IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 25, 2019, accepted April 17, 2019, date of publication April 29, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913547

T-SR: A Location Privacy Protection Algorithm Based on POI Query

LI KUANG, (Member, IEEE), SHUAI HE, YUYOU FAN, HUAN ZHANG, AND RUYI SHI
School of Computer Science and Engineering, Central South University, Changsha 410075, China

Corresponding author: Li Kuang (kuangli@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772560, and in part by the National Key Research and Development Program of China under Grant 2018YFB1003800.

ABSTRACT With the popularity of location-based services in the field of mobile network applications, users enjoy the convenience on one side, they may face the risk of privacy disclosure on the other side. Attackers can easily dig out the user's home address, occupation, and other personal privacy from the data of location-based services. For the problem of user location privacy disclosure based on semantic query of point of interest (POI) in the road network environment, the previous research seldom paid attention to the temporal association relationship between the query semantic and the user's location, as well, seldom considered whether the constructed anonymity sets of fake-locations can fulfill the property of reciprocity. Therefore, in this paper, we propose a location privacy protection algorithm T-SR based on POI query. The algorithm can be divided into three steps. First, the whole road network is divided into multiple areas based on the Voronoi structure, so that the areas are independent and non-overlapping, and the road grid that contains user's real location will be located according to the level of privacy protection H. Second, the temporal association between the POI semantics are mined from the real check-in data based on Gaussian distribution σ rule, and the POIs which have weak association with the POI of user's location will be filtered to resist the temporal association attack. Third, the POI semantics within the grid will be split into several buckets according to query times record, anonymity degree k, and semantic degree l, then the dummy locations will be selected from various kinds of POIs of the bucket where the user's location is. The resulted anonymity set can defend against replay attack, inference attack, and temporal association attack. The theoretical analysis and experimental evaluation prove that the proposed solution can protect user's location privacy efficiently and effectively.

INDEX TERMS Location privacy, reciprocity, temporal incidence relation, Voronoi.

I. INTRODUCTION

In recent years, with the rapid development of wireless communication technologies and location services, applications based on location-based services (LBS) have been used more and more frequently. Among them, LBS based on POI query is one of most widely used services. However, while people enjoy the convenience brought by LBS, they also expose their privacy information to malicious attackers invisibly, and it may even harm their personal safety or property. For example, if a user Anny navigates to company (query POI: an office building) from home (current location POI: a residential area) at 8:00am every working day, and after analysis based on big data, the attacker can infer Anny's home address, even Anny's occupation and living habits. Therefore, in this paper, we aim to study how to construct an algorithm to protect user's location privacy in POI query service. The algorithm should guarantee that the query quality of service will not decrease much, while the constructed anonymity set can resist from replay attack, inference attack and temporal association attack.

In the past, many privacy protection models for road network in POI query have been proposed to protect location privacy, but there are some problems with the existing solutions:

(1) They seldom consider whether the constructed anonymity set can fulfill the property of reciprocity. Reciprocity means, within a certain period of time, there are no

The associate editor coordinating the review of this manuscript and approving it for publication was Shuiguang Deng.

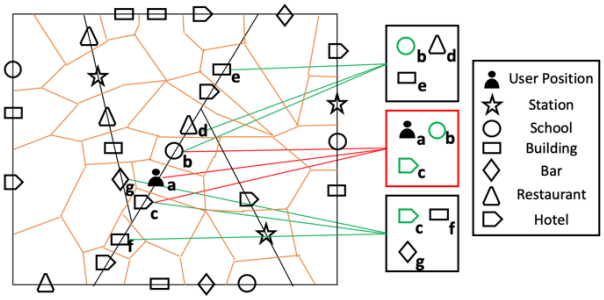


FIGURE 1. An example of anonymity sets which cannot fulfill the property of reciprocity.

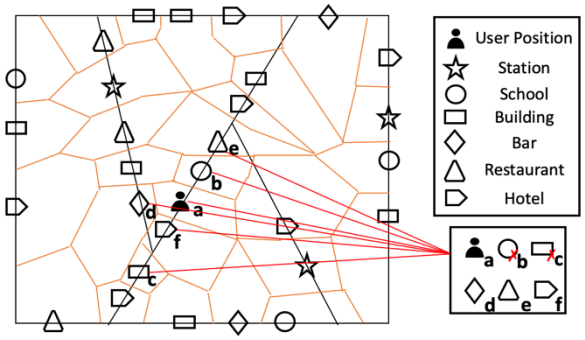


FIGURE 2. An example of temporal association attack.

difference between the anonymity sets which are re-generated according to the true position or the fake position in the original anonymity set. If reciprocity cannot be fulfilled, the anonymity set will not be able to resist from replay attack, since attacker can locate the real user through multiple queries and related background information. For example, as shown in Figure 1, the attacker filches the anonymity set of a query Q as $\{a, b, c\}$, and the corresponding location semantics are {train station, school, hotel}. If the attacker gangs up with the users at the school and hotel, knowing that the anonymity sets submitted by the user in the two locations are $\{b, d, e\}$ and $\{c, f, g\}$, respectively, then through the intersection and reasoning, the attacker can infer the real user's location in current query Q is a.

(2) They seldom consider the temporal association relationship between the query semantic and user's location, which can also lead to privacy disclosure. As shown in Figure 2, it is assumed that user Anny invokes location query service at 2:00 am near the train station, and she would like to query the hotel nearby. If the attacker not only filches the anonymity set $\{a, b, c, d, e, f\}$ submitted to the LBS, the corresponding location semantics of which are {"station", "school", "office building", "bar", "restaurant", "hotel"}, but also knows the temporal association relationships between the semantics category, which indicates that the association between "school" and "hotel" and the association between "office building" and "hotel" are weak at 2:00 am, it is possible to filter out the two fake positions b and c with a high probability.

In order to solve the problems mentioned above, in this paper, we propose a location privacy protection algorithm T-SR based on POI query. The algorithm consists of three steps. First, by constructing the Delaunay triangulation, the road network is divided into independent and non-overlapping Voronoi Cells (VC for short) [1]–[7] based on POIs. Second, based on the POI check-in data, the association relationships between POI categories are defined by the temporal transition from one POI to another, and then the weak associations between semantic categories can be identified according to the Gaussian distribution σ rule, and the result will be used as a basis when filtering POI candidates from the anonymity set. Third, for each query, the road grid that contains user's real location will be located according to the level of privacy protection h first, and then the POIs which have weak association with the POI of user's location will be filtered, next, the POIs within the road grid will be split into several buckets based on query times record, anonymity degree k and semantic degree l, and then the dummy locations will be selected from various kinds of POIs in the bucket where the user's location is. The generated anonymity set can satisfy the location k-anonymity [6]–[11] and l-diversity, the property of reciprocity, and the temporal association with user's location.

The main contributions of this paper are summarized as follows:

(1) We propose the temporal association attack model, and correspondingly, we dig out the weak association relations between POI categories based on the real check-in data, and then the results are used to filter the candidates of the anonymity set.

(2) The proposed T-SR algorithm divides the road network by VC, and splits the POIs in the user's grid into several buckets according to query times record, anonymity degree k and semantic degree l, making the constructed anonymity set cannot be distinguished in terms of POI semantics and query times, thus fulfill the property of reciprocity.

(3) Theoretical and experimental results prove that the proposed T-SR algorithm can resist from the attack models including replay attack, inference attack and temporal association attack, and the anonymity set generated by T-SR algorithm has advantages over other algorithms in terms of verisimilitude, the number of weak associations, and the area of anonymity region.

The rest of this paper is organized as follows. In Section 2, we review the related work on location privacy protection. In Section 3, we introduce the attack models, and related definitions. The details of T-SR algorithm and the security analysis are presented in Section 4 and Section 5 respectively. Then the proposed algorithm is evaluated with an extensive experiment on a public dataset Oldenburg in Section 6. Finally, conclusions and future work are given in Section 7.

II. RELATED WORK

To solve privacy leakage problems in location services, many researches have presented various problem scenarios and

solutions, and in particular, different solutions and protection technologies are specialized to its corresponding situations. In recent years, considering that real road network environment has a higher similarity to the real scene than Euclidean space, an increasing number of researchers focuses on privacy protection based on location semantics, and distortion method and encryption method are two common LBS privacy protection technologies in road network environment among those algorithms.

One basic idea in privacy protection is space cloaking methods [8]–[14]. Gruteser and Grunwald [8] proposed a concept, location k-anonymity, which can make location of users who send the service request indistinguishable from other $k-1$ locations, so the probability of identification successfully becomes lower than $1/k$. Chow *et al.* [9] proposed Casper Cloak algorithm which divides the space into H layers by a quad tree and each layer maintains the information of whole spatial structure. Sun *et al.* [11] designed a geolocation tag to distinguish sensitive locations and common locations in order to minimize the response time of the requesting service and protect the user's location privacy through spatial stealth technology. Li and Zhu [12] have proposed partition method for cloaking region. Voronoi diagrams have been used in a large number of LPPMs, for both real time and offline techniques. The work of Hashem *et al.* [13] presents a technique to aggregate data using Voronoi diagrams, in order to guarantee a defined level of K-anonymity in the dataset.

Dummy position generation methods [14]–[25]. Li and Palanisamy [14] assumed that attackers may already know background knowledge about user's behavior pattern and trajectory similarity. Their algorithm generates false trajectories from real trajectory by rotation, which can ensure trajectory similarity, and then adjusts the points on the trajectory to meet the trajectory request probability. The DUMMY-Q model [15] uses context of query and user's motion model to protect query semantics in consecutive requests, which directly leads to a result that a more diversified fake query brings a securer user privacy protection. Guo *et al.* [17], [18] proposed a method using dynamic pseudonym change mechanism and user-controllable fake location generation mechanism with combining the geometric transformation algorithm to achieve privacy protection. Huo *et al.* [20] designed a trajectory anonymity algorithm YCWA for constructing a trajectory k-anonymous set, which is easier to control information loss degree based on graph partitioning. Gao *et al.* [21] proposed a privacy protection method based on personalized trajectory privacy graph model, which fully considers influencing factors such as trajectory direction, time and spatial distance, and it finds an approximate optimal trajectory k-anonymous set constructed by minimum spanning tree. Pan et al. [23] take users' moving speed and direction as affecting factors and a distortion function is defined to measure the temporal query distortion of a cluster in continuous queries, and they also have proposed a method which maintains maximal cliques to deal with location dependent attacks [25]. Though these algorithm has its own merits, there are still some disadvantages like a slight discrepancy between anonymity sets constructed by such algorithms in the case of multiple queries. In addition, most privacy protection technologies do not protect the query information, so attackers can infer the user's potential real semantic information according to the POI semantics relationship between fake locations and the query location, resulting in privacy leakage.

For encryption-based location privacy protection technology [26]–[32], it can send encrypted requests directly to the LBSs without using a third-party trusted server, and then decrypt the result on the client. This method is highly secure, while the computational cost is huge, the deployment is complicated, and the algorithm also needs to be optimized. Khoshgozaran and Shahabi [26] proposed an encryption method based on Hilbert curve, which can transfer the user's position and user POIs from two-dimensional coordinates into one-dimensional encryption space. The one-dimensional encryption space transformed by two Hilbert curves with different parameters still maintains the proximity in the two-dimensional space, so that k-nearest neighbor query and range query can also be performed in one-dimensional encryption space. PIR (Private Information Retrieval) [27] replaces the homomorphic comparison step with unconscious transfer to achieve a safer solution for the sender and receiver, making the privacy protection more robust. The PIR technique used by Kalnis *et al.* [27] relies on the quadratic remainder theorem, namely that it is difficult to obtain a prime factorization of a large number from calculation, and the location service provider can respond to the query result without knowing the query information. Papadopoulos *et al.* [29] introduced a query plan mechanism to defend against pattern attacks, making requests with different access frequencies difficult to distinguish. Reference [30] proposed an enhanced location privacy preserving (ELPP) scheme, which utilizes a function generator to generate the transforming parameters, and the anonymizer knows nothing about a user's real location unless it knows the parameters of function generator. However, the burden on clients of such technologies is relatively heavy, and there is no effective trade-off between service quality and privacy protection requirements. Moreover, a large part of technologies doesn't consider the privacy leakage caused by attacker's possession of side information (such as query frequency).

In summary, the main limitations of extant work are: (1) Seldom consider whether the generated anonymity set have reciprocity [32], and have the risk of resisting replay attack failed. (2) The extant works only consider protecting the user's location, without protecting user's query semantic, so the attacker can infer some fake locations after considering temporal association relationship between location semantic and query semantics, resulting in privacy leakage. (3) The existing works seldom find a good tradeoff between location privacy protection and the quality of service simultaneously.

III. PRELIMINARIE

A. ATTACK MODEL

In general, there are two types of attackers in location service requests, active and passive attackers. Active attackers can steal communication information between entities, whereas passive attackers steal encryption algorithms of constructing anonymity set and other database information by attacking the server. Different attackers have different background knowledge and different attack directions. We assume that a malicious attacker in the location service is both active and passive type to steal location privacy, namely: (I)The attacker utilizes the LBS, intercepts the request query, and analyzes out the anonymity set. (II)The attacker can steal anonymous algorithms on anonymous server. (III)The attacker can obtain relevant supplementary information by attacking the LBS such as road network structure, query frequency, historical query and location table. The goal of our approach is not only to resist replay attack and inference attack from common types of attackers, but also temporal association attack. Attack models are described as follows:

1) **Replay attack**: the attacker steals the anonymous algorithm and the anonymity set that send to the LBS, and then re-anonymizes the fake location from the anonymity set to get a new anonymous set. After that, the attacker compares the new anonymity set with the original anonymity set, and finally infers the user's real location. The example without reciprocity mentioned in the introduction is one of the typical cases that attackers use replay attack to steal location privacy.
2) **Inference attack**: the attacker steals the anonymous algorithm and relevant supplementary information in the anonymizer, and infers user's real location from the anonymity set, by comparing, mapping and inferring the original anonymity set. For example, the attacker steals the two location semantics of hospital, bank} in the anonymity set, when the user Anny requests the location service. If the attacker has the supplementary information of semantic query frequency and through comparing information knows that the query frequency of user in the hospital is much larger than in the bank, then he might reasonably infer that the true location semantic of Anny's query is in the hospital. In another case, the attacker owns the road network information, and by mapping the fake locations in the anonymity set to the road network structure, it can be found that most of the fake locations are distributed in the hospital area, while only a small part is distributed around the bank, so it can be inferred that the semantic of user's true location is the hospital.
3) **Temporal association attack**: as mentioned in the introduction, there are different visit and query probabilities between the semantics of POI in each specific time period. Temporal association attack means that there may have huge difference in the probability of accessing the specific query semantics for the location semantics corresponding to the fake location of the anonymity set. If this difference is ignored, the attacker can easily filter out some of the fake locations, making there a risk of privacy disclosure.

B. RELATED DEFINITION

Definition 1 (Query Request): The request query sent by the user is the tuple $Q_u < uid, t, uloc, quloc >$, and the corresponding attributes respectively represent the user ID, the current query time, the location of the user (corresponding to the location semantic), and the semantic of the user query request (corresponding to the query semantic).

The anonymizer anonymously processes the user's real location uloc to construct the anonymity set CR, and then sends the anonymous query $Q_c < uid, t, CR, quloc >$ to the LBS for query request.

Definition 2 (Location Privacy Requirements): Different users and location request services have various privacy protection focuses. We use triplet $< K, L, H >$ to represent privacy protection requirements. The tuple attributes are defined as follows: (1) K is the anonymity degree k, which means that the probability that an attacker can locate a real user without considering background knowledge no more than 1/k. (2) L is the semantic degree l, which means the location of the anonymity set at least l semantic categories, so the probability that an attacker can derive true position semantic based on anonymity set semantic attributes is no more than 1/l. (3) H is the privacy protection level coefficient. The smaller h means the smaller the grid area of the anonymity set selected by the algorithm, the higher the service quality, and the lower the privacy protection level respectively. Conversely, it would be a lower quality of service and a higher privacy protection level.

Definition 3 (Reciprocity): The anonymity set has reciprocity means that the anonymity set constructed are identical when two users in the same set request location services. For example, the anonymity set constructed by the user U requesting the POI query at a certain moment is $AS^k(U)$, where k is an anonymity degree, and for $\forall V \in AS^k(U), AS^k(V) = AS^k(U)$ is satisfied.

C. SYSTEM FRAMEWORK

Most of the existing privacy protection frameworks use a trusted central anonymizer to bridge users and LBS servers, instead of the traditional privacy protection framework CS. Anonymizer can quickly respond to query service requests, reduce client computing overhead, improve service quality, and avoid privacy disclosure when LBS is not trusted. The workflow of our privacy protection framework is shown in Figure 3, and the process sequence is identified by Arabic numbers 1 to 4. At the beginning, the client submits the query request Q_u to the anonymizer. Then, the anonymizer responds to the request, constructs the anonymity set CR after anonymizing the real location, and sends the anonymous request Q_c to the LBS. In the third step, the anonymizer

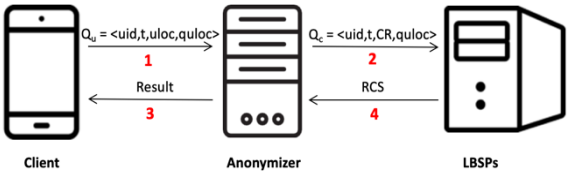


FIGURE 3. Privacy protection framework.

receives the query result set RCS returned by the LBS; finally, the anonymizer refines the query result, and returns the result of the query request to the user.

D. PROBLEM DEFINITION

As mentioned in the introduction, the purpose of this paper is to solve the problem of how to protect the privacy of users' location based on the query of POI. The core content of the algorithm is the process from the first step to the second step in the privacy protection framework, namely, how the anonymizer anonymously processes the user-initiated query request Q_u to construct the anonymity set CR. The goal of our algorithm is to construct a reciprocal anonymity set based on the user-defined privacy protection triples, and the algorithm can resist the above three types of attack models.

IV. ANONYMOUS ALGORITHM

A. OVERVIEW OF THE METHOD

Figure 4 shows a brief process describing the construction of the anonymity set by our algorithm. The number in the figure is the execution order of the anonymity set CR's generation. There is a possibility to partition the geographical area in a more natural way for Voronoi [1]–[7] diagrams than a grid. Before the user requests the query, the system has generated the road network V map, the semantic weak incidence relation table, and the POI query frequency table. The construction of the anonymity set is divided into the following five steps: First, the anonymizer maps the user's location *uloc* to the corresponding road network grid according to the user' privacy protection level h and the road network V map. Second, according to the weak incidence relation table, the current query time t and the query semantic quloc, the anonymizer filters the POIs which have weak incidence relation with the semantic of quloc in the grid in the current time period. Third, based on the query frequency table, the anonymity degree k and semantic degree l, anonymizer classifies the POIs in the grid into buckets, so that the POIs in each bucket cannot be distinguished through the query frequency, and the principle of reciprocity is satisfied. Fourth, the anonymizer selects the fake location in the area of the VC in the bucket corresponding to the user, constructing the anonymity set CR; finally, return CR to the anonymizer.

The symbols used in this article are shown in Table 1.

B. V MAP CONSTRUCTION AND ROAD NETWORK DIVISION

We use the undirected graph $G = (D, E)$ to represent the road network model, where D is the set of the road segments' intersections, and E is the collection of network segments

TABLE 1. Symbol definition.

Name	Description
$G = (D, E)$	Road Network Model
$V(P_i)$	The Set of the Points with the Smallest Distance to POI P_i
$V(P)$	All the VC of POIs in the set P
N_{ij}	User grid
$\mathrm{SPOI_i}$	POI semantic i
$Pset_{all}$	All semantic category set
$Pset_{filter}$	Filtered semantic category set
V_{left}	Set of POI after filtering
num_{remain}	Uneven remaining POIs
R_i	Number of road segments of V_i

of the road. The algorithm takes the POI as the seed node and divides the road network into independent sub-areas. The divided road network is called VM (network Voronoi map). *VM* is a geometric structure widely used in spatial segmentation, and can reasonably and effectively represent the neighboring relationship between spatial targets. Suppose that there is a set P containing n POIs in the road network area A, $P = \{P_1, P_2, \ldots, P_n\}$. Define the $\mathrm{V(P_i)}$ of P_i as the set of the points in A with the smallest distance to P_i: $\{V(P_i) = b | d(b, P_i) \leq d(b, P_j), b \in A, j \neq i, j = 1, 2, \ldots, n\}$, that is, the distance of the point in the area of P_i's *VC* to the P_i is shorter than any distance to other POI. We refer to the road network structure composed of all the *VC* of POIs in the set P as *VM*, which is denoted as $V(P) = \{V(P_1), V(P_2), \ldots, V(P_n)\}$, as shown in Figure 5. The location of a POI in the road network environment is fixed, so the *VM* of the road network mapped by the grid corresponding to the POI set is also uniquely determined.

Delaunay triangulation [1], [7] is a kind of method to construct *VM* proposed by Delaunay, a Russian mathematician. The key step that how to construct the Delaunay triangulation network is described in algorithm 1.

As shown in Figure 6, we divide the road network according to the user' privacy protection level $\mathrm{h}(0 < \mathrm{h} \leq \mathrm{n})$. By default, when $\mathrm{h} = 0$, the entire road network is divided into $2^\mathrm{n} \times 2^\mathrm{n}$ grids, and the coordinates corresponding to each grid are $< i, j >, i \leq i \leq 2^n, 1 \leq j \leq 2^n$. Accordingly, when h is 1, the entire road network is divided into $2^{\mathrm{n}-1} \times 2^{\mathrm{n}-1}$ grids, and when h is 2, the whole space is divided into $2^{\mathrm{n}-2} \times 2^{\mathrm{n}-2}$ grids, and so on. Normally, if the POIs in the grid N_{ij} corresponding to the user position satisfy the degree k and l, N_{ij} can be regarded as the selected area of the final anonymity set. Conversely, if the POIs in N_{ij} do not satisfy the degree k or l, the h will be upgraded one layer by default, expanding the area of anonymity set selected until meeting the privacy protection requirements. The smaller the h is, the lower the

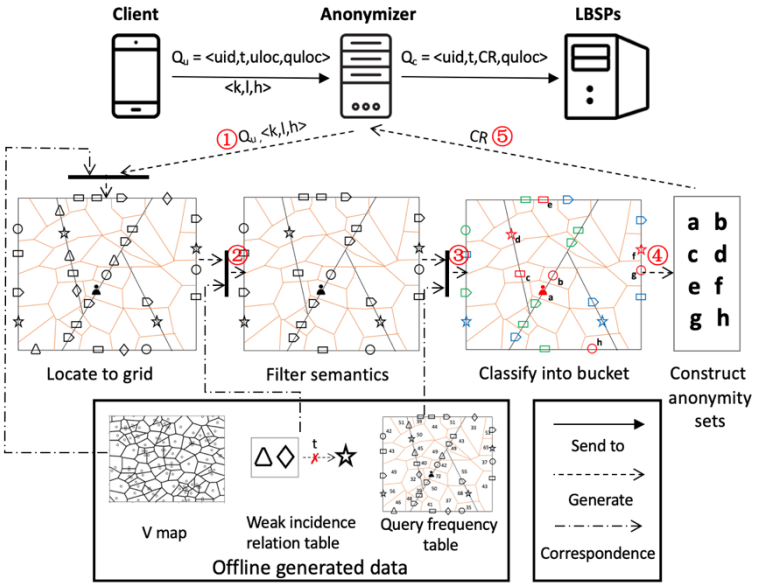


FIGURE 4. The diagram of constructing anonymity set.

Algorithm 1 Delaunay Triangulation Network Division

Input: The set of POI **pois**
Output: Delaunay triangulation **allT**
1. **for** poi in **pois**:
2. **for** t in **allT**:
3. **traverse** to find the affected triangle named influencedT
4. **delete** the affected triangle in **allT**
5. **find** common edges of influencedT
6. **exclude** newT formed by the common edge in influencedT
7. **local optimization** of newT
8. **for** nT in newT:
9. **add** nT in **allT**
10. **return allT**

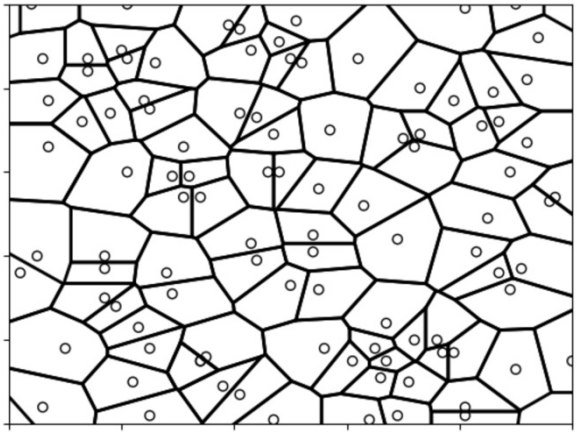


FIGURE 5. An example of VM.

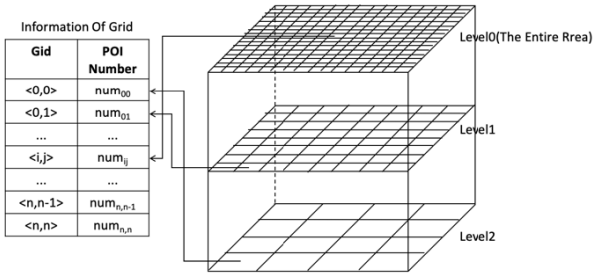


FIGURE 6. Road network division diagram.

privacy protection level is, and the higher the service quality is. On the contrary, the larger the h is, the higher the privacy protection level and the lower the service quality respectively is.

C. POI FILTERING BASED ON TEMPORAL ASSOCIATION RELATIONSHIP

Existing methods of privacy protection ignore the temporal association relationship between the query semantic and the user's location when generating anonymity set. However, according to our experience in life, bars are usually not open at 10:00 am, shopping malls are usually closed at 2:00 am, and a person are most probable to go home or hotels from a train station at 10:00 pm, are less probable to go to bars, and impossible to go to shopping malls from a train station at 2:00 am. In a word, there are different transition probability from one POI to another at different time periods. As shown in Table 2, if the temporal association relationship from user's current location to his destination is neglected, attackers can quickly use common knowledge or related statistical data to identify the fake positions in the anonymity set.

When a user initiates a query, the query POI corresponds to the destination semantic, and the positions in the anonymity

TABLE 2. Examples of weak incidence relation scenarios.

Query Time	Query Semantic	Fake Position Semantic	Weak Incidence Relation Semantic Abandonment Scenario
9am	Office Building	Cinema	This is not the prime time for viewing movies. The probability of users launching location service query for office buildings in the location of cinemas is small.
1pm	Train station	Bar	At this time, the bar is not open, and the probability of query the train station at the location of the bar is small.
2am	Hotel	Market	At this time, the market has been closed down, and the probability of users launching service in the location of market is small.

set correspond to the source semantic. In this paper, we aim to figure out the temporal association relationship from the source to the destination and identify the weak association set, so that attackers cannot distinguish true or fake locations from the anonymity set, since the algorithm will filter out the sources with weak association to the destination. In order to achieve the goal, we first count the transition frequency between different POI semantics, and then identify the weak association.

1) COUNTING THE TRANSITION FREQUENCY BETWEEN DIFFERENT POI SEMANTICS

We use POI sign-in data set to obtain the relationship of POI semantics, then derive the weak association set. We assume that the user must query the POI before signing in. Therefore, the semantics s and d, the start positions and destinations respectively in a sign-in data, have an orderly association at the current time, $s \rightarrow d$.

First, we divide the POIs into different semantic categories. Then we choose a period of time and figure out which POIs that the same user has signed in. Correspondingly, these POIs have a temporal association at the time period. For example, if user Anny signs in a residential area at 8:00 am and signs in an office building at 9:00 am, there is an association between the residential area and the office building during this time period. Third, we will count the transition frequency between different POI semantics in all time periods.

2) IDENTIFYING THE WEAK ASSOCIATIONS BETWEEN POI CATEGORIES

There are different access frequencies in different kinds of POIs, and there are also different association degrees between POI categories. Association rule mining algorithm can be used to mine frequent item sets, and here we use the association rule mining algorithm for reference to identify the weak associations between POI categories. We set the

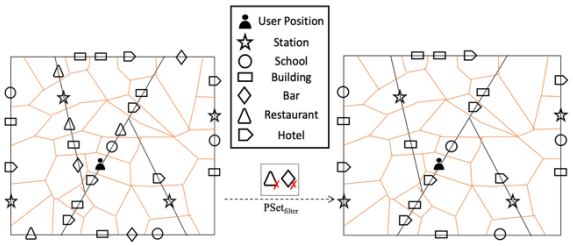


FIGURE 7. Filtering POIs with weak association relation.

support threshold minsup to 0, and the confidence threshold minconf is defined according to the σ rule of Gaussian distribution.

We believe that the distribution of association degree values between different POI categories follows a Gaussian distribution $N(\mu, \sigma)$, where μ and σ represent the mean and standard deviation of the association degrees between the source POI and all the destination POIs respectively. The 3σ criterion is a method for eliminating coarse error data, based on the repeated measurement with equal-precision on the normal distribution data. According to the σ rule of the Gaussian distribution, the probability that the association degree falls in the range of $[\mu - \sigma, \mu + \sigma]$ is 68.27%, while the probability that in the range of $[-\infty, \mu - \sigma]$ is 15.87%.

Therefore, we set $\mu - \sigma$ as the threshold of confidence, that is, if the association degree between the source and a destination POI falls in the range of $[-\infty, \mu - \sigma]$, we determine it as a weak association relation; otherwise, if it falls in the range of $[\mu - \sigma, +\infty]$, we determine it as a strong association relation. The association degree between the two POI categories can be calculated using formula (1), as shown at the bottom of the next page, while the strong and weak association relationship between two POI categories is defined by formula (2), as shown at the bottom of the next page.

In formula (1), Where *Fnum* denotes the access frequency of the two POI semantics $SPOI_i$ and $SPOI_j$ in the time period t, where n represents the number of POI categories, and function T denotes the association degree between two POI categories determined by the ratio of the access frequency from $SPOI_i$ to $SPOI_j$ and the total access frequency from $SPOI_i$ to all the other POIs. In formula (2), where *minconf* is the confidence threshold $\mu - \sigma$, and function *AR* denotes the strong or weak association relationship determined by the value of *minconf* and T between two POI semantics $SPOI_i$ and $SPOI_j$ in the time period t. Some examples of weak incidence relation scenarios are shown as follows:

As shown in Figure 7, based on the query time t and the weak incidence relation table, the anonymizer exports the semantic category set $\text{Pset}_{\text{filter}}$ that has a weak incidence relation with the query semantic in the time period after receiving the query request Q_u sent by the user, and then filter out the POIs that match the set of semantic category $Pset_{filter}$ in the grid. And the anonymity set will be selected in the remaining set of POIs V_{left}. The pseudo code of $V'_{left}s$ construction algorithm is described in algorithm 2.

Algorithm 2 Filtering Pois with Weak Association

Input: Sign in data **Data**, Set of POIs $\mathbf{V}_{all}$, Query semantic **qv**, Query time **t**
Output: Remaining set of POIs $\mathbf{V}_{\mathbf{left}}$
1. for data in **Data**:
2. **count** the $T(SPOI_i, SPOI_j, t)$ between semantics
3. mine weak associations between semantic categories with Gaussian distribution
4. Get the weak associations $\text{PSet}_{\text{filter}}$ of **qv** in **t**
5. for v in $\mathbf{V}_{\mathbf{all}}$:
6. **for** p in $\text{PSet}_{\text{filter}}$:
7. **if** P(v) and p are not equal:
8. **add** v into $\mathbf{V}_{\mathbf{left}}$
9. return $\mathbf{V}_{\mathbf{left}}$

D. CLASSIFICATION OF INTEREST POINTS

As mentioned in the problem description, the objective of our algorithm is to select a reciprocal anonymity set according to user's privacy protection requirements. In this section, the goal is to select k − 1 POIs whose query frequencies are similar to the POI corresponding to the user's real location, and ensure selected POIs can be divided into at least l semantic categories. The anonymity set would be constructed based on the classification of the POI corresponding to the user's real location, and POI corresponding to this classification is already determined. Therefore, even if attacker submits malicious query in replay attack, the user's real semantic information cannot be inferred.

The brief steps of the algorithm are as follows: First, select POI that satisfies the strong association to form POI sets V_{left}, and compute the size of V_{left} as sum. Second, divide V_{left} into b buckets based on query frequency from high to low, $b = \lfloor sum/k \rfloor$. In order to ensure that each bucket contains k POIs and l semantics, the ideal state is that each bucket contains exactly k POIs, but there is a possibility that the distribution of POIs will be uneven. That is, there are still some POIs left after classification, and $num_{remain}(num_{remain} = sum - k * b)$ is greater than 0, so the algorithm allocates a POI from the remain POIs to the previous num_{remain} bucket until the last POI is allocated. The time complexity of the algorithm described in algorithm 3 is $O(n*log(n))$, where n is the number of POI sets V_{left}.

Assume that the road network of the user's real location is shown in Figure 8. The road network includes 6 POI categories, namely station, school, office building, bar, restaurant, and hotel. There is sum = 30 POIs in the entire grid. In an example, the triplet $< K, L, H >$ set by the user is $< 10, 4, 3 >$. Based on the protection coefficient k = 10, l = 4, the algorithm divides the entire grid into b = 3 buckets. To ensure that each bucket contains 10 POIs and at least 4 POI categories, these POIs are sorted based on query frequency, and then sequentially divided into the buckets until all POIs are allocated. This result is shown in Figure 8, coloring some POIs with same color represents that those POIs are divided into the same bucket. In this case, the POIs in the bucket containing the POI corresponding to the user's real location are {a1, b1, c1, d1, e1, f1, g1, h1, I1, j1}.

Algorithm 3 Classifying POIs into Buckets

Input: Set of POIs $\mathbf{V}_{\mathbf{left}}$, Anonymity degree **k**, Semantic degree **l**, Query frequency table **FTable**
Output: Classified **buckets**
1. Sort $\mathbf{V}_{\mathbf{left}}$ by **FTable**
2. for v in $\mathbf{V}_{\mathbf{left}}$:
3. **if** len(bucket) < k **and** the semantic category bCatg of bucket < **l**:
4. **if** v.catg is **not in** bucket.bCatg **then**:
5. **add** v **into** bucket
6. **else:**
7. **continue**
8. **else if** len(bucket) < k **and** the semantic category bCatg **of** bucket >= **l**:
9. **add** v **into** bucket
10. **else**:
11. **add** bucket **into buckets**
12. bucket = null
13. return buckets

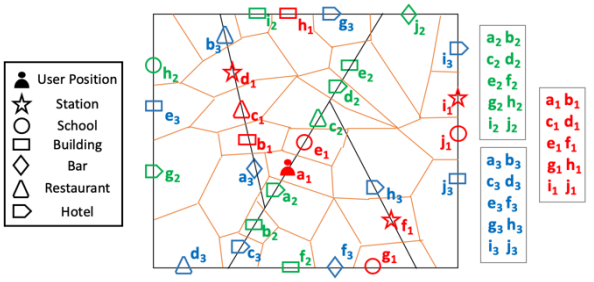


FIGURE 8. An example of classifying POIs into buckets.

E. FAKE POSITION SELECTION

In order to ensure that the constructed anonymity set further satisfies the reciprocity, we select the fake position in the other k − 1 VC based on the road segment number where

$$T(SPOI_i, SPOI_j, t) = Fnum(SPOI_i, SPOI_j, t))/\sum_{l=1}^{n} Fnum(SPOI_i, SPOI_l, t)) \quad (1)$$

$$AR(SPOI_i, SPOI_j, t) = \begin{cases} strong\ association\ relationship\ if\ T(SPOI_i, SPOI_j, t) > minconf \\ weak\ association\ relationship\ if\ T(SPOI_i, SPOI_j, t) < minconf \end{cases}$$
$$i = 1, 2, \ldots, \quad n; j = 1, 2, \ldots, n; i \neq j \quad (2)$$

Algorithm 4 Selecting Fake Positions

Input: Grid POIs set $\mathbf{V_{left}}$
Output: Anonymity set **CR**
1. Get R_{max} **from** $\mathbf{V_{left}}$
2. for v **in** $\mathbf{V_{left}}$:
3. **ranking** v roads by clockwise
4. **if** the R_v of v $<$ R_{max}:
5. **for** r in R_v:
6. the roads are **divided** into two sections until $R_v == R_{max}$
7. for v in $\mathbf{V_{left}}$:
8. **select** the new query position qpos in v as fake position
9. **add** qpos into **CR**
10. return CR

the real user location is located. The fake position is the latest historical query point of the current road segment, and the fake position combines with real user location to form anonymity set CR.

The algorithm will mark the *VC* road segments in clockwise order. If the user's true location is on the x-road, then select the fake position in the x-road of the other *VC* in the same bucket. The fake position is the latest historical query point of the current road. Different VC have different number of road segments, and in order to avoid the absence of x road segments, we divide the number of road segments in the bucket into R_{max} which is the maximum number of VC's road segments in the bucket. $R_{max} = Max\,(R_i)\,, i = 1, 2, \ldots, k$, R_i is the number of road segments corresponding to the i-th *VC*. The number of road segments that need to be divided for each VC is $R_{max} - R_i$. The algorithm divides the road into two according to the road segments number, until the number is R_{max}. The time complexity of the algorithm is $O(n * R_{max})$, where n is the number of POI sets V_{left}, and the CR generation algorithm is described in algorithm 4.

V. SECURITY ANALYSIS

A. RESISTANCE AGAINST REPLAY ATTACK

Theorem 1: TSR can resist replay attack.

Proof: In order to prove that TSR algorithm can resist replay attack, it needs to be proved that no matter how many times the user performs a query at a certain location, the generated anonymity set is always invariant in another k-1 fake locations, namely, the generated anonymity set satisfies reciprocity. Assume that the anonymous algorithm proposed in this paper is A (*), and the anonymous server generates an anonymity set $\boldsymbol{CR} = \{\boldsymbol{a_1}, \boldsymbol{a_1}, \ldots, \boldsymbol{a_k}$ using the anonymous algorithm A(*). Run the algorithm A(*) for each fake position $\boldsymbol{a_i}(\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{k})$ in the $\boldsymbol{CR}$, generate a new anonymity set $\boldsymbol{CR'}$, and record the query frequency of the POI where the user is located. In the T-SR anonymous algorithm, the algorithm classifies the POI according to the query frequency and the semantic category, and the k-invariant POIs that satisfy the privacy protection triple and whom the query frequency is close are always divided into the same bucket and the fake positions are selected from the same road segment. Therefore, in a certain period of time, no matter how many queries a user requests, the anonymity set constructed by the algorithm is always satisfies the reciprocity, namely, $\boldsymbol{CR'} = \boldsymbol{CR} = \{\boldsymbol{a_1}, \boldsymbol{a_1}, \ldots, \boldsymbol{a_k}$. Therefore, our proposed solution can resist replay attack.

B. RESISTANCE AGAINST INFERENCE ATTACK

Theorem 2: TSR can resist inference attack.

Proof: In order to prove that the proposed anonymous algorithm can resist inference attack, it needs to be proved that the attacker has the relevant background knowledge of the road network, namely, the historical query frequency information of each POI in the road network and still can't distinguish the user's real location from k locations. In the T-SR anonymous algorithm, the POI corresponding to the fake location in the anonymous set has a relatively close query frequency, and attackers cannot filter out fake positions from background knowledge, and that is to say the probability of recognizing real position is no more than $\mathbf{1}/\boldsymbol{k}$. Therefore, our proposed solution can resist inference attack.

C. RESISTANCE AGAINST TEMPORAL ASSOCIATION ATTACK

Theorem 3: TSR can resist against temporal association attack.

Proof: In order to prove that the proposed anonymous algorithm can resist the time semantic association attack, it is necessary to explain that the attacker has the relevant background knowledge, that is, the association relationship between the POI categories in the road network, and TSR algorithm still guarantees that the attacker cannot identify the user's location semantic information through query semantics. Assume that the current query semantic is ***quloc***, the attacker not only can steal the complete anonymity set CR, but also obtain the association relationship between POI categories, that is, attacker can infer the probability that POI semantic corresponding to the fake location request the current query semantic. However, in the anonymous algorithm of this paper, the POI semantic corresponding to the fake location in CR is strongly related to the query semantic. Therefore, the attacker can't infer the real position semantic information through the semantic association relationship, namely, the probability that the attacker infers the true positional semantic is no more than $1/l$.

VI. EXPERIMENT ANALYSIS

A. OVERVIEW OF THE METHOD EXPERIMENTAL ENVIRONMENT AND DATA

The experiment was written in Python, running in **Interl** $(\boldsymbol{R})$ ***Core*** $(\boldsymbol{TM})$ **i5** $-$ **4590 CPU** with 8GB of 64-bit Windows 10 operating system. This paper uses road network data from Oldenburg, Germany, including 6,105 vertices and 7,035 roads. According to the distribution of POIs in

TABLE 3. Experimental data set parameters.

Parameter	Number
Number of nodes	6105
Number of edges	7035
Number of POIs every road	4-9
Number of query frequency	2500000

the data of FourSquare, the algorithm randomly generates 4-9 categories (23 categories in total) of positional semantic information on each road in the original road network of Oldenburg. By default, 2.5 million query frequencies are used to simulate the query request initiated from different users, and some parameters of the data set are shown in Table 3.

B. EXPERIMENTAL DEVELOPMENT AND METRICS

The experiments in this paper includes: First, compare the difference in $\mathbf{Q^2}$ (pseudo-variance) and **APD** (average path distance) between the T-SR algorithm, the HilbertCurve algorithm, the MinDis algorithm and the Random algorithm through three sets of contrast experiments. In this experiment, we focus on the discriminative degree between the true position and the fake position constructed by each algorithm. The experiment shows that show that the T-SR algorithm can effectively resist replay attack and inference attack. The specific formula of each metric is described as follows. Second, compare the number of weak incidence relation in the anonymity set constructed by each algorithm. In this experiment, the number of weak incidence relation in the anonymity set constructed by T-SR algorithm is much smaller than the other algorithms. The experiment shows that T-SR algorithm can effectively resist temporal association attack. Third, compare query time in the anonymity set constructed by T-SR algorithm when setting different privacy protection requirements. Among these four types of algorithms, HilbertCurve is a variant of the X-star anonymous algorithm. The idea of the algorithm is to construct a bucket based on the query frequency and select a reciprocal anonymity set in the bucket, but it does not take the incidence relation and semantic difference of the POIs into account. The MinDis algorithm is an algorithm based on Euclidean distance to select the anonymity set. The constructed anonymity set is smaller and more vulnerable to regional attack. Random algorithm is an algorithm for randomly selecting fake positions on the road network. T-SR is the algorithm proposed in this paper. It is a kind of algorithm that considers the difference of query frequency, POIs incidence relation and reciprocity, and can satisfy the user's privacy protection requirements triplet $< \boldsymbol{K}, \boldsymbol{L}, \boldsymbol{H} >$.

$$Q^2 = \frac{\sum_{i=1}^{k-1} (P_{ij} - P_{uj})^2}{(k-1) * l} \tag{3}$$

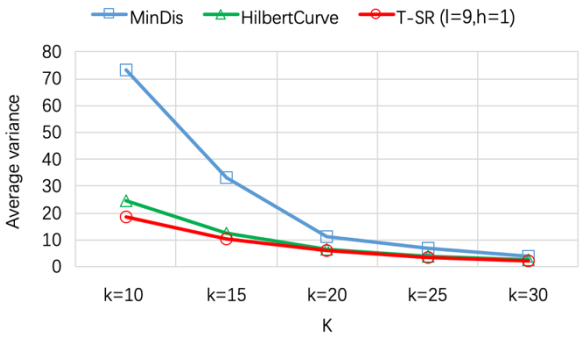


FIGURE 9. Comparison of algorithmic pseudo-variance.

$$APD = \frac{\sum_{i=1}^{k-1} dis(uloc, ucr)}{k-1} \tag{4}$$

$\boldsymbol{P_{uj}}$ is the query frequency of the POI corresponding to the user's real location, $\boldsymbol{P_{ij}}$ is the query frequency of the POI corresponding to the i-th fake location, $\boldsymbol{dis(uloc, ucr)}$ is the distance between the user's location and the fake location, k is the K-anonymity coefficient, and l is the L-disparity coefficient.

C. EXPERIMENTAL RESULTS

1) COMPARISON OF ALGORITHMIC PSEUDO-VARIANCE

Q^2 reflects the degree of deviation between the fake position and the true position. The smaller Q^2 is, the higher the uncertainty of the anonymity set and the more realistic the fake position is. Figure 9 reflects the difference in Q^2 of the three algorithms when value k is different. It can be seen that regardless of the value of k, T-SR algorithm is always better than the other two algorithms in Q^2. Therefore, the fake position selected by the T-SR algorithm is more realistic and T-SR algorithm has better performance for resisting replay attack and inference attack than the other two algorithms. When k is set to 10, Q^2 of the T-SR algorithm and the HilbertCurve algorithm is much smaller than that of the MinDis algorithm. This is because the two algorithms take the difference of the query frequency of the POIs into account when selecting the anonymity set. As the value of k increases, it is worth noting that the difference of the three algorithms in Q^2 becomes smaller and smaller. This is because when the privacy protection level h is fixed, the grid area for constructing anonymity set is fixed, and the candidate set for fake position is also fixed, so with the increase of the coefficient k, the three algorithms will select more identical fake positions in the anonymity set, and the difference of Q^2 between several algorithms gets smaller.

2) COMPARISON OF ALGORITHMIC APD

APD is used to measure the average road network distance from the fake position to the real position. The larger the distance is, the more scattered the fake position is, and the more difficult it is for the attacker to obtain semantic information of real position from anonymity set. Figure 10 reflects the

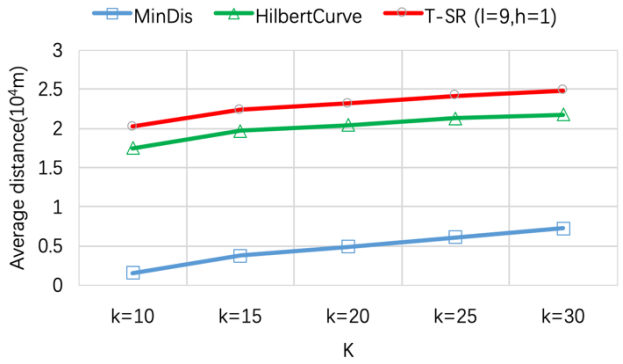


FIGURE 10. Comparison of algorithmic APD.

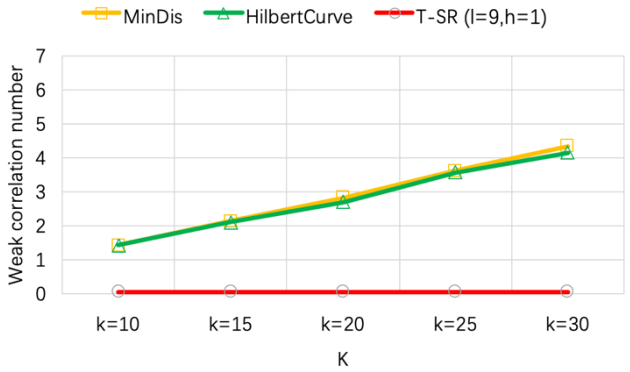


FIGURE 11. Comparison of algorithmic weak association relation numbers.

difference in ***APD*** when anonymous coefficient k is fixed. It can be seen that ***APD*** corresponding to the T-SR algorithm is larger than the other two algorithms, which makes the distribution of the fake position more dispersed. And as the value of k increases, the difference in ***APD*** between the three algorithms will become smaller and smaller. This is because the size of the grid area in our experiment where the user is located is determined. Therefore, as the value of k increases, our proposed T-SR algorithm will select more identical fake positions, so the difference of three algorithms in ***APD*** will be smaller.

3) COMPARISON OF ALGORITHMIC WEAK INCIDENCE RELATION NUMBERS

Weak incidence relation number refers to the number of weak incidence relation between the POI corresponding to the fake position and the query POI in the anonymity set. As shown in Figure 11, we respectively count the weak incidence relation number of MinDis algorithm, HilbertCurve algorithm and T-SR algorithm. The T-SR algorithm has filtered out the POIs that have weak incidence relation with the query POI before constructing anonymity set. Therefore, the weak incidence relation number of the T-SR algorithm is 0, so it is much smaller than the other two algorithms, and T-SR algorithm can resist the temporal association attack better. The weak incidence relation number of the HilbertCurve

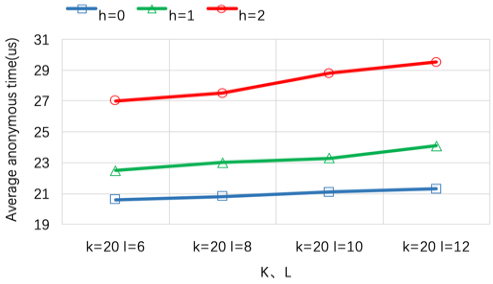


FIGURE 12. Difference in query time when K is determined.

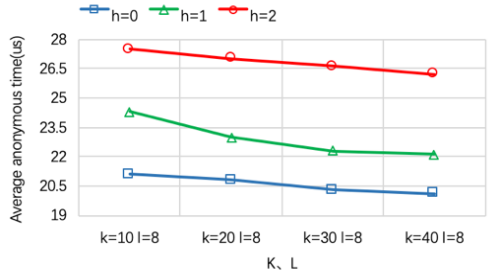


FIGURE 13. Difference in query time when L is determined.

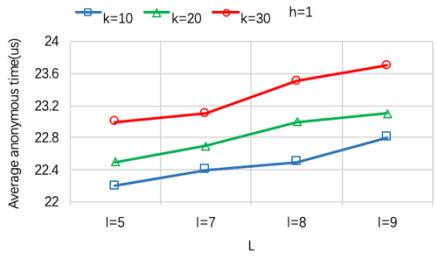


FIGURE 14. Difference in query time when H is determined.

algorithm is slightly smaller than that of MinDis, because the query frequency of POIs with the same semantic category is close and the POIs with the same semantic category generate strong association by default.

4) COMPARISON OF DIFFERENCES BETWEEN TRIADS AND DIFFERENT ALGORITHMS

Figure 12-14 shows the difference in the query time for the T-SR algorithm when setting up different privacy protection requirements triples. As shown in Figure 12, when the k-anonymity coefficient is fixed, as l and h increase, the area of the anonymity set selection and the number of POI categories increase, so the query time increases. As shown in Figure 14, with the increase of K and L, the query time increases proportionally when the privacy protection level H is fixed. On the contrary, in Figure 13, when the semantic coefficient L is fixed, the query time to satisfy the l-diversity in a bucket is constant. Therefore, the larger k is, the faster the algorithm is when POIs are classified into different buckets, and the less query time is consumed when privacy protection level h is same.

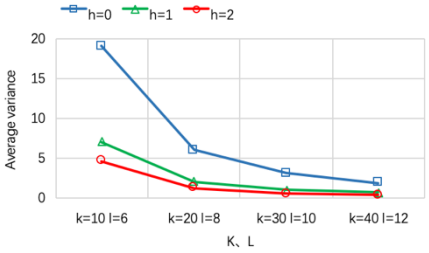


FIGURE 15. Difference in pseudo-variance when H is determined.

Figure 15 shows the difference in Q^2 for T-SR algorithm when setting up different privacy protection requirements triples. When the triplet $< K, L, H >$ parameter increases, it can be seen that the grid area selected by the anonymity set increases, so it is more likely to select the fake positions whose query frequency is close to the true position, and Q^2 is correspondingly reduced.

VII. CONCLUSION

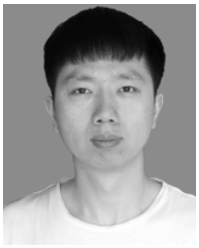
In this paper, we propose a location privacy protection T-SR algorithm for POI query in road network environment. The algorithm takes the temporal association relationship between semantics into account, defines the triplet of privacy protection requirements, and considers the historical query frequency of POIs and the reciprocity of anonymity set. Therefore, the algorithm makes it difficult for an attacker to distinguish the real location from the constructed anonymity set, makes the distribution of the fake position more dispersed, so as to resist replay attack, inferred attack and temporal association attack model. Experiments show that the proposed algorithm has obvious advantages in pseudo-variance, entropy, *APD* and the number of weak incidence relation, and the algorithm has good scalability which can effectively protect user location privacy. In the future work, we will take the strength and weakness of semantic categories into account from a multi-dimensional perspective to protect user location privacy more accurately.

REFERENCES

[1] C. L. Lawson, "Software for C^1 surface interpolation," in *Mathematical Software*. New York, NY, USA: Academic, 1977, pp. 161–194.
[2] C. G. Ma, C.-L. Zhou, S.-T. Yang, and Y.-L. Zhao, "Location privacy-preserving method in LBS based on Voronoi division," *J. Commun.*, vol. 36, pp. 5–16, May 2015.
[3] S. Deng, H. Wu, W. Tan, Z. Xiang, and Z. Wu, "Mobile service selection for composition: An energy consumption perspective," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1478–1490, Jul. 2017.
[4] R. Lu, X. Lin, Z. Shi, and J. Shao, "PLAM: A privacy-preserving framework for local-area mobile social networks," in *Proc. INFOCOM*, 2014, pp. 763–771.
[5] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2681–2691, Dec. 2015.
[6] L. Kuang *et al.*, "An improved privacy-preserving framework for location-based services based on double cloaking regions with supplementary information constraints," *Secur. Commun. Netw.*, vol. 2017, Nov. 2017, Art. no. 7495974. doi: 10.1155/2017/7495974.
[7] M. Kolahdouzan and C. Shahabi, "Voronoi-based K nearest neighbor search for spatial network databases," in *Proc. 30th Int. Conf. Very Large Data Bases*, vol. 30, 2004, pp. 840–851.
[8] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. 1st Int. Conf. Mobile Syst., Appl. Services*, 2003, pp. 31–42.
[9] C.-Y. Chow, M. F. Mokbel, and W. G. Aref, "Casper*: Query processing for location services without compromising privacy," *ACM Trans. Database Syst.*, vol. 34, no. 4, 2009, Art. no. 24.
[10] S. Deng, L. Huang, J. Taheri, J. Yin, M. Zhou, and A. Y. Zomaya, "Mobility-aware service composition in mobile communities," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 3, pp. 555–568, Mar. 2017.
[11] G. Sun, D. Liao, H. Li, H. Yu, and V. Chang, "L2P2: A location-label based approach for privacy preserving in LBS," *Future Gener. Comput. Syst.*, vol. 74, pp. 375–384, Sep. 2016.
[12] T. C. Li and W. T. Zhu, "Protecting user anonymity in location-based services with fragmented cloaking region," in *Proc. IEEE Int. Conf. Comput. Sci. Automat. Eng. (CSAE)*, vol. 3, May 2012, pp. 227–231.
[13] T. Hashem, L. Kulik, and R. Zhang, "Countering overlapping rectangle privacy attack for moving *k*NN queries," *Inf. Syst.*, vol. 38, no. 3, pp. 430–453, 2013.
[14] C. Li and B. Palanisamy, "De-anonymizable location cloaking for privacy-controlled mobile systems," in *Proc. Int. Conf. Netw. Syst. Secur.* Cham, Switzerland: Springer, 2015, pp. 449–458.
[15] A. Pingley, N. Zhang, X. Fu, H.-A. Choi, S. Subramaniam, and W. Zhao, "Protection of query privacy for continuous location based services," in *Proc. INFOCOM*, 2011, pp. 1710–1718.
[16] L. Kuang, Y. Zhu, S. Li, X. Yan, H. Yan, and S. Deng, "A privacy protection model of data publication based on game theory," *Secur. Commun. Netw.*, vol. 2018, Oct. 2018, Art. no. 3486529. doi: 10.1155/2018/3486529.
[17] M. Guo, N. Pissinou, and S. S. Iyengar, "Pseudonym-based anonymity zone generation for mobile service with strong adversary model," in *Proc. 12th Annu. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2015, pp. 335–340.
[18] K. Dou, B. Guo, and L. Kuang, "A privacy-preserving multimedia recommendation in the context of social network based on weighted noise injection," *Multimedia Tools Appl.*, pp. 1–20, Jan. 2017. doi: 10.1007/s11042-017-4352-3.
[19] Y. Yin, L. Chen, Y. Xu, and J. Wan, "Location-aware service recommendation with enhanced probabilistic matrix factorization," *IEEE Access*, vol. 6, pp. 62815–62825, 2018.
[20] Z. Huo, X. Meng, H. Hu, and Y. Huang, "You can walk alone: Trajectory privacy-preserving through significant stays protection," in *Proc. Int. Conf. Database Syst. Adv. Appl.* Berlin, Germany: Springer, 2012, pp. 351–366.
[21] S. Gao, J. Ma, C. Sun, and X. Li, "Balancing trajectory privacy and data utility using a personalized anonymization model," *J. Netw. Comput. Appl.*, vol. 38, pp. 125–134, Feb. 2014.
[22] S. Deng *et al.*, "Toward risk reduction for mobile service composition," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1807–1816, Aug. 2016.
[23] X. Pan, X. Meng, and J. Xu, "Distortion-based anonymity for continuous queries in location-based mobile services," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2009, pp. 256–265.
[24] S. Deng, Z. Xiang, J. Yin, J. Taheri, and A. Y. Zomaya, "Composition-driven IoT service provisioning in distributed edges," *IEEE Access*, vol. 6, pp. 54258–54269, 2018.
[25] X. Pan, J. Xu, and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1506–1519, Aug. 2012.
[26] A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," in *Advances in Spatial and Temporal Databases*. Boston, MA, USA: SSTD, 2007.
[27] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 12, pp. 1719–1733, Dec. 2007.
[28] Y. Yin, F. Yu, Y. Xu, L. Yu, and J. Mu, "Network location-aware service recommendation with random walk in cyber-physical systems," *Sensors*, vol. 17, no. 9, p. 2059, 2017.
[29] S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbor search with strong location privacy," *Proc. VLDB Endowment*, vol. 3, nos. 1–2, pp. 619–629, 2010.
[30] T. Peng, Q. Liu, and G. Wang, "Enhanced location privacy preserving scheme in location-based services," *IEEE Syst. J.*, vol. 11, no. 1, pp. 219–230, Mar. 2017.
[31] S. Deng, L. Huang, D. Hu, J. L. Zhao, and Z. Wu, "Mobility-enabled service selection for composite services," *IEEE Trans. Services Comput.*, vol. 9, no. 3, pp. 394–407, May/Jun. 2016.

[32] H. C. Liang, B. Wang, N. N. Cui, K. Yang, and X. C. Yang, "Privacy preserving method for point-of-interest query on road network," (in Chinese), *Ruan Jian Xue Bao/J. Softw.*, vol. 29, no. 3, pp. 703–720, 2018. [Online]. Available: http://www.jos.org.cn/1000-9825/5451.htm
[33] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz, "On the semantic annotation of places in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 520–528.
[34] F. Tian, X.-L. Gui, X.-J. Zhang, J.-W. Yang, P. Yang, and S. Yu, "Privacy-Preserving approach for outsourced spatial data based on POI distribution," *Chin. J. Comput.*, vol. 37, no. 1, pp. 123–138, 2014.
[35] B. Bamba, L. Liu, P. Pesti, and T. Wang, "Supporting anonymous location queries in mobile environments with privacygrid," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 237–246.
[36] K. Mouratidis and M. L. Yiu, "Anonymous query processing in road networks," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 1, pp. 2–15, Jan. 2010.
[37] X. J. Zhang, X. L. Gui, and Z. D. Wu, "Privacy preservation for location-based services: A survey," (in Chinese), *Ruan Jian Xue Bao/ J. Softw.*, vol. 26, no. 9, pp. 2373–2395, 2015. [Online]. Available: http://www.jos.org.cn/1000-9825/4857.htm
[38] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Hoboken, NJ, USA: Wiley, 2009.
[39] E. Yigitoglu, M. L. Damiani, O. Abul, and C. Silvestri, "Privacy-preserving sharing of sensitive semantic locations under road-network constraints," in *Proc. IEEE 13th Int. Conf. Mobile Data Manage.*, Jul. 2012, pp. 186–195.
[40] B. Palanisamy and L. Liu, "Attack-resilient mix-zones over road networks: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 495–508, Mar. 2015.
[41] Y. K. Kim, A. Hossain, A.-A. Hossain, and J.-W. Chang, "Hilbert-order based spatial cloaking algorithm in road network," *Concurrency Comput., Pract. Exper.*, vol. 25, no. 1, pp. 143–158, 2013.

LI KUANG received the Ph.D. degree in computer science from Zhejiang University, China, in 2009. She is currently a Professor with the School of Computer Science and Engineering, Central South University. Her research interests include service computing, mobile computing, and privacy preserving.



SHUAI HE received the B.S. degree from Changsha University, China. He is currently pursuing the master's degree with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include location privacy, mobile computing, and machine learning.

YUYOU FAN is currently a Senior Student with the School of Computer Science and Engineering, Central South University, Changsha, China. He is interested in programming language and algorithms.

HUAN ZHANG received the B.Sc. degree from Central South University, Changsha, China, in 2018, where she is currently pursuing the master's degree with the School of Computer Science and Engineering. Her research interests include software ecosystems, mobile computing, and service computing.

RUYI SHI received the B.Sc. degree from Central South University, Changsha, China, in 2018, where she is currently pursuing the master's degree with the School of Computer Science and Engineering. Her research interests include services computing and crowd computing.