A Semantic-Based Dual Location Privacy-Preserving Approach

Xudong YANG†, Member, Ling GAO†, Yan LI†, Jipeng XU†, Jie ZHENG†, Hai WANG†, and Quanli GAO†, Nonmembers

SUMMARY With the popularity and development of Location-Based Services (LBS), location privacy-preservation has become a hot research topic in recent years, especially research on k-anonymity. Although previous studies have done a lot of work on anonymity-based privacy protection, there are still several challenges far from being perfectly solved, such as the negative impact on the security of anonymity by the semantic information, which from anonymous locations and query content. To address these semantic challenges, we propose a dual privacy preservation scheme based on the architecture of multi-anonymizers in this paper. Different from existing approaches, our method enhanced location privacy by integrating location anonymity and the encrypted query. First, the query encryption method that combines improved shamir mechanism and multi-anonymizers is proposed to enhance query safety. Second, we design an anonymity method that enhances semantic location privacy through anonymous locations that satisfy personal semantic diversity and replace sensitive semantic locations. Finally, the experiment on the real dataset shows that our algorithms provide much better privacy and use than previous solutions.

key words: location privacy, semantic attack, multi-anonymizers

1. Introduction

With the advent of the 5G era, terminal-based Applications and services have been significantly developed [1]. The extensive application and development of terminal service also make the most comprehensive application, such as Location-Based Service (LBS) has attracted people’s attention. With the location-based service, users obtain the current location through the mobile terminal and send it to the LSP and the service request. Through the above process, the LBS can provide the service that fulfills the user’s request. Location-based service improves the quality and efficiency of service from the aspects of diet (Meituan App), travel (Didi App), and entertainment (Dazhong App), because of its feature of providing nearby service.

However, while enjoying the convenience brought by the LBS, users also leave a lot of move tracks data on the service platforms. By analyzing these data, including user’s locations stored in the service platform, the attacker can obtain the user’s privacy information by combining relevant background knowledge of his has [2]. For example, the attacker can infer the user’s state of health by combining his query near the location of the hospital. Thus, we need to pay more attention to location privacy in LBS.

To reduce the risk of privacy disclosure in LBS, many studies have been proposed and achieve the goal of privacy security, including:

(1) Obfuscation [3], [4]: By the location obfuscation technology, the actual request location is sent to the LSP to enjoy service after the operation of adding noise or another location substitution. Although this method has the advantage of being simple to implement, it has poor availability due to changing the original request location.

(2) Offset [5], [6]: To overcome the problem of low usability the obfuscation has, the offset method chooses an identified location around the request location to replace the user’s real location to be sent to LSP and enjoy LBS.

(3) Dummy [7], [8]: Dummy locations generation scheme is a method where the service provider generates dummy locations in addition to the user’s real location to provide a more efficient privacy service than offset.

Although the above method alleviates the risk of location privacy to some extent, it leads to the problem of low efficiency and poor availability.

To overcome the above problems, the relevant research on k-anonymity has become a hot topic in recent years [9]–[18], [28]. The user enjoys the LBS service by sending an anonymity set that includes k-1 anonymous locations adjacent to the requested location by using the k-anonymity method. At present, most of the privacy protection studies based on k-anonymity mainly focus on the following aspects: Research on improving the privacy degree of the anonymous location [9], [10] and satisfying the personalized privacy preference [11], [12], research on resisting the security threat in the process of anonymity set generation [13], and improve the efficiency of anonymity [14], [15]. In addition, with the diversification of attack methods in recent years, location privacy preservation to defend the semantic attack has also received academic attention [16]–[18]. However, these previous works research on privacy mainly focus on geographic semantic privacy, ignoring the impact of other semantic information, including queries and sensitive semantic locations on location privacy security, including the following cases:

(1) Compromise the security of request location by the query content leak: Considering the user’s movement pattern and user habits, there is a specific correlation between the user’s query content and the requested location. For example, the user will query the nearby food at home and the
to its good balance between availability and privacy.

2. Related Work

Although the obfuscation, offset, and dummy methods have been proposed to protect location privacy, the k-anonymity technique has become the mainstream in current studies due to its good balance between availability and privacy.

The Literature [9] was first proposed the k-anonymity method for location privacy, and achieved by the real location was generalized by adjacent k-1 locations selected. Although this method can realize the basic idea of k-anonymity, it fails to consider the threat of background attacks on privacy. To resist the background attack, the anonymous location selection based on the background knowledge about the location history request probability and similarity was proposed in [10]. However, it lacks consideration for the personalized privacy requirements of different users. The [12] proposed a location privacy-preserving scheme through the user-defined grid in LBSs to satisfy the customized needs of users. But, it doesn’t consider the security during anonymity collection generation. The authors of [11] designed a framework for cooperative privacy preservation in LBS, which is based on the game model between users and cooperative users and blockchain technology to ensure communication security between users in building anonymity sets. The above works can provide security for location information in LBS, but they ignore the effects of the anonymity generation.

To further improve the efficiency of anonymity, Niu et al. completed the anonymity set pre-cached that users might need in the future by introducing the local caching mechanism [13]. Therefore, the time cost of anonymous requests was cut back and improved the efficiency of anonymity. However, a problem of a low cache hit ratio still exists with this work. Zhang et al. adopted a cache prediction model based on the user’s historical movement trajectory to improve the matching degree between user requirements and cached location. They cached them for offering help for anonymity requests the next time [14]. To make full use of the computing resources of terminals owned by each user, Literature [15] proposed a user collaboration privacy protection method for LBS. This method does not need to change the LBS server architecture and meet the personal privacy demand through the collaboration among users and sharing of the personal anonymity set. However, all these studies have a significant disadvantage, that is, it is difficult to achieve a good balance between efficiency and safety, especially for the defense of semantic attacks.

To solve the problem of privacy protection under the semantic attack, the researcher proposed a privacy protection method based on location semantics in the environment constrains of road networks [16]. This study established a privacy evaluation model based on location semantics based on location popularity, sensitivity, and privacy degree. They completed the construction of the anonymity set by the anonymous road selection based on the privacy evaluation results. Literature [17] designed a dummy location generation mechanism based on location semantic information and met the semantic k-anonymity. Under the k-anonymity constraint of location semantic diversity, uniform distribution of geographical locations was guaranteed under the condition that the probabilities of anonymous location queries were similar. The authors of [18] proposed a location semantic model based on map semantic information and adopted this model to choose the anonymous location.
3. Motivation and Architecture

3.1 Motivation

The existing anonymity-based method may fail in a new situation where the adversary obtains semantic knowledge by analyzing the query content and sensitive semantic location. When the adversary launches the attack at only one anonymous server, he can steal the user queries. In this case, even if the k-anonymity method is used to protect the real request location, the user’s privacy will be leaked due to query content being related to the user’s semantic of request locations, and this relationship is depicted in Fig. 1 (a). The attacker can infer the user’s request location semantics by analyzing the relationship between the query content and the request location semantics. As shown in Fig. 2 (a), when the adversary known Alice’s query content about where is good food nearby, he can infer Alice’s maybe locate at semantic of home, school, and company which is closely related to daily life, rather than places that have little to do with the probability of inquiry such as cinemas, restaurants, and bars. Thereby, his 5-anonymity is failed because the anonymous location includes cinemas, restaurants can be filtered out.

As previously researchers discovered that, sensitive semantic locations could identify the access patterns of different users since sensitive semantic locations are related to the user’s role and preferences [22]. As shown in Fig. 1 (b), it depicts that different users have significantly different access frequencies to user-sensitive semantic locations in the public dataset. When an attacker has this knowledge, it can infer the anonymous locations that will most likely not be accessed based on the user’s access habits, thus compromising anonymity security. For example, Fig. 2 (b) shows that the anonymity set includes Alice’s sensitive semantic location bookstore. The attacker can still filter out the school and snack bar since they are not in the semantic type related to the sensitive semantic location. This failure security is due to the exposed personal visit pattern as Alice’s sensitive locations are identified.

To address the above issues, both query and location privacy need to be emphasized in the anonymization process. To this end, we propose a dual semantic privacy protection mechanism (DSPPM) based on k-anonymity and multiple anonymization servers. The query information is prevented from being obtained by attackers through a query encryption scheme based on the Shamir mechanism. Moreover, by constructing anonymity sets that satisfy the semantic diversity of individuals, the replacement of sensitive semantic locations is accomplished, and thus the security of anonymous sets is improved.

3.2 Architecture

As Fig. 3 shows, our privacy-aware location framework’s architecture includes three critical components: users, the anonymous server (AS), and the location-based service provider (LSP).

Users: as LBS service initiator, the user sends his current location and service query request to the anonymous server (AS) through RAN; Before sending the query, it is necessary to encrypt the query based on the Shamir sharing mechanism, and then combine the request locations and send it to the anonymous server.

Multi-Anonymizers (MA): There are $n_A$ anonymous server in our privacy framework, called Multi-Anonymizers, and two main parts in each anonymous server, including the computing module for the Lagrange Factor generation and location anonymity generation function, and the storage module for storing the service response. For one anonymous server in our architecture, when it receives the user’s request by RAN transmission, the computing module generates the corresponding Lagrange Factors according to the Lagrange generation method of the Shamir mechanism. It generates the anonymity set combined with the anonymity set generation method proposed in this paper. After completing the above operations, the anonymity set and the Lagrange factors will be sent to the LSP.

Location Server Provide (LSP): The LSP can provide various services for users based on their query, such as...
location-based navigation. When the LSP receives the sub-queries satisfying the quantity constraint and the anonymous set, the LSP restores the original query request according to the recovery operation in the Shamir process and then returns the corresponding service result to users.

4. A Dual Semantic Privacy Protection Mechanism

4.1 Overview

As shown in Fig. 4, DSPPM first performs query encryption. The improved Shamir scheme is adopted to query partitioning based on the number of anonymous servers and encrypt sub-query (EQ). The encrypted sub-query and request location (RL) is then sent to corresponding servers in Multi-Anonymizers. Then, UPDDM achieves anonymity based on the RL sent by the user. The anonymous server is randomly selected from multi-anonymizers (MA) and generates the anonymity set (AS), which considers both personal semantic privacy and sensitive semantic privacy-related user’s role. In the final step, decryption, LSP obtains decrypts query (DQ) after receiving sub-query by request Lagrange factors (LA) generated by each sub-secret EQ. Then it generates results of the service (SRs) according to the decryption query (DQ) and the anonymity set (AS) and returns it to the user.

4.2 Query Encryption and Decryption

As a secret sharing mechanism [19], [20], the Shamir mechanism has been widely used in daily life. the Shamir secret sharing algorithm is a multi-user secret sharing scheme based on Lagrange interpolation, which is traditionally called (t, n)-threshold scheme. In this scheme, secrets are divided into n sub-secrets for secure sharing, so as to resist privacy disclosure in the centralized sharing of secret information that can be obtained by attacking only one target.

In this section, we introduce the implementation method of providing users with secure query requests based on the Shamir sharing mechanism. Details of its implementation include the following query encryption and decryption steps.

**Query Encryption:** First, To divide the user’s original query Q into n-pieces which from the number of nA anonymous servers, t-1 (t=nA) elements ai (i=1,2,3...t-1), x' are randomly selected in the finite field GF(P) and first construct the following (t-1)-order polynomial:

\[ F(x) = a_0 + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \]  

Then, introduce the query Q as secret S into Eq. (1) as the first term. Then the Eq. (1) containing the secret S can be expressed

\[ F(x) = (S + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1}) \mod p \]  

With the Eq. (2) and the large prime p, we can generate nA encrypted sub-query, called sub-secrets \( EQ_j (1 < j < n) \) as

\[ EQ_j = F(x_j) = (S + a_{j} x_j^1 + a_2 x_j^2 + \cdots + a_{t-1} x_j^{t-1}) \mod p \]  

Although the above encrypted query can prevent query information from being stolen by a single anonymous server, it cannot effectively provide security with the untrusted anonymous server in multi-anonymous servers. By masquerading as a server of multiple-anonymizer, the attacker steals t pieces of the encrypted sub-query and then...
random choose number of \( k \) each polynomial can be expressed as following two aspects: 1) Through the process of generating method proposed in the Sharmin sharing mechanism, the specific process is as follows:

Different from the classical Shamir theory, the user first random choose number of \( k \) \((k > n_A - 1)\) non-zero coefficients \( a_{im}; (1 \leq i \leq k', 0 \leq m \leq t - 1)\) in finite field GF(P), and construct \( k \) polynomials of order \((t - 1)\) accordingly, each polynomial can be expressed as

\[
f_i(x) = \sum_{m=0}^{t-1} a_{im} = a_0 + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \tag{4}
\]

Base on this polynomials, the generated \( n_A \) sub-secrets \( E_{Qj} \) can be computed as

\[
E_{Qj} = f_i(x) = \left( \sum_{m=1}^{t-1} a_{1m} x^m, \sum_{m=1}^{t-1} a_{2m} x^m, \cdots, \sum_{m=1}^{t-1} a_{km} x^m \right)
\]

For the sub-query \( E_{Qj} \), any integer \( d, w \) can be found in a finite field, and the random \( S \) satisfies that \( S \) is equal to a linear combination of all crypto subqueries \( E_{Qj} \) as

\[
S = \sum_{i=1}^{k'} d_i f_i(w^m) = (d_1 f_1(w^m) + d_2 f_2(w^m) + \cdots + d_k f_k(w^m)) \tag{5}
\]

After encryption operation above, the recovery of the original query requires not only relying on sub-secrets \( E_{Qj} \) but also Lagrange factors \( LA_j \). The Lagrange factors and parameter \( d, w \) permission is granted to all LSP servers only. At this point, the process of query encryption is described in algorithm 1.

---

**Algorithm 1 Query Encryption**

**Input:**
- The original query, \( Q \);
- The parameter, \( k', t, r \);
- The encrypted subquery, \( E_Q \);

**Output:**
- The parameter set, \( w, d \);
- The encrypted subquery, \( E_{Qj} \);

1. Initialize the \( w = [] \); \( d = {} \);
2. Initialize the \( S \)
3. Randomly choose the number of \( t - 1 \) non-zero constants \( x \);
4. Randomly choose the number of \( k' \) non-zero constants \( a \)
5. \( \text{while each } i \text{ in } k' \text{ do} \)
6. \( \text{the encrypted subquery } f_i(x^m) \) obtained based on Equation (5);
7. \( E_{Qj} = f_i(x^m) \)
8. \( \text{end while} \)
9. \( w, d = \text{Findparameter}(f_i(x^m)) \)
10. \( \text{return } E_Q, w, d \);

**Query Decryption:** Unlike the traditional decryption method proposed in the Sharmin sharing mechanism, the improved sharing mechanism is mainly reflected in the following two aspects: 1) Through the process of generating Lagrange factors, it verifies the validity of the sub-secrets owned by different servers and solves the attack of the attacker disguising the secret; 2) At the same time, the secret decryption stage is limited to the communication between the anonymous server and the LSP, which avoids the Lagrange factor being obtained by the attacker for secret recovery. The specific process of the decryption method is as follows and shown in Algorithm 2:

1) The LSP first send the decrypt request to MA after received encrypted query \( EQ(f_k(x^m))=E_Q \).
2) Then, the AS received request can generate its own unique Lagrange factors according to Eq. (7) and return its own LF to LSP:

\[
LF_i = \sum_{j=1}^{k} d_j f_k(x^m) \prod_{t=1}^{t-1} \frac{w_j - w_i}{x_j - x_i} \tag{7}
\]

3) When LSP receives the number of encrypted LF satisfying the quantitative constraint \( t \), it performs the original secret recovery by combining the LF of the response with Eq. (8). After the query decryption is complete, the service result is generated and returned to the user in combination with the anonymous location set.

\[
EQ = S = \sum_{i=1}^{t-1} LF_i \tag{8}
\]

**Algorithm 2 Query Decryption**

**Input:**
- The encrypted subqueries, \( EQ \);
- The minimum number of recovery parameter, \( min \);
- The encrypted parameters, \( w, b, r \);

**Output:**
- the original query, \( S \);

1. Compute the received the number \(|EQ| \) of encrypted subqueries \( EQ; \)
2. if \( t \leqslant |EQ| \)
3. \( \text{while each } i \text{ in } |EQ| \text{ do} \)
4. \( \text{LSP send a decryption request for } E_{Qj} \)
5. \( \text{end while} \)
6. \( \text{The anonymous server compute the Lagrange factor } LF_i \text{ of the crypto subquery } E_{Qj} \text{ their owned by equation(7)} \)
7. \( \text{while each } i \text{ in num do} \)
8. \( \text{receive the Lagrange factor } LF \text{ for } E_{Qj} \)
9. \( \text{end while} \)
10. with \( w, h \) LSP recovery of the original query \( S \) by equation(8)
11. \( \text{end if} \)
12. \( \text{return } S \);

**Theorem 1:** The decrypted query is equal to the original query.

To verify the correctness of the encryption method, we will prove it in this subsection. First, according to the above encryption method, the subquery of encrypted query can be expressed as

\[
f_i(x_r) = \left( \sum_{m=0}^{t-1} a_{1m} x_r^m, \sum_{m=0}^{t-1} a_{2m} x_r^m, \cdots, \sum_{m=0}^{t-1} a_{km} x_r^m \right) \tag{9}
\]

Then, LSP combines Lagrange factor in formula (7) with numbers of \( k \) encrypted subquery to decrypting, and the decrypted query can be expressed as

\[
f_i(x_r) = \left( \sum_{m=0}^{t-1} a_{1m} x^m, \sum_{m=0}^{t-1} a_{2m} x^m, \cdots, \sum_{m=0}^{t-1} a_{km} x^m \right)
\]
\[ s = \sum_{r=1}^{t} L F_{r}^{i} = \sum_{r=1}^{t} \sum_{l=1}^{t} d_{l} f_{l}(x_{l}) \prod_{i=1, i \neq r}^{t} \frac{u_{i} - x_{e}}{x_{r} - x_{e}} \]
\[ = \sum_{r=1}^{t} \left( d_{1} f_{1}(x_{l}) \prod_{i=1, i \neq r}^{t} \frac{u_{i} - x_{e}}{x_{r} - x_{e}} + d_{2} f_{2}(x_{l}) \prod_{i=1, i \neq r}^{t} \frac{u_{i} - x_{e}}{x_{r} - x_{e}} + \ldots + d_{1} f_{1}(x_{l}) \prod_{i=1, i \neq r}^{t} \frac{u_{i} - x_{e}}{x_{r} - x_{e}} \right) \]

(10)

Meanwhile, according to the Lagrange interpolation principle, \( f_{l}(x) \) can be expressed as
\[ f_{l}(x) = \sum_{r=1}^{t} f_{l}(x_{l}) \prod_{i=1, i \neq r}^{t} \frac{x - x_{e}}{x_{r} - x_{e}} \]

(11)

After substituting formula (12) into Formula (6), the decrypted query can be expressed as
\[ S = \sum_{l=1}^{t} d_{l} f_{l}(w^{m}) = (d_{1} f_{1}(w^{m}) + d_{2} f_{2}(w^{m}) + \ldots + d_{c} f_{c}(w^{m})) \]

(12)

Based on the above analysis, the encrypted query (Eq. (12)) is equal to the original query (Eq. (6)), which shows the correctness of our proposed query encryption method.

4.3 Anonymity

This section first extracted the user’s location semantic features based on their mobile trace. Then the personal semantic hierarchy tree is then constructed based on these semantic features, thus getting the location semantic fine-grained classification. Finally, the anonymity set is created based on anonymous locations that satisfy the semantic diversity and semantic sensitive location replacement in this semantic tree.

4.3.1 Extract Personal Semantic Feature

In the user’s movement, their moving track can be express as an trajectory \( T \) and expressed as
\[ T = l_{0} \rightarrow l_{1} \rightarrow \ldots \rightarrow l_{n-1} \rightarrow l_{n} \]

(13)

Where, \( l_{i} \) in the trajectory represents the location node, and each location node contains \( \text{lng}, \text{lat}, \text{t}. \) \( \text{lng} \) represents the longitude information of the user’s sampling location, \( \text{lat} \) represents the latitude information of the user’s sampling location, and \( \text{t} \) represents the sampling time of the user sending a query.

To avoid the anonymous location and the user’s real behavior does not match the occurrence of the situation, the semantic feature should be related user’s mobile. A location semantic model is proposed based on long stay location, which is called stay-point, to obtain location semantic information related to users’ mobile behavior. The stay point \( s \) is a geographic region where a user stayed over a time threshold \( \theta \), within a distance threshold \( \theta_{d} \). Thus, the stay-point can be get when the location \( l_{i}, l_{j} \) in \( T \) satisfies time constraint \( \text{len}(l_{i}, l_{j}) > \theta \) and distance constraint \( \text{dis}(l_{i}, l_{j}) > \theta_{d} \) at same time. At this point, the center of the stop area is the stay point \( S \).

Then, the basic semantic feature (BSF) of \( S \) can be described by geographic semantic information such as hospitals and schools belonging to the point of interest within the same stay area. Inspired by the TF-IDF technology[21], we take the original semantic type of the location as the word, the stay area as the document, and the mobile area of all users as the thesaurus, then the weight of the PoI of type \( i \) can be expressed as
\[ s_{fi} = \frac{m_{i}^{n}}{M_{r}} \times \frac{M_{p}}{m_{ip}} \]

(14)

Where, \( m_{i}^{n} \) and \( m_{ip} \) represent the number of type \( i \) request locations appearing in the stay area and the user’s trajectory data set, respectively. \( M_{p}, M_{r} \) represent the total number of request locations similar to \( M_{p} \) and similar locations, respectively. In Eq. (12), the first part represents the frequency of category occurrence, while the second part represents the reverse frequency of the category. Then, The semantic feature vector of the stay point is obtained by semantic type weights for all locations in the stay area \( \text{BSF} = \{s_{f1}, s_{f2} \ldots s_{fn}\} \).

The residence time in different semantic type locations is also an essential factor in the semantic description because the different stay times in the same semantic type location can reflect the user’s different semantic purpose of mobile user behavior. For example, the gas station’s location where the longer stay time maybe for refueling, while the shorter stay time maybe for going to the bathroom. Thus, the temporal semantic features (TSF) are critical factors for semantic description and can be computed by the result of the user’s departure time subtract arrive time as follow:
\[ \text{TSF}_{i} = t_{\text{leave}} - t_{\text{arrive}} \]

(15)

Where, \( t_{\text{arrive}} \) represents the time when the user arrives at position \( i \), and \( t_{\text{leave}} \) represents the time when the user leaves position \( i \).

4.3.2 Construct Personal Semantic Tree

Considering the semantic feature, which includes BSF and TSF, represented as \( F = \{\text{BSF}, \text{TSF}\} \), the hierarchical semantic tree can be constructed to describe the semantic information of each location, which is represented as \( G = (V, E, F) \). Where \( V \) represents a tree node information, the node’s value is equal to the different users of different granularity relevant mobile location, such as stay-location information (internal node), real request location information (leaf node), etc. These nodes can be divided into leaf nodes and internal nodes. The user’s request location represents by the leaf nodes, and leaf nodes belonging to the same ancestor.
reflect these locations belong to the same semantic type, the internal node can represent the stay-point, aggregation point position of the semantic description of coarse-grained said. E represents a side of the tree structure, each edge represents the relationship between the two adjacent nodes.

The basic idea of the semantic hierarchical tree construction is first to build the basic layer and then extend it up and down, respectively, as shown in Fig. 5. Then the semantic tree can be constructed by the following step:

1) We first cluster the stay-point based on its BSF and the stay-point nearest the cluster center can be taken as the node of the basic layer.

2) Then, we perform the operation of upwards from the basic layer iteratively based on BSF of stay-point. The node of this tree can be obtained by the stay-area merge method in [22]. Also, this iterative process ends when the root node is obtained and upwards process as shown in algorithm 3.

3) Finally, we operated downwards from the basic layer iteratively, according to the TSF of locations. The EM-cluster method [23] is introduced and adopted for this cluster. The downwards process is shown in algorithm 4.

4.3.3 Anonymous Set Construction

To prevent privacy leakage caused by sensitive semantic location, this paper proposes an anonymity set construction method that must meet the following constraints: 1) All anonymous locations are not associated with the user visit. That is, they are non-sensitive semantic locations to users. 2) the constraint of semantic diversity among anonymous locations.

To satisfy constraint 1, we design a sensitive semantic location replacement method. We find that a more personal pattern of the user’s visits to a particular semantic category implies that the semantic category is more sensitive to the user. According to this finding, the frequency of users’ access to the semantic location of different categories can represent the sensitivity of location. For one user, when a semantic location appears more frequently in the user’s access track and less frequently in the other user’s access track, it is sensitive to this user.

To describe this sensitivity of location, the TF-IDF ap-

Algorithm 3 Upward

Input:
The Stay-Region set, $SR$;
The Basic Semantic Feature set, $SC$;
The Similarity threshold, $\lambda_{sim}$;

Output:
The clustered stay-point set, $S*$;

1: Initialize the $S*, SR^*$;
2: while each $i$ in range($SC$) do
3: if $S* == null$ then
4: $j = i + 1, k = 1$;
5: $SR^{k}_{t} = SR_{t}$;
6: end if
7: $sim_{ij} = getSim(SR^{k}_{t}, SC)$; // similarity computing by [25]
8: if $sim_{ij} < \lambda_{sim}$ then
9: $SR^{k+1}_{t} = SR_{t}$;
10: end if
11: if $sim_{ij} > \lambda_{sim}$ then
12: $S^{k} = getS(SR^{k}_{t})$;
13: $k = k + 1$;
14: $SR^{k}_{t} = SR_{t}$;
15: $j = j + 1$;
16: end if
17: end while
18: return $S*$;

Algorithm 4 Downward

Input:
The location set, $L$;
The Temporal Semantic Feature set, $TSF$;
Number of Iterations, $N$;
Convergence parameters, $\lambda$;

Output:
The clustered location set, $C = \{C_i\}, (i = 1, 2, 3...n)$;

1: Initialize the cluster center $c = c_j$;
2: while each $i$ in $N$ do
3: E Step:
4: Compute the $d_{ij} = ||TSF_{i} - C_j||^2$ and choose the locations $L_i \in L$ to join the Cluster $C_j$ based on $d_{ij}$;
5: Compute the likelihood function $Q_n(z^n) := P(z^n | TSF_{i}, c^n) P(c^n)$ of the clustering result in each iteration
6: M Step:
7: Update the cluster center based on $c_{n+1} = \sum_{TSF \in C_i} TSF_{i} / \sum c_{i}$
8: $c_{n+1} := arg\max \sum c_{i} Q(t | z^n) \log \frac{P(z^n | TSF_{i})}{P(c^n)}$
9: end while
10: if $c_{n+1} - c_n < \lambda$ then
11: break;
12: end if
13: return $C$;

4 plies in this case. According to the TF-IDF, we first take the semantic type of the current user’s location as the word, the semantic type of all the user’s trajectory data as the document. The word frequency can be calculated by proportioning the semantic type of the current location to the frequency of the user’s trajectory. The frequency of the reverse file is calculated according to the frequency of this semantic type in all other user trajectories, and finally, the semantic sensitivity of the user’s current location is evaluated by multiplying the two

$$ss_{i} = \frac{n_{w}}{N_{a}} \times \log \frac{N}{n_{i}} \tag{16}$$
shown in Fig. 6 (The red sensitive semantic node characterized by semantic distance not include sensitive locations related to users. The weights of all positions the semantic obtained, which does not include sensitive locations related to users. Moreover, these selected anonymous locations should be near the requested location. Therefore, we perform security analysis in the case of both of these attacks occurring in this section.

**Algorithm 5 Construct Anonymity Set**

*Input:*  
The Personal Request Location, RL;  
The Personal semantic tree, Tree;  
The threshold, $\theta_r, \theta_s, k$;  

*Output:*  
The Anonymous Set, AS;  

1. Initialize ACS, AS, k, count;  
2. **for** Each Node $\in$ Tree.LeafNode **do**  
   3. Compute ss based on Equation(14) of Node;  
   4. **end for**  
   5. Find the sensitive node with the ss by outlier detection method in [24];  
   6. Choose the replacement node by equation(14) and then complete replacement;  

**Construct the anonymity candidate set**  

1. **for** Each Node $\in$ Tree.LeafNode **do**  
   2. Choose Node from different clusters;  
   3. ACS=ACS $\cup$ Node;  
   4. **for** ACS $\in$ ACS **do**  
       5. ACS=ACS $\cup$ Node;  
       6. Remove the Node from the Tree;  
   7. **end for**  
   8. Check ACS satisfies semantic diversity constrain and distance constraints;  
   9. Choose ACS that meet distance between it and anonymous Locations more than 6;  
   10. **if** Dis($\text{CS, RL}$)$<\theta_r$ and Dis($\text{CS, RL}$)$>\theta_s$ **then**  
       11. AS=AS $\cup$ CS;  
       12. Compute Dis($\text{CS, RL}$) $\times$ $\text{Sim(\text{CS, RL})}$;  
       13. **end if**  
   14. **if** $m == k\text{-}1$ **then**  
       15. Break;  
   16. **end if**  
   17. **end for**  

5. **Construct the anonymity set**  

18. **for** Each CS $\in$ ACS **do**  
   19. **if** $\text{Dis(\text{CS, RL})}<\theta_r$ and Dis($\text{CS, RL}$)$>\theta_s$ **then**  
       20. AS=AS $\cup$ CS;  
       21. Remove the CS from the ACS;  
       22. $\text{count} = \text{count} + 1$;  
   23. **end if**  
   24. **if** $m == k\text{-}1$ **then**  
       25. Break;  
   26. **end if**  
   27. **end for**  
   28. return AS;

---

Where $n_i$ represents the number of the type i semantic that appears in the personal movement, and $n_i$ represents the total number of the i-th semantic class that appears in all request locations. $N_a$ is the number of types i locations in the individual track, and $N$ is the total number of all request locations. In Eq. (14), the first half represents the occurrence frequency of the i-th category in the user’s movement, and the second half represents the reciprocal of the frequency of this category in all POIs. The height $h$ of the upward backtracking) and shown in Fig. 6 (The red sensitive semantic node $ssi$ is replaced by the node indicated by the arrow).

$$h = \left(1 - \frac{ss_i}{SS}\right) O_t + \frac{ss_i}{SS} O_r$$  

(17)

where $O_t, O_r$ represents the replacement range specified by the user, and $SS$ represents the sum of the sensitive weights of all positions the semantic obtained, which does not include sensitive locations related to users. To satisfy constraint 2, the anonymous locations are characterized by semantic differences between the request location and each other. Moreover, these selected anonymous locations should be near the requested location.

Considering the above constraints, the construction of an anonymity set consists of three steps and is shown in algorithm 5:

1) Sensitive Node Replacement: we took the personal sensitive location as an outlier and replaced it by outlier detection in [24], as shown in line 6 to line 10;  
2) Find Anonymity Candidate Set: To ensure that semantic diversity is met and ACS is added by traversing and selecting leaf nodes in different clusters, as shown in line 11 to line 19;  
3) Construct Anonymity Set: The anonymity set is constructed by choosing a location that meets the semantic difference and near the RL, as shown in line 20 to line 27;
constraint and providing the corresponding Lagrangian Factors. Therefore, we analyze the security from these two stages.

After executing an attack against anonymous servers, the attacker can take possession of \( n - 1 \) of the anonymous servers and obtain the encrypted queries therein. In this process, the attacker can recover directly based on the obtained query information and steal the Lagrangian Factor for recovery.

When the attacker possesses \( n - 1 \) anonymous servers, he can use EQ’s encrypted query information to generate the following \( k't \)-order equations for query reconstruction.

\[
\begin{align*}
    f_1(x) &= \sum_{m=0}^{t-1} a_{1m}x^m = a_{10} + a_{11}x + \cdots + a_{1(t-1)}x^{t-1} \\
    f_2(x) &= \sum_{m=0}^{t-1} a_{2m}x^m = a_{20} + a_{21}x + \cdots + a_{2(t-1)}x^{t-1} \\
    \vdots \\
    f_k(x) &= \sum_{m=0}^{t-1} a_{km}x^m = a_{k0} + a_{k1}x + \cdots + a_{k(t-1)}x^{t-1} \\
\end{align*}
\]

(18)

The above set of equations includes \( k't \) polynomial coefficients. Since \( k't > n - 1 \) and it doesn’t match the condition of the linear equation system solution, the above set of equations does not get a unique solution. Thus, the original query \( Q \) is not be recovered.

Further, the attacker obtains \( n - 1 \) Lagrangian Factors of other users for query reconstruction by disguise and last sent Lagrangian Factor (LF). After obtaining \( n - 1 \) LF, the attacker can build \( n - 1 \) equations, which contain \( k't \) equation coefficients. Since \( k't > n - 1 \), and it doesn’t meet the condition of the solution of the linear equation system that \( n - 1 \) equations do not get a unique solution. In this context, only the linear correlation result of the original query \( Q \) can be obtained, even though no information is obtained at all.

Based on the above analysis, the value of \( \text{PR}(Q|\text{EQ}) \) is close to zero and thus the proposed method is proved to be effective against the anonymous server attack.

**Semantic Attack:** The semantic attack represents the probability \( \text{PR}(\text{RL}|\text{AS, BK}) \) that the attacker inferred the semantic category of the real location (RL), based on the semantic background knowledge (BK), including temporal semantic knowledge, geographic semantic knowledge and sensitive semantic knowledge) that he has about the real user, after obtaining the anonymity set by stealing. Specifically, the operation \( \text{PR}(\text{RL}|\text{AS, BK}) \) that improves the probability of inferred semantic of real location by filtering the anonymous locations that are not related to the background knowledge one possesses.

**Theorem 3:** Our method can ensure security under semantic attack.

The anonymous set constructed based on the privacy-preserving method in this paper satisfies two semantic differences, including the semantic difference of geography and the temporal semantic difference. Thus, the anonymous set contains \( m_1 \) anonymous locations with different semantic categories of geography and \( m_2 \) anonymous locations with the temporal semantic difference, where \( m_2 \geq m_1 \).

When the attacker launching a semantic attack to an anonymous set based on geographic knowledge BK1, it includes \( m_1 \) geographically distinct semantic locations, and thus the probability that the attacker identifies the real requested location based on the semantic knowledge is \( 0<\text{Pr}(\text{RL}|\text{AS, BK1})<1/m_1 \). Moreover, since the semantic description vector contains information about the geographic semantics and the anonymous location is chosen to maximize the semantic difference of the description vector, the probability of \( m_1 \) is large and \( \text{Pr}(\text{RL}|\text{AS, BK1}) \) is small, \( \text{Pr}(\text{RL}|\text{AS, BK1}) < \text{neg1} \), where \( \text{neg1} \) is a very small probability.

When the attacker launching a semantic attack to an anonymity set based on temporal knowledge BK2, since the anonymity set contains \( m_2 \) temporal semantic locations of different classes, the probability that the attacker identifies the real request location based on the semantic knowledge is \( 0<\text{Pr}(\text{RL}|\text{AS, BK2})<1/m_2 \). Because the anonymous location is selected by considering maximizing the temporal semantic difference location, \( m_2 \) is large and \( \text{Pr}(\text{RL}|\text{AS, BK2}) < \text{neg2} \), where \( \text{neg2} \) is a very small probability.

Also, the probability of identifying a sensitive semantic location in the anonymity set at this time is small and less than the very small probability \( \text{neg3} \) due to the substitution operation performed on the semantic sensitive location.

According to above analysis, the probability \( \text{Pr}(\text{RL}|\text{AS, BK})+\text{neg3}<\text{neg1}+\text{neg2}+\text{neg3} \) (BK=BK1+BK2) after performing the semantic attack and it is difficult to occur.

### 6. Experiment

#### 6.1 Experiment Setting

To verify the method proposed in this paper, we implemented all the methods in Python language and ran them on a 3.20GHz Intel octa-core GPU, 16GB RAM, and Windows 10 computer. The experimental data came from the domain recognized TDS data set [26], which included more than 1500 users’ request locations generated by Brinkhoff software simulation and extracted 500 users’ data and added their simulated query in our experiment. The example of a dataset is shown in Fig. 7. In addition, we use the API provided by Baidu Map to add semantic information features of TDS data. Specific includes six main categories and 44 subcategories, 110 meters and 10 minutes were used as the thresholds for stop points (i.e., more than 10 minutes away from 110 meters), and 1822 stay-points were extracted from the TDS for the construction of a personal semantic tree.

Based on the association knowledge of the semantic location and user identity information that the attacker understands, combined with the Bayesian inference model, the inference attack on the anonymity set can be carried out. In order to avoid accidental and special cases, the experimental results are all from the results of 500 operations and the main parameters setting, as shown in Table 1. Finally, To evaluate the proposed method, we conduct experiments
from two aspects utility and privacy.

To verify availability, the AA and TC metric are used to measure.

1) Anonymity Area (AA) [17]: We use the area of the smallest circle (R represent its radius) that can cover all anonymous locations in each anonymity set to evaluate the availability of the proposed method as

\[ AA = \pi R^2 \]  

(19)

2) Time Cost (TC) [28]: We use the time overhead of the proposed method to complete the requested service, including the total time of Encryption of query (t₁), generation of anonymity set (t₂) and decryption of query and return of service (t₃), to evaluate the time availability of the method and represent as

\[ TC = t_1 + t_2 + t_3 \]  

(20)

To verify privacy, the ASD and ROT metric are used to measure method privacy for user under semantic attack and PERL is used for measure its privacy in a common scenario.

1) Probability of Exposing Real Location (PERL) [27]: We use the ratio of anonymous locations filtered out \( N_f \) by the attacker to all the anonymous locations \( N \) to evaluate the privacy of proposed method, and express as

\[ PERL = \frac{1}{N - N_f} \]  

(21)

2) Average Semantic Distance (ASD) is calculated by averaging the semantic difference distance of the anonymous location and this semantic difference distance as the number of edges between two positions in the semantic tree \( d_s() \).

\[ ASD = \frac{1}{k} \sum_{i=1}^{k} d_s(l_i, l_{i+1}) \]  

(22)

3) Ratio Of Sensitive Location (ROS): We use the ratio between sensitive semantic locations relevant to the user \( N_s \) and the number of anonymous locations in 100 anonymity sets \( N_{sum} \) to Verify privacy under semantic attack ROS

\[ ROS = \frac{N_s}{N_{sum}} \]  

(23)

6.2 Efficiency and Analysis

In this section, we analyze the efficiency of our method in two aspects: Usability and Privacy. In particular, we consider the evaluation index under semantic attack to verify privacy security.

1) Utility Analysis: In this section, we analyze the effect of different parameters on the efficiency of our method based on Time Cost (TC). Figure 8 (a) shows that with the number of k increased, the TC curve has a low increase trend. This is due to more privacy levels needing our method to find more anonymous locations to satisfy semantic differences and thus result in the TC. Moreover, since the higher \( n_a \) needs more anonymous server collaborative to generate and take more time in this process, higher \( n_a \) has longer protection.

Figure 8 (b) shows that with the number of anonymous servers increasing, the TC curve has increased trend. This is because, as the anonymous server increases, the query encryption process needs to take more time (including partition time and secret computation time). Thus it increases the total time cost. Moreover, due to the privacy degree k
increase need time for selecting more anonymous location, the curve of larger k with the higher the time cost than lower.

As shown in Fig. 8 (c), with the increase of privacy demand k, the AA curve presents an upward trend. And this growth trend is gradually accelerating. This is because the higher privacy requirements make the number of locations in the anonymity set grow. Also, since we consider sensitive language locations for replacement, these selected locations are far from the requested location. Moreover, this increase in the number of locations to be replaced leads to accelerated growth of AA as the privacy requirement k increases. Therefore, we need to choose the appropriate privacy requirements and thus obtain the tradeoff between availability and privacy.

As shown in Fig. 8 (d) shows that with the number of anonymous servers increasing, the AA curve has decreased trend. And this downtrend is gradually decelerating. This is because the more anonymous servers we have, the more efficient our method can be, and the closer the anonymous location is to ensure that our semantic privacy needs are met.

2) Privacy Analysis We first verify the security of our method based on PERL, which is used to common metrics for verifying privacy security. To further verify the proposed method’s security, we analyze privacy based on ASD, ROS while adversaries know sensitive semantic information about users. Specifically, the attack is carried out by sensitive semantic location-tagged data as prior knowledge and then attacked by the Bayesian model. ASD and ROS have recognized metrics for verifying privacy security against semantic attacks.

As shown in Fig. 9 (a), with the increase of privacy demand k, the ASD curve presents an upward trend. This is because the higher privacy requirements increase the number of locations belonging to different semantic types in the anonymity set. When the anonymous locations are chosen for anonymity, our method considering its semantic type, should be different from the chosen location. To improve semantic diversity further, chosen locations should not be in the same semantic class as the real location. This reason made the ASD increase and ensured the diversity of anonymous locations for preventing the semantic attack.

As shown in Fig. 9 (b), the ROS curve presents a downward trend with the increase of privacy demand k. This is because the higher privacy requirements make the number of locations in the anonymity set grow. Also, since we consider sensitive language locations for replacement, it leads to these selected locations for replacement and thus reduces the number of sensitive locations. The above analysis shows that our method can effectively avoid the privacy damage caused by sensitive semantic exposure.

Figure 9 (c) shows that with the increase of privacy demand k, the PERL curve presents a slow downward trend. And this trend slows down gradually. The reason for this phenomenon is twofold: 1) The higher privacy requirements make it more difficult for the attack, which makes the PERL lower. 2) At the same time since our approach considers both sensitive semantics and the effect of semantic differences for anonymous location selection, it makes it further difficult for the requested location to be identified even if the attacker has relevant background knowledge.

6.3 Comparison and Analysis

In this section, in order to further illustrate the effectiveness of our method, the sensitive location diversity anonymity set construction method, called STA-LPPM in [17], the anonymity set construction method based on road network semantic diversity called LPRN[16] and the location privacy protection method called NLP in [8], which combines simple semantic of locations and query probability are choose for comparison.

AS shown in Fig. 10(a), with the increase of k value, the curve of TC shows a slow upward trend, and our method has achieved the second best effect after NLP. The reason for this phenomenon is the cost time of our method, including encryption and decryption of queries, the anonymous communication between the server and the user role-sensitive semantic model generation, and anonymity collection construction. Although it is optimized in our ap-
As shown in Fig. 11 (a), PERL of all the methods shows an upward trend with the increase of k value, and our scheme achieves a lower PERL than the other three methods, especially in the case of increased user privacy requirements. The main reason for this phenomenon is that this paper adopts the anonymity set construction method based on sensitive semantics, and the best effect is achieved. And this advantage is becoming more and more evident as the need for privacy increases. With the increasing demand for anonymity, our location semantic modeling method can provide more choices and achieve better results than the traditional basic semantic annotation. Since STA-LPPM considers the availability and privacy of semantic anonymity construction methods, the initial effect is similar to that of this paper. However, with the increase of privacy demand, it is challenging to meet the privacy demand because it only considers the simple semantic model. Since LPRN considers simple network semantics but does not have balanced optimization, it achieves the third most effective result. However, NLP does not consider the anonymity of semantic location, so it has the highest probability of leakage.

As shown in Fig. 11 (b), ASD of all the methods shows an upward trend with the increase of k value, and our scheme achieves the highest ASD than the other three methods, especially in the case of increased user privacy requirements. This trend is because our anonymity set construction method considers the semantic diversity that combines geography and user-related semantic information. Since STA-LPPM adopts semantic relativity, representing the multidimensional difference of semantic to describe fine-grained semantic information of location and thus generates more categorical candidate locations for choosing, good results were obtained. Since LPRN considers more comprehensive semantic information with the constrain of road network than NLP, it doesn’t think the semantic information, it has better ASD than NLP.

As shown in Fig. 11 (c), the ROS of all the methods shows a downward trend with the increase of k value, and our scheme achieves a lower ROS than the other three methods. The main reason for this phenomenon is that our method considers describing sensitivity semantics location and replacement for anonymity and thus avoids revealing user’s sensitive location. Because the STA-LPPM considers the simple sensitivity of locations not user-related. It achieves a second effective result since the other two anonymity methods do not consider the negative effects of sensitive semantic locations and thus achieve close and worst outcomes.

7. Conclusion

Considering the sensitive semantic locations releasable user identifier and single anonymous server attack, based on multiple-anonymous servers, we propose a user identical-correlation location semantic privacy protection method, including query encryption and anonymity methods. To resist the attack of anonymous servers, a query encryption method combining the Shamir mechanism and multiple-anonymous servers is proposed. In order to improve the ability to re-
sist attacks on sensitive semantic locations, we adopt the personal semantic tree based on mobile patterns and the replacement of sensitive semantic locations for anonymity. The experiments result show that our method outperforms previous methods regarding privacy and usability.

Acknowledgments

This research was partially supported by China National Key Research and Development Program under Grant No.2019YFC1521400, and the National Science Foundation of China (NSF) under grants No.62072362, No.61672426, No.61572401.

References

[1] G. Beigi and H. Liu, “A Survey on Privacy in Social Media: Identification, Mitigation, and Applications,” ACM Transactions on Data Ence, vol.1, no.1, pp.1–38, 2020.
[2] B. Liu, W. Zhou, T. Zhu, L. Gao, and Y. Xiang, “Location privacy and its applications: a systematic study,” IEEE Access, vol.6, no.17600–17624, 2018.
[3] S. Barker and G.J. Ahn, “Location privacy protection through obfuscation-based techniques,” Lecture Notes in Computer Science, vol.4602, Chapter 4, pp.47–60, 2007.
[4] R.-H. Hwang, Y.-L. Hsieh, and H.-W. Chung, “A Novel Time-Obfuscated Algorithm for Trajectory Privacy Protection,” IEEE Transactions on Services Computing, vol.7, no.2, pp.126–139, 2014.
[5] J.I. Hong and J.A. Landay, “An architecture for privacy-sensitive ubiquitous computing,” MobiSys 2004, 94720-1776, 2004.
[6] M.L. Yiu, C.S. Jensen, X. Huang, and H. Lu, “SpaceTwist: Managing the Trade-Offs Among Location Privacy, Query Performance, and Query Accuracy in Mobile Services,” IEEE International Conference on Data Engineering, IEEE, 2008.
[7] S. Hayashida, D. Amagata, T. Hara, and X. Xie, “Dummy Generation Based on User-Movement Estimation for Location Privacy Protection,” IEEE Access, vol.6, no.22958–22969, 2018.
[8] T. Hara, A. Suzuki, M. Iwata, Y. Arase, and X. Xie, “Dummy-Based User Location Anonymization Under Real-World Constraints,” IEEE Access, vol.4, no.673–687, 2016.
[9] M. Gritn and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” Proceedings of the 1st international conference on Mobile systems, applications and services, ACM, pp.31–42, 2015.
[10] Z. Hao, X. Meng, H. Hu, and Y. Huang, “You can walk alone: trajectory privacy-preserving through significant stays protection,” International conference on database systems for advanced applications, Springer, Berlin, Heidelberg, vol.7238, pp.351–366, 2012.
[11] H. Liu, Xing-Hua L I, B. Luo, et al., Distributed K-Anonymity Location Privacy Protection Scheme Based on Blockchain, 2019.
[12] S. Zhang, K.-K.R. Choo, L. Qin, and G. Wang, “Enhancing privacy through uniform grid and caching in location-based services,” Future Generation Computer Systems, vol.86, pp.881–892, 2017.
[13] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, “Enhancing privacy through caching in location-based services,” 2015 IEEE Conference on Computer Communications (INFOCOM), IEEE, 2015.
[14] S. Zhang, L. Qin, and G. Wang, “A Caching-Based Privacy-Preserving Scheme for Continuous Location-Based Services,” International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, Springer International Publishing, vol.10067, pp.73–82, 2016.
[15] L. Zhu, C. Xu, J. Guan, Y. Liu, and H. Zhang, “A preference-aware trajectory privacy-preserving scheme in location-based social networks,” IEEE INFOCOM 2017 -IEEE Conference on Computer Communications Workshops (INFOCOM WKSHP), IEEE, 2017.
[16] C. Tian, H. Xu, T. Lu, R. Jiang, and Y. Kuang, “Semantic and Trade-Off Aware Location Privacy Protection in Road Networks Via Improved Multi-Objective Particle Swarm Optimization,” IEEE Access, vol.9, pp.54264–54275, 2021. doi: 10.1109/ACCESS.2021.3071407.
[17] 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, no.184965–184975, 2020.
[18] W.U. Zhong-Zhong, L.V. Xin, and L.I. Xin, Query Probability Based Dummy Location Selection Algorithm, Computer Science, 2018.
[19] C. Hs, L. Harm, Z. Xia, and M. Zhang, “Non-Interactive Dealer-Free Dynamic Threshold Secret Sharing Based on Standard Shamir’s SS for 5G Networks,” IEEE Access, vol.8, pp.203965–203971, 2020. doi: 10.1109/ACCESS.2020.3035278.
[20] R. Steinfeld, J. Pieprzyk, and H. Wang, “Lattice-Based Threshold Changeability for Standard Shamir Secret-Sharing Schemes,” IEEE Transactions on Information Theory, vol.53, no.7, pp.2542–2559, July 2007. doi: 10.1109/TIT.2007.895941.
[21] U. Erra, S. Senatore, F. Minnella, and G. Caggianese, “Approximate TF-IDF based on topic extraction from massive message stream using the GPU,” Information Sciences, vol.292, no.143–161, 2015.
[22] G. Qiu, D. Guo, Y. Shen, G. Tang, and S. Chen, “Mobile Semantic-aware Trajectory for Personalized Location Privacy Preservation,” IEEE Internet of Things Journal, vol.8, no.21, pp.16165–16180, 2020.
[23] J. Silva, "Turismo, crescimento e desenvolvimento: uma análise urbano-regional baseada em cluster," Urban Public Economics Review, vol.3, 2005.
[24] W. Ni, G. Chen, J. Lu, et al., "Local entropy based weighted subspace outlier mining algorithm," Journal of Computer Research and Development, 2008.
[25] F. Wei, P. Vijayakumar, N. Kumar, R. Zhang, and Q. Cheng, “Privacy-Preserving Implicit Authentication Protocol Using Cosine Similarity for Internet of Things,” IEEE Internet of Things Journal, vol.8, no.7, pp.5599–5606, 2021.
[26] Thomas brinkhoff, “Network-based generator of moving objects,” http://iupg.jade-hs.de/personen/brinkhoff/generator/
[27] X.Y. Xia, Z.H. Bai, L.I. Jie, et al., “A location cloaking algorithm based on dummy and Stackelberg game,” Chinese Journal of Computers, 2019.
[28] S. Zhang, G. Wang, et al., “Trajectory privacy protection method based on multi-anonymizer,” Computer research and development, vol.056, no.003, pp.576–584, 2019.
[29] X. Yang, L. Gao, J. Zheng, and W. Wei, “Location Privacy Preservation Mechanism for Location-Based Service With Incomplete Location Data,” IEEE Access, vol.8, pp.3958–39584, 2020. doi: 10.1109/ACCESS.2020.2995504.

XuDong Yang received the M.S. degree from Northwest University, China in 2018. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Northwest University. His research interests include privacy protection, edge computing.
Ling Gao is currently a professor of computer science at Northwest University, Xi’an, China. He received his BS degree in Computer Science from Hunan University and his MS degree in Computer Science from Northwest University, in 1985 and 1988 respectively. In 2005, he received PhD degree in Computer Science from Xi’an Jiaotong University, Xi’an, China. His research includes Network Security and Management, Embedded Internet Service. Prof. Gao is the director of China Higher Educational Information Academy, vice chairman of China Computer Federation Network and Data Communications Technical Committee, member of IEEE, CAET (China Association for Educational Technology) director.

Yan Li is currently pursuing the M.S. degree with the Department of Computer Science, Northwest University. His research interests include privacy protection, edge computing.

JiPeng Xu is currently pursuing the M.S. degree with the Department of Computer Science, Northwest University. His research interests include privacy protection, edge computing.

Jie Zheng received the Ph.D. degree in computer science from Xi’an Dianzi University, Xian, in 2008. He is currently a associate Professor of the Department of Computer Science, Northwest University. His research was sponsored by several the National Science Foundation. His research interests include privacy protection, edge computing.

Hai Wang received the Ph.D. degree in computer science from Xi’an Jiaotong University, Xian, in 2003. He is currently a associate Professor of the Department of Computer Science, Northwest University. His research was sponsored by several the National Science Foundation. His research interests include privacy protection, edge computing.

Quanli Gao is currently a PhD candidate in the department of Computer Application Technology at Northwest University, Xi’an, China. His research interests are in areas of Recommender system, Artificial Intelligence.