, & x \ge h \end{cases} , \quad (6)$$

$$A_3(x) = \begin{cases} 0, & x \le m \\ \dfrac{x-m}{h-m}, & m < x < h \\ 1, & h \le x \le 1 \end{cases} . \quad (7)$$

Formula (5), (6) and (7) respectively represent the membership functions for evaluating the small type (mild sensitivity), middle type (moderate sensitivity), and large type (high sensitivity) levels of sensitivity, where x denotes the sensitivity level set by the user.

Protecting only the locations deemed sensitive by users may still lead to privacy leakage. For example, when the number of sensitive locations is much smaller than the number of insensitive locations, the likelihood of an attacker inferring the user's true trajectory through insensitive locations increases. To provide more comprehensive privacy protection, this method combines subjective sensitivity with objective historical query probabilities to construct the sensitivity of a given location's semantics for the user as follows:

$$S_i(x_i, y_i) = \omega_1 SL + \omega_2 qp_i . \quad (8)$$

where SL denotes a user's subjective sensitivity to a location, and qp_i denotes the location's historical query probability, reflecting how frequently it has been queried. ω_1 and ω_2 are weight indicators for SL and qp_i. The weights were determined using the following fuzzy comprehensive evaluation model.

The matrix form is used to express the relative importance of the two factors of subjective sensitivity level and objective historical query probability on user sensitivity [17], to establish a fuzzy judgment matrix $F=(f_{ij})_{n\times n}$, and to determine the matrix elements according to the following fuzzy logic, e_i, e_j, denote the two factors of subjective sensitivity level and objective historical query probability, and f_{ij} denote the degree of influence of the e_i factor on the e_j factor, denoted as:

$$f_{ij} = \begin{cases} 1 & e_i \textit{ is more important than } e_j \\ 0.5 & e_i \textit{ and } e_j \textit{ are equally important} \\ 0 & e_j \textit{ is more important than } e_i \end{cases}, \tag{9}$$

Then, the fuzzy consistency judgment matrix R is constructed to measure the consistency between the elements in the fuzzy matrix, and the fuzzy judgment matrix F is summed by rows, denoted as:

$$r_i = \sum_{j=1}^{n} f_{ij} , \tag{10}$$

Perform the following mathematical transformations:

$$r_{ij} = \frac{r_i - r_j}{2n} + 0.5 , \tag{11}$$

Then, the resulting matrix is a fuzzy, consistent matrix.

Finally, the fuzzy agreement matrix is normalized and the weight ω_i of factor e_i is calculated as:

$$\omega_i = \frac{\sum_{j=1}^{n}(r_{ij}) - 0.5}{\sum_{i=1}^{n}(\sum_{j=1}^{n}(r_{ij}) - 0.5)} . \tag{12}$$

Algorithm 1. Sensitivity determination algorithm

Input: location semantic information, historical query probability, $sp_i\ (x_i, y_i)$

Output: $S_i\ (x_i, y_i)$ of the sensitive location point $sp_i\ (x_i, y_i)$

1. User select a set of sensitive location semantics from the global location semantic information based on personal preferences
   $SEN_u = \{(stype_1, deg_1), (stype_2, deg_2), ..., (stype_k, deg_k)\}$
2. Determine the deg_j based on the $stype_j$ of the location point $sp_i\ (x_i, y_i)$
3. max = $max(max(A_1(\text{deg}_j), A_2(\text{deg}_j)), A_3(\text{deg}_j))$
4. **if** max == $A_1\ (deg_j)$ **then**
5. $\quad SL = l$
6. **end if**
7. **if** max == $A_2\ (deg_j)$ **then**
8. $\quad SL = m$
9. **end if**
10. **if** max == $A_3\ (deg_j)$ **then**
11. $\quad SL = h$
12. **end if**
13. Calculate the qp_i of the location unit where the $sp_i\ (x_i, y_i)$ is located
14. Normalize the calculated sensitivity level SL and query probability qp_i.
15. Calculate the weight ω_i of sensitivity level and query probability
16. Calculate the user sensitivity level $S_i\ (x_i, y_i)$.
17. **return** $S_i\ (x_i, y_i)$

Algorithm 1-3 lines determine the subjective sensitivity degree deg_i of sensitive location point $sp_i\ (x_i, y_i)$, and use the membership function to calculate the membership level of deg_i for three levels. The fourth line calculates the maximum value among the membership functions for the three types: small type, middle type, and large type. Lines 5-13 determine the sensitivity level based on the calculated maximum membership degree. Lines 14-18 calculate the final sensitivity $S_i\ (x_i, y_i)$ based on the weight of sensitivity level and query probability and return it.

3.3 Optimal Path Selection Algorithm

This section selects a route with minimal privacy leakage and low driving cost for road network users based on four attributes: travel distance, traffic congestion, road classification, and total sensitive distance. Due to the differing scales of these attributes, it is necessary to preprocess and normalize the data before evaluation. The final values for all attributes are normalized to the range [0, 1]. Then, weights are assigned to each attribute according to the user's personal preferences. Finally, a route that satisfies the user's requirements is selected. The calculations for the four attributes are as follows:

1) Travel distance. The travel distance is the sum of distances between the starting location point of the current route and the location of the queried content, expressed as:

$$a_1^j = \sum_{i=1}^{m} \sqrt{(x_{i+1}^j - x_i^j)^2 + (y_{i+1}^j - y_i^j)^2}\,,\ j = 1, 2, \ldots, r\,, \quad (\mathbf{13})$$

where j represents one of the r recommended routes by the navigation system; m represents the number of road sections on the jth route.

2) Congestion situation. Evaluation indicators for traffic congestion generally include queue length, average speed, etc. In this section, the average speed (km/h) is used to measure the congestion situation, denoted as:

$$\overline{v}^j = \frac{1}{m}\sum_{i=1}^{m} \overline{v}_i\,,\ j = 1, 2, \ldots, r\,. \quad (\mathbf{14})$$

where m represents the total number of road sections on the current route j; $\overline{v_i}$ is the average speed on the road section i. The congestion level is categorized into four states: smooth, light congestion, moderate congestion, and severe congestion based on the average speed [18]. The calculation for the congestion situation is as follows:

$$a_2^j = \begin{cases} 1 & \overline{v}^j > 30 \\ 0.75 & 20 < \overline{v}^j \le 30 \\ 0.5 & 10 < \overline{v}^j \le 20 \\ 0.25 & \overline{v}^j < 10 \end{cases}, \quad (\mathbf{15})$$

For ease of calculation, the four congestion situations are divided based on the intervals of the average speed. A higher attribute value corresponds to a higher average speed, indicating smoother road conditions.

3) Road grade. According to the road grade classification, this section divides the road grade into three levels: arterial roads, collector roads, and local roads [19]. When selecting a path, it is recommended to choose a section with a higher grade as much as possible. The calculation is shown in (16):

$$a_3^j = \frac{R_i}{\max\{R_i\}}\,, \quad (\mathbf{16})$$

where R_i represents the three levels of the road, with values of 3, 2, and 1, where a higher grade corresponds to a larger value.

4) Total sensitive distance. The total sensitive distance refers to the sum of distances from all requested location points on the current route to their nearest sensitive location points. The calculation process is shown in (17):

$$a_4^j = \sum_{i,k=1}^{n,m} \lambda dis(rp_i^j(x_i, y_i), sp_k^j(x_k, y_k)),\ j = 1, 2, \ldots, r\,, \quad (\mathbf{17})$$

where *m* represents the number of sensitive location points selected by the user; *n* represents the number of request location points; $dis(rp_i^j, sp_k^j)$ represents the distance from the *i*th requested location point on route *j* to its nearest sensitive location point; The sensitive location closest to the requested location point is set to $\lambda = 1$, while the rest are set to $\lambda = 0$.

After determining the four attributes, first construct a multi-attribute decision matrix *X* for $r \times 4$ based on the recommended routes of the navigation system; Secondly, forward processing for the extremely small driving distance; Then, the standardization matrix *A* is obtained by eliminating the dimensions of the four indicators through standardization processing, as follows:

$$A=\begin{bmatrix} a_1^1 & a_2^1 & a_3^1 & a_4^1 \\ a_1^2 & a_2^2 & a_3^2 & a_4^2 \\ \cdots & \cdots & \cdots & \cdots \\ a_1^r & a_2^r & a_3^r & a_4^r \end{bmatrix}. \tag{18}$$

Finally, weights ω_1, ω_2, ω_3, ω_4 are assigned according to the user's emphasis on different attributes according to the weight determination method proposed in Section 3. These weights reflect the user's focus on four attributes, namely, driving distance, congestion, road class, and sensitive total distance. Then, the score s is calculated for each route and s reflects the combined performance of the *j*th route for the four attributes.

The route with the highest score is selected as the best because it has the largest score value, implying that it performs optimally on the four attributes and best meets the user's privacy preferences and driving cost minimization requirements.

Algorithm 2. Optimal path selection

Input: Recommended *r* routes by the navigation system

Output: The route with the highest score *j*

1. Calculate the travel distance, congestion situation, road grade, and total sensitive distance of route *j*
2. Determine the standardization matrix *A* that has undergone forward normalization and standardization processing.
3. Calculate each attribute weight value ω_i according to formula (12) and calculate the maximum value $A_i^+ = max\{a_i^1, a_i^2, ..., a_i^r\}$ and minimum value $A_i^- = min\{a_i^1, a_i^2, ..., a_i^r\}$ in column i, i.e., the maximum and minimum value of each attribute in all routes.
4. Calculate the distance $D_j^+ = \sqrt{\omega_i \sum_{i=1}^{4} (A_i^+ - a_i^j)^2}$ between each attribute value in the *j*th route and its maximum value A_i^+
5. Calculate the distance $D_j^- = \sqrt{\omega_i \sum_{i=1}^{4} (A_i^- - a_i^j)^2}$ between each attribute value on the *j*th route and its minimum value A_i^-
6. Calculate the score $S_j = \dfrac{D_j^-}{D_j^+ + D_j^-}$ for the *j*th route and calculate the normalized result $\tilde{S}_j = \dfrac{S_j}{\sum_{j=1}^{r} S_j}$
7. Sort the calculated $\tilde{S}_j$ of all routes and return the highest rated route *j*

In algorithm 2, four attributes that affect path selection were identified. Then, the user was selected from the r routes recommended routes by the navigation system to choose the route with the highest score. The first line determines the calculation method for four attributes: travel distance, congestion situation, road grade, and total sensitive distance. The second line performs forward normalization on the decision matrix and eliminates dimensionality through standardization to obtain the standardized matrix *A*.

3.4 Personalized Location Privacy Protection Algorithm

Building on the route selection process outlined in Section 3.2, which prioritizes low privacy leakage and low travel costs, this section introduces perturbation by adding Laplace noise, in accordance with the differential privacy mechanism, to the users' true requested location points. Existing studies usually assign the same privacy budget on all requested location points or only on points of interest, which may lead to under or over protection of certain points. To address this issue, we propose a personalized privacy budget allocation algorithm. This algorithm adheres to differential privacy requirements by adding Laplace noise to perturb the requested location points. This approach delivers personalized location privacy protection for users while maintaining service quality.

In order to satisfy user's personalized privacy requirement, we allocate the privacy budget considering both the user's sensitivity to sensitive location points and the distance between the requested location point and the nearest sensitive location point. The total privacy budget is allocated in a way that takes into account the different needs of each requested location point, aiming to improve service quality while effectively protecting users' location privacy. The formula for allocating privacy budgets to different requested location points is as follows:

$$\varepsilon_i = [\alpha_1 S_k(x_k, y_k) + \alpha_2 \frac{dis(rp_i(x_i, y_i), sp_k(x_k, y_k))}{\sum_i^n dis(rp_i(x_i, y_i), sp_k(x_k, y_k))}]\varepsilon_{total} . \tag{19}$$

where α_1 and α_2 represent the degree of importance that users attach to the two factors, $S_k(x_k, y_k)$ represents the sensitivity of the sensitive location closest to the requested location point, $dis(rp_i(x_i, y_i), sp_k(x_k, y_k))$ represent the distance between the requested location point and its nearest sensitive location point, and ε_{total} represents the total privacy budget.

Algorithm 3. Personalized privacy budget allocation algorithm

Input: Optimal path j; Sensitive location point closest to the requested location point $sp_k(x_k, y_k)$; ε_{total}

Output: Privacy budget ε_i for location point $rp_i(x_i, y_i)$

1. Determine the sensitivity $S_k(x_k, y_k)$ of sensitive location point $sp_k(x_k, y_k)$
2. Calculate the ratio of $dis(rp_i(x_i, y_i), sp_k(x_k, y_k))$ to $\sum_i^n dis(rp_i(x_i, y_i), sp_k(x_k, y_k))$
3. Normalize the calculated sensitivity and distance ratio
4. Set weight indicators α_1 and α_2 according to formula (12)
5. Calculate the privacy budget ε_i of the requested location point according to formula (19)
6. **return** ε_i

4 Experiment and Analysis

4.1 Experimental Environment

To verify the effectiveness of the personalized location privacy protection algorithm against background knowledge attacks, experiments were conducted using the Geolife real dataset. This dataset, recorded by Microsoft Research, contains 17,621 mobile trajectory data points from 182 users over a period of 5 years. The experiment extracts trajectory data from the Geolife dataset, specifically from users whose longitude values range from [116.2, 116.6] and latitude values range from [39.8, 40]. Miller's projection is then used to convert the longitude and latitude values into X and Y coordinates in the Cartesian coordinate system on the plane. The trajectory map of the experimental data is shown in Fig. 2, where blue points represent the distribution of user trajectories. Considering that each user has different requirements for sensitive locations, some points are randomly selected from these location distribution points and labeled red to indicate sensitive locations for users in the current region.

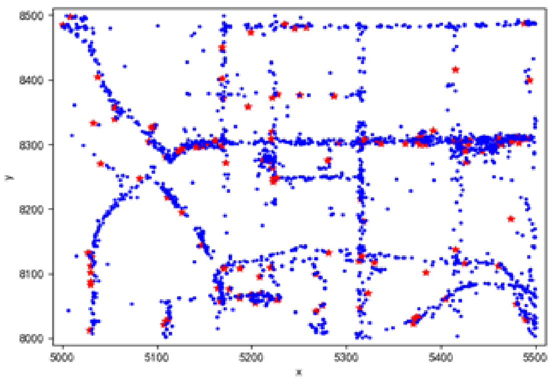


Fig. 2. Trajectory map of the experimental data

The experiment was developed using PyCharm as the development platform and implemented using Python language programming. The hardware environment is Intel (R) Core (TM) i5-7300HQ CPU @ 2.50GHz, memory of 8.00GB, and operating system of Windows 10.

4.2 Evaluation Metrics

Privacy Quality (PQ): Privacy quality is an important metric for evaluating the effectiveness of trajectory privacy-preserving methods based on differential privacy. The greater the distance between the perturbed trajectory points and the original trajectory points, the higher the privacy quality of the data. Therefore, the distance between the perturbed and original trajectory points is used to quantify the obfuscation quality. Privacy quality is defined as:

$$PQ = \frac{1}{n}\sum_{i=1}^{n} d(D_i, D_i') . \tag{20}$$

Where n is the number of position points in the trajectory, D_i represents the original trajectory point, and D_i' represents the data after adding Laplace noise to the original real data .

Query Error (QE): QE is a common method of usability metrics, which refers to the average relative error between the processed dataset D' and the actual results in the original dataset D, which is calculated as shown:

$$QE = \frac{|Q(D\ ) - Q(D')|}{\max\{Q(D\ ), b\}} . \tag{21}$$

Where D is the original real data; D' denotes the original real data after adding Laplace noise processing; and b is the threshold value set to prevent the denominator from zero. By selectivity of the query is meant the number of all records satisfying the query conditions as a percentage of the total number. Where max{} denotes the maximum sequence of the query. When an in-vehicle user obtains a service from an LBS provider, the actual location is replaced by a false location, which results in the location of the service request sent by the in-vehicle user to the LBS deviating from the real location, i.e., the service information provided by the LBS provider to the user is inaccurate. Therefore, the quality of service provided by LBS to in-vehicle users is determined by the similarity between the false dataset D' and the original dataset D, i.e., the loss of service quality for users is determined by the average relative error QE. If the similarity between the real dataset D and the noise dataset D' is higher, the service quality is higher; otherwise, the service quality is worse.

Root Mean Square Error (RMSE): RMSE measures the accuracy of the data by calculating the square root of the arithmetic mean of the squared differences between the true location data and the noisy location data. A smaller RMSE value indicates that the noisy location data is closer to the true location data, thus resulting in a higher quality of service. The RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} dis(D_i - D_i')^2}{n}} . \quad \textbf{(22)}$$

where n represents the number of requested location points, D_i represents the original real data, D_i' represents the data after adding Laplace noise to the original real data, and $dis(D_i - D_i')$ is the distance between D_i and D_i'.

4.3 Experimental Result

To verify the effectiveness of our method, comprehensive experiments will be compared with the traditional average allocation method, PPBA algorithm [20] and SDLDP algorithm [21] in terms of privacy protection and quality of service. The traditional average allocation method allocates the privacy budget equally to each sensitive location. The PPBA algorithm allocates the appropriate privacy budget to the user as a percentage of the sensitive distance based on the sensitive location set by the user and the acceptable error value, and does not allocate the privacy budget to the non-sensitive locations. The SDLDP algorithm calculates the privacy score of a sensitive point based on its density score, the size of the cluster it is in, and the distance between other sensitive points in the cluster, and then assigns DP budgets to the sensitive points based on the privacy scores, and does not do anything with the other trajectory points.

1) Privacy Effects Analysis

Theorem 1. Personalized location privacy protection algorithm to resist background knowledge attacks satisfies ε–*DP*.

Proof: The algorithm assigns budgets ε_i to n request location points on the selected optimal path, so $\varepsilon = \sum_{i=1}^{n} \varepsilon_i$. Let the random algorithm for each request location point $rp_i(x_i, y_i)$ be $M_i : D_i \to \mathfrak{R}$ and any two random algorithms M_i are independent of each other. The output result of algorithm M is $O = \{r_1, r_2, ..., r_n\}$. For $\forall O \in \mathfrak{R}$, according to definition 2, there are

$$\begin{aligned} \Pr[M(D) = O] &= \prod_{i=1}^{n} \Pr[M_i(D_i) = O_i] \\ &\le \prod_{i=1}^{n} (e^{\varepsilon_i \times |D_i \oplus D_i'|} \times \Pr[M_i(D_i') = O_i]) \cdot \\ &= e^{\sum_{i=1}^{n} \varepsilon_i \times |D_i \oplus D_i'|} \times \Pr[M(D') = O] \cdot \end{aligned} \quad \textbf{(23)}$$

It can be inferred from

$$\begin{aligned} & |D_i \oplus D_i'| = 1 \\ & \Pr[M(D) = O] \le e^{\sum_{i=1}^{n} \varepsilon_i} \times \Pr[M(D') = O] \\ & \qquad = e^{\varepsilon} \times \Pr[M(D') = O] \end{aligned} . \quad \textbf{(24)}$$

Therefore, personalized location privacy protection algorithms that resist background knowledge attacks satisfies ε–*DP* and can provide privacy assurance for users.

2) Comparison of real trajectory privacy protection

We extracted two real trajectories from the dataset and compared them with the trajectories generated by the four algorithms to evaluate the privacy-preserving strength of the algorithms. Fig. 3 illustrates the comparison between one real trajectory and four perturbed trajectories.

As can be seen from Fig. 3, the average distribution method deviates from the real location points regardless of whether there are sensitive location points around the current request location point, and cannot meet the user's personalized privacy needs. The PPBA algorithm and the SDLDP algorithm perturb the request location points that are close to the sensitive location points, and do not perturb the request location points that are not surrounded by the sensitive location points, and there are the location points that overlap with the real trajectory in the perturbed trajectory generated. The generated perturbed trajectory has position points that overlap with the real trajectory, and when the number of requested position points is much larger than the number of sensitive position points, the attacker can infer the user's real trajectory with high probability. The method in this paper combines the subjective sensitivity degree of the user and the objective history query probability to determine the user sensitivity degree, and perturbs the requested location points that are far away from the sensitive location points with the history query probability, which has a higher protection strength compared with other algorithms.

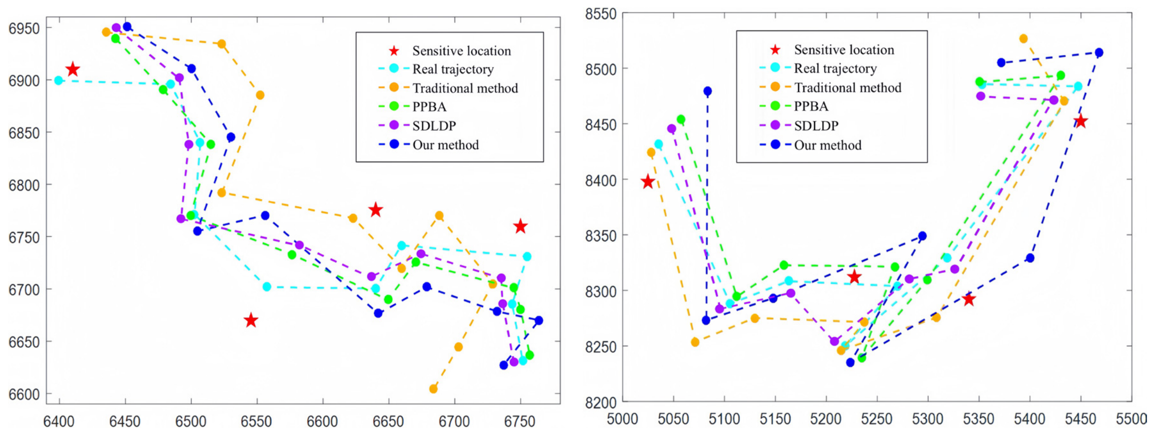


Fig. 3. Trajectory comparison

3) PQ under DP budgets

Fig. 4 shows the variation of PQ with privacy budget for different trajectory privacy-preserving methods. With the same privacy budget, the PQ of the present method is similar to the traditional average distribution privacy-preserving method. However, the present method protects the privacy of sensitive location points more effectively because it includes additional background information and adds a larger perturbation to these points. In addition, the PQ value of the present method is higher compared to the PPBA method and SDLDP method which also consider sensitive location points. This is because the present method not only considers the user's subjective sensitivity, but also calculates the sensitivity objectively by combining the historical query probability and adds perturbations to the non-sensitive location points as well. Compared with other methods, the method in this paper provides more comprehensive privacy protection.

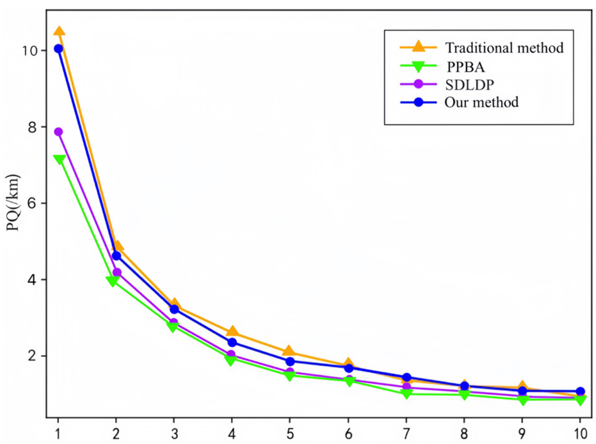


Fig. 4. PQ under DP budgets

Analysis of Service Quality

The experiments measure the impact on the quality of service under privacy protection in terms of QE and RMSE. The experiments respectively set the privacy budget a to change from 1 to 10 to take the average of 100 times to verify the effectiveness of the algorithms under DP budgets; the experiments respectively set the number of requested location points to change from 5 to 50 to take the average of 100 times to verify the impact of the different number of requested location points on the performance of location privacy protection;

1) QE under DP budgets

Fig. 5 compares the QE between real data and noise processed data under DP budgets for this paper's method, the traditional average distribution method and, the PPBA algorithm and the SDLDP algorithm. It can be seen from the figure that the query errors of all four algorithms are on a decreasing trend as the privacy budget continues to increase. This is due to the fact that the larger the privacy budget the less privacy-preserving the effect is, i.e., the smaller the noise is added, and thus the query error decreases gradually. Under the same privacy budget, the traditional average distribution method is significantly higher than the other three methods, this is because the traditional method adds the same privacy protection to all request points, which leads to over-protection and lower data availability. The query error of this paper's method is slightly higher compared with PPBA algorithm and SDLDP algorithm, this is because the proposed method not only considers the subjective sensitivity provided by the user, but also objectively calculates the sensitivity by combining the historical query probability, and privacy protects the location of the non-sensitive points as well, and the method also considers various complexities of the road network during the re-route selection, so that it ensures the service quality while it also protects the user's privacy in a more comprehensive way.

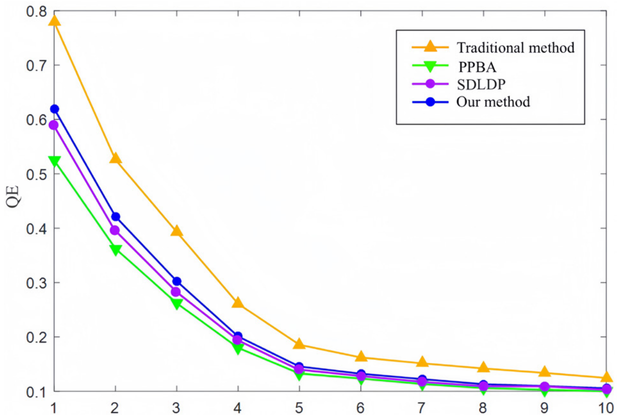


Fig. 5. QE under DP budgets

2) RMSE under DP budgets

Fig. 6 compares the RMSE between real data and noise-processed data for DP budgets for this paper's method, the traditional average distribution method and, the PPBA algorithm and the SDLDP algorithm.

As can be seen from Fig. 6, the RMSE of all four algorithms shows a decreasing trend as the privacy budget increases. With the same privacy budget, the RMSE of the traditional average allocation method is higher than the other methods. This is due to the fact that the average allocation method adds the same amount of noise to some sensitive points that are of less concern to users, resulting in excessive privacy protection. The RMSEs of the PPBA algorithm and the SDLDP algorithm are not much different from those of this paper's method, which is due to the fact that this method allocates the privacy budget based on the probability of the historical query, which allocates a certain degree of privacy protection to the location points that are not considered to be sensitive by the users' subjective opinion. In contrast, the other two methods ignore the background knowledge available to the attacker for non-sensitive location points. In the presence of multiple non-sensitive location points, the user's trajectory privacy can be easily compromised. Therefore, despite the slightly lower RMSE, this method provides stronger privacy protection than the PPBA algorithm.

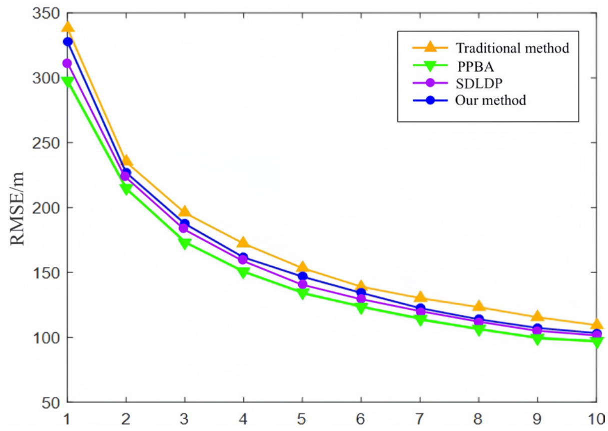


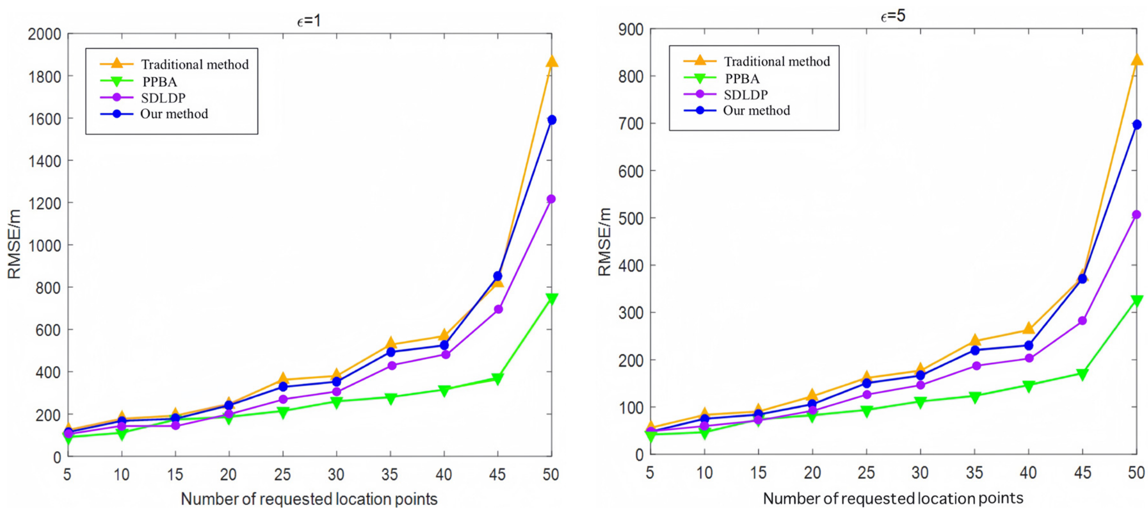
Fig. 6. RMSE under DP budgets

3) RMSE under different number of request location points

Fig. 7 compares this paper's approach, the traditional approach, the PPBA algorithm, and the SDLDP algorithm, all of which utilize differential privacy mechanisms to protect location privacy. The comparison is based on the RMSE between real data and data protected by these differential privacy mechanisms, under varying numbers of requested location points. This analysis aims to illustrate the relationship between the number of requested location points and service quality.

Fig. 7 shows an increasing trend in RMSE for all four algorithms as the number of requested location points increases under three DP budgets. This phenomenon can be attributed to the fact that under a fixed privacy budget, the increase in the number of location points leads to a corresponding decrease in the privacy budget allocated to each location point, which introduces more noise and thus increases the RMSE.

With the same privacy budget and number of location points, the traditional method has the highest RMSE. This is due to the fact that the traditional method adopts the strategy of distributing the privacy budget equally by assigning the same privacy budget to all location points, which may lead to excessive privacy protection for some location points. On the contrary, this paper's method personalizes the privacy budget allocation based on the sensitivity and distance of the location points, and thus balances privacy protection and service quality more effectively.



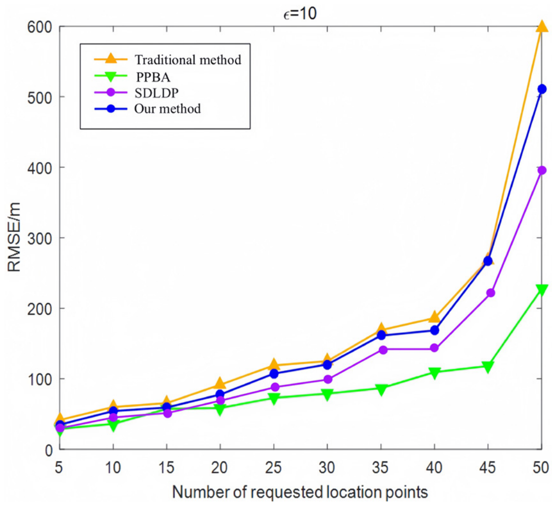


Fig. 7. RMSE under different number of request location points

The RMSE of this paper's method is slightly higher than that of the PPBA algorithm and the SDLDP algorithm, and this difference is mainly due to the fact that as the number of location points increases, the number of non-sensitive location points also increases. Since both the PPBA algorithm and the SDLDP algorithm only add noise to the sensitive location points, this results in their relatively low RMSE. However, as the number of non-sensitive location points increases, not protecting these location points increases the likelihood of an attacker inferring the user's true trajectory.

In summary, compared with the PPBA algorithm and the SDLDP algorithm, this paper's approach takes into account users' subjective privacy needs and provides users with a higher intensity of personalized privacy protection while ensuring low travel costs and high service quality.

5 Conclusion

To meet users' needs for personalized privacy protection, this paper proposes a personalized location privacy protection algorithm to defend against background knowledge attacks. Our approach first calculates the sensitivity to satisfy the user's privacy needs based on their subjective preference for sensitive locations and objective historical query probability. Subsequently, it recommends a privacy-protecting, low-cost path for the user based on the characteristics of the road network environment. Finally, it assigns a personalized privacy budget based on the sensitivity of the path and the distance to sensitive locations, and adds Laplace noise that complies with differential privacy mechanisms to each requested location point, providing personalized privacy protection. Experimental results and analysis show that the algorithm offers higher privacy protection strength and satisfies users' personalized privacy needs while ensuring service quality. However, our approach also has limitations in that it focuses on background knowledge attacks and may be less effective against more sophisticated attacks such as federated attacks or adversarial attacks. Future research could explore more comprehensive privacy-preserving mechanisms, such as methods that combine differential privacy with secure multi-party computation or homomorphic encryption, for more sophisticated attacks.

6 Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (61762058, 62366029), the Key Research and Development Program of Gansu Province (25YFFA089), the Industrial Support Project of Gansu Provincial Department of Education (2022CYZC-38), and the Gansu Provincial Natural Science Foundation Program (21JR7RA282, 23JRRA855).

References

[1] D. Liao, X. Huang, V. Anand, G. Sun, H. Yu, k-DLCA: An efficient approach for location privacy preservation in location-based services, in: Proc. 2016 IEEE International Conference on Communications (ICC), 2016. https://doi.org/10.1109/ICC.2016.7510687

[2] X. Zhang, X. Gui, Z. Wu, Privacy preservation for location-based services: a survey, Journal of Software 26(9)(2015) 2373-2395. https://doi.org/10.13328/j.cnki.jos.004857

[3] C. Dwork, Differential privacy, in: Proc. of the 33rd international conference on Automata, Languages and Programming, 2006. https://doi.org/10.1007/11787006_1

[4] Y. Li, X. Li, X. Luo, Z. Li, H. Li, X. Zhang, AnotherMe: A Location Privacy Protection System Based on Online Virtual Trajectory Generation, IEEE Transactions on Dependable and Secure Computing 21(4)(2024) 2552-2567. https://doi.org/10.1109/TDSC.2023.3314200

[5] J. Zhang, F. Yang, Z. Ma, Z. Wang, X. Liu, J. Ma, A Decentralized Location Privacy-Preserving Spatial Crowdsourcing for Internet of Vehicles, IEEE Transactions on Intelligent Transportation Systems 22(4)(2021) 2299-2313. https://doi.org/10.1109/TITS.2020.3010288

[6] J. Wang, C. Wang, J. Ma, H. Li, Dummy location selection algorithm based on location semantics and query probability, Journal on Communications 41(3)(2020) 53-61. https://doi.org/10.11959/j.issn.1000-436x.2020061

[7] X. Zhang, H. Yang, Z. Li, F. He, J. Gai, J. Bao, Differentially Private Location Privacy-preserving Scheme with Semantic Location, Computer Science 48(8)(2021) 300-308. https://doi.org/10.11896/jsjkx.200900198

[8] X. Sun, Q. Ye, H. Hu, J. Duan, Q. Xue, T. Wo, Generating Location Traces With Semantic-Constrained Local Differential Privacy, IEEE Transactions on Information Forensics and Security 19(2024) 9850-9865. https://doi.org/10.1109/TIFS.2024.3480712

[9] C. Chen, X. Hu, Y. Li, Q. Tang, Optimization of Privacy Budget Allocation In Differential Privacy-Based Public Transit Trajectory Data Publishing for Smart Mobility Applications, IEEE Transactions on Intelligent Transportation Systems 24(12)(2023) 15158-15168. https://doi.org/10.1109/TITS.2023.3309783

[10] H. Li, X. Ren, J. Wang, J. Ma, Continuous location privacy protection mechanism based on differential privacy, Journal on Communications 42(8)(2021) 164-175. https://doi.org/10.11959/j.issn.1000-436x.2021123

[11] C. Xu, Y. Ding, C. Chen, Y. Ding, W. Zhou, S. Wen, Personalized Location Privacy Protection for Location-Based Services in Vehicular Networks, IEEE Transactions on Intelligent Transportation Systems 24(1)(2023) 1163-1177. https://doi.org/10.1109/TITS.2022.3182019

[12] C. Wang, Y. Zhao, A Novel Approach for Trajectory Partition Privacy in Location-Based Services, in: Proc. 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2023. https://doi.org/10.1109/TrustCom60117.2023.00200

[13] H. Yuan, L. Wu, L. Xu, L. Ban, H. Wang, Y. Su, W. Meng, SCTP: Achieving Semantic Correlation Trajectory Privacy-Preserving With Differential Privacy, IEEE Transactions on Vehicular Technology 74(4)(2025) 5856-5870. https://doi.org/10.1109/TVT.2024.3505200

[14] L. Wu, C. Qin, Z. Xu, Y. Guan, R. Lu, TCPP: Achieving Privacy-Preserving Trajectory Correlation With Differential Privacy, IEEE Transactions on Information Forensics and Security 18(2023) 4006-4020. https://doi.org/10.1109/TIFS.2023.3290486

[15] C. Zhou, Y. Chen, H. Tian, S. Cai, Location Privacy and Query Privacy Preserving Method for K-nearest Neighbor Query in Road Networks, Journal of Software 31(2)(2020) 471-492. https://doi.org/10.13328/j.cnki.jos.005679

[16] S. Li, X. Ji, W. You, A Personalized Differential Privacy Protection Method for Repeated Queries, in: Proc. 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), 2019. https://doi.org/10.1109/ICBDA.2019.8713224

[17] X. Jiao, X. Li, Processing Parameters Optimization of the Automobile Panels Based on Fuzzy Analysis, Machinery Design & Manufacture 384(2)(2023) 112-116. https://doi.org/10.3969/j.issn.1001-3997.2023.02.022

[18] J. Lv, Study on Evaluation and Judgment Method of Urban Road Traffic Congestion, Construction & Design for Engineering 10(9)(2022) 74-76.

[19] J. Yao, C. Shao, P. Wang, Y. Wei, Research on Route Selection Model of Emergency Vehicles in Urban Road Network, Logistics Sci-Tech 43(4)(2022) 83-99. https://doi.org/10.13714/j.cnki.1002-3100.2020.04.022

[20] C. Xu, Y. Ding, L. Luo, S. Liu, L. Liu, G. Zhao, Personalized Location Privacy Protection for Location-based Services in Vehicular Networks, Journal of Software 33(2)(2022) 699-716.
https://doi.org/10.13328/j.cnki.jos.006157
[21] W. Zhi, X. Gong, Y. Wang, Personalized Differential Privacy Preservation Method for Trajectory Based on Regional Density Analysis, in: Proc. 2023 2nd International Conference on Big Data, Information and Computer Network (BDICN), 2023.
https://doi.org/10.1109/BDICN58493.2023.00016