Privacy Protection Scheme of POI Query Based on Semantic and Temporal Association

Kaizhong Zuo, Jun Zhao, Peng Hu, Zhangyi Shen, and Xixi Chu
(Corresponding author: Peng Hu)

School of Computer and Information, Anhui Normal University, Wuhu 241002, China
Anhui Provincial Key Laboratory of Network and Information Security, Wuhu 241002, China
Email: hupeng@ahnu.edu.cn

(Received Feb. 16, 2022; Revised and Accepted June 28, 2022; First Online July 3, 2022)

Abstract

In a digital society, location-based service (LBS) is a necessary function for mobile devices. However, while enjoying the convenience, users also face the risk of privacy disclosure. A novel location privacy protection scheme based on semantic and temporal association is proposed in this paper to solve this problem. Specifically, this scheme establishes a temporal association model according to users' historical query requests to measure the temporal association relationship. Then, the scheme constructs an optimal association table at the current moment based on association entropy. Finally, based on users' personalized privacy requirements and the optimal association table, the scheme builds a dummy location set distributed on different road segments. In addition, we have applied this scheme to the Android platform to implement a novel location privacy protection app called QueryWithMe. Theoretical analysis and experimental results show that the scheme can resist inference and temporal association attacks.

Keywords: Location Privacy; Mobile Application; POI Query; Temporal Association

1 Introduction

With the rapid development of mobile internet, LBS has been widely used. Among them, the POI query has always been a widely used service. The POI query refers to certain unique buildings on the electronic map, such as restaurants, hospitals, and schools [10]. Typical POI queries include "K hospitals closest to me." and "restaurants near me.".

However, LBS not only brings convenience to people, but may also cause privacy disclosure-related problems at the same time [19]. For example, college student Alice has lunch with her friends every weekend, so every time she queries the nearby restaurants in her dormitory at around 10:00 am on Saturday or Sunday. By intercepting and analyzing the content of Alice's requests, attackers can infer her living habits (going out for lunch every weekend), occupation (student), and even dietary preferences (she often queries western restaurants and may like western food). Therefore, the purpose of our scheme is to protect the privacy of users' locations while using POI query service in real life.

In recent years, some researchers have proposed many location privacy protection schemes used in POI query. However, there are still attack vulnerabilities in existing works:

1) The attackers can obtain the relevant supplementary information from the central anonymous server, such as historical query data, road network, and location semantics. For example, assuming that attackers can intercept the anonymous set, which only contains a single semantic information "hospital". It is possible to infer that the user's physical condition is poor. If attackers analyze the historical queries of the anonymous set and calculate the historical query frequency of these locations, they can also easily filter out some unrealistic dummy locations. Furthermore, if attackers have information about road network topology, they can find that most of the locations are located in the hospital area, so they can learn that the user's current location is the hospital;

2) The attackers can utilize temporal association relationship between the semantics of query content and locations. For example, around 6:00 pm, Bob wants to invite colleagues to dinner when he is at a company, so he requests POI query service. It is known that the information submitted by Bob to the anonymous server is $(id_{carol}, loc, restaurant)$. If attackers intercept the dummy locations anonymous set $\{loc, dummy_1, dummy_2, dummy_3, dummy_4\}$, generated by the anonymous server and submitted to the location-based service provider (LSP). The location semantics corresponding to the anonymous set are "company", "park", "primary school", "restaurant", "bar". At around 6:00 pm, there is a weak temporal

association between "primary school" and "restaurant" and between "park" and "restaurant". Thus, they can filter out dummy locations $dummy_1$ and $dummy_2$ with a higher probability, and the probability of revealing the user's location increases from $\frac{1}{5}$ to $\frac{1}{3}$.

To overcome the above privacy security issues in road network environment, we proposed a privacy protection scheme based on semantic and temporal association for POI query. First, this scheme measures the temporal association between query content semantics and location semantics by establishing a temporal association model. Second, the scheme constructs a optimal association table at the current moment by defining association entropy. Finally, according to the user's personalized privacy requirements and the optimal association table, the scheme selects dummy locations with similar historical query frequencies distributed on different road segments and adds them to the anonymous set. To verify the effectiveness and practicability of the scheme, we implemented a location privacy protection application called QueryWithMe on Android platform, which can protect users' location privacy when querying POI based on our proposed scheme.

When a user requests POI query, the application automatically selects decoy query on the dummy locations to confuse the user's location and query intention. In this way, although attackers obtained multiple queries from the user, they could not confirm the real POI query. Thus, this app can protect the privacy of the user's location while requesting POI query service. Specifically, the contribution of this paper mainly includes four aspects:

1) The temporal association model that we propose can measure temporal association relationship. Through this model, the temporal association probability between the semantics of query content and location semantics can be calculated. Then, we define an association entropy to calculate the indistinguishability of temporal association relationship;

2) We propose a method that selects dummy locations according to personalized privacy requirements. By formalizing user-defined privacy requirements, we propose a method that selects dummy locations on different road segments;

3) Combining with the proposed scheme, we design and implement QueryWithMe, an application that protects the location privacy of mobile users during POI query in real-time;

4) Through theoretical analysis, we conduct a lot of experiments, and the results show that our scheme can defend against temporal association attack and inference attack. The tests and analysis of QueryWithMe show that the proposed scheme has good practicability and application value.

The structure of the rest paper is as follows. In Section 2, we introduce the location privacy protection of LBS by related researchers in recent years. In Section 3, we give the system framework, attack model, some basic definitions, and temporal association model. The details of the algorithm, experimental analysis, and practical application are presented in Section 4 and Section 5 respectively. Finally, conclusions and future work are given in Section 6.

2 Related Works

Many methods have been proposed to solve the problem of location privacy disclosure. Among them, the most commonly used schemes for location privacy protection include selecting dummy locations [15, 21, 22] and generating cloaking regions [16, 20].

Initially, the location K-anonymity was proposed by Gruteser *et al.* [8] On this basis, considering the query frequency calculated from user history query records and the over-concentrated distribution of dummy locations, literature [15] selected dummy locations based on the entropy metric. Zhao *et al.* [21] proposed a novel mechanism for generating dummy locations based on Super Concept-based Distance. Alotaibi *et al.* [1] designed a user-based location selection scheme that uses dummy locations to hide users' real location, where the selection of dummy location is based on the existing location of users. The scheme has better performance in terms of entropy and cloaking region. Yang *et al.* [17] proposed a location privacy protection scheme based on Q-Tree storage, which can protect user's location and query privacy by selecting POI with a higher query probability as the query content of anonymous location units.

The above work is mainly applicable to Euclidean space, but the location privacy protection scheme in the road network environment is closer to real life and has received full attention. Taking into account the personalized privacy need, Chen *et al.* [6] proposed a road network location privacy protection method based on location semantics, which generates an anonymous road segment set to prevent the disclosure of sensitive information caused by location semantics. On this basis, literature [12] calculates the privacy of adjacent road segments according to user-defined sensitivity and selects a set of road segments as candidate road segments. Considering the distribution of the semantic location of road segments, Liu *et al.* [11] proposed a privacy protection method based on the similarity of semantic distribution. In the location privacy protection of continuous query, Wang *et al.* [16] introduced semantic sensitivity into the Voronoi unit. If a user is in a sensitive semantic position, an anonymous area that satisfies K-anonymity and L-sensitivity is constructed for the user. Otherwise, it is constructed to satisfy K-anonymity anonymous area while using the dynamic pseudonym mechanism to update user identification information.

Aiming at the fact that attackers can infer real location by filtering out the semantically incompatible locations in the cloaking area, Bostanipour *et al.* [4] proposed a joint obfuscation method for location obfuscation and provided a new framework to evaluate the joint and non-joint methods. To resist the attack of semantic inference, Ma *et al.* [13] introduced the concept of (θ, δ)-diversity, and calculated the semantic similarity between real location and other locations through Earth Mover Distance.

In recent years, privacy protection schemes based on Android platform have gradually attracted people's attention. Aiming at the problem of location privacy disclosure caused by Android security architecture defects, Amukelani *et al.* [2] proposed a security model that protects against harmful applications for application permissions. Zhang *et al.* [18] proposed a new location privacy protection mechanism called ShiftRoute for smartphone map services, and applied it to Android platform to track users. Hu *et al.* [9] proposed SAMLDroid, an Android application recognition method that combines static code analysis and machine learning. First, this method uses static analysis to review the source code of the app to get the location. Second, it uses classifiers and integrates multiple application functions to dynamically analyze patterns. Finally, the experiments show that SAMLDroid has a higher accuracy rate. Recently, to protect the location privacy of android users in offline mode, Arshad *et al.* [3] proposed a scheme which protects the actual location coordinates by keeping the simulated location within a seemingly realistic span radius and automatically adjusts the location coordinates. Nieminen *et al.* [14] proposed a practical privacy protection scheme based on secure two-party calculations, and the experiments show that the scheme has high feasibility in certain types of indoor positioning applications.

3 Preliminaries

3.1 System Framework

This paper adopts the framework of the central anonymous server. The central anonymous server can reduce the system overhead of mobile clients, improve the quality of service, and prevent untrusted LSPs from maliciously obtaining user privacy information.

Figure 1 shows the system framework of our scheme. One POI query includes the following four steps:

1) Users use smart devices to send query request Qu;

2) The central anonymous server receives the POI query from the querying user and constructs the dummy locations set AS according to user-defined privacy requirements. After that, the central anonymous server sends an anonymous POI query request Qc to the LSP;

3) The LSP obtains the query result set RS by searching its location database and returns it to the central anonymous server;

4) The central anonymous server filters the accurate query result from RS and sends it to the querying user.

Figure 1: System framework

3.2 Attack Model

For the above system framework, we assume that users who use POI query service suffer two malicious attacks:

1) Active attackers usually target the communication between the central anonymous server and the LBS server as the main target;

2) The central anonymous server stores some database information, including road network topology, semantic information, etc. Once it is successfully obtained by passive attackers, it will be used as supplementary information to infer user privacy.

Defending against temporal correlation attacks is the goal of our scheme. At the same time, our scheme also needs to prevent attackers from reasoning about the location information of users. Specifically, these two attack models are described below.

1) Inference attack: Through calculating historical query frequency, attackers filter out some locations that are impossible to query and locations with low query frequency. It increases the possibility that the user's real location is leaked. Meanwhile, attackers can easily obtain users' real location by analyzing the location semantics and road network topology;

2) Temporal association attack: The location semantics in dummy anonymous set may change significantly the probability of accessing specific query content semantics over time. Taking advantage of this difference, attackers can easily filter out some weak or unrelated dummy locations to improve the probability of obtaining location privacy.

3.3 Related Definition

In order to better resist inference attack and temporal association attack during POI query, we propose a location privacy protection scheme based on semantic and temporal association. The following work mainly introduces the relevant definitions in the scheme:

Definition 1. POI query. *The POI query can be represented by tuple* $Qu = \langle uid, t, uloc_t, qs_t \rangle$, *where uid is the unique identifier when a user requests POI query service at time t,* $uloc_t$ *is the user's location, and the semantic of query content is represented by* qs_t.

The central anonymous server constructs dummy anonymous set AS to hide the user's location, where AS contains the user's location and $k-1$ dummy locations selected by anonymous algorithm, then sends anonymous query $Qc = \langle uid, t, AS, qs_t \rangle$ to the LSP.

Definition 2. User personalized privacy requirement. *User's personalized privacy requirement PR is expressed in tuple ¡k,l,s¿. Specifically, the definition of each attribute in this tuple is introduced below:*

1) *k is the level of privacy protection that can be set by users. It means that, in the absence of other additional knowledge, the probability that attackers will infer users' location is no more than* $\frac{1}{k}$;

2) *l is the number of road segments that users privacy requirements with personalized, which means that the dummy locations are at least distributed on l road segments to avoid a single road segment attack caused by excessively concentrated dummy locations;*

3) *s is the number of semantic types of users' privacy requirements with personalized privacy. That is, the dummy locations set construct in our scheme needs to contain at least s semantics to prevent attackers from inferring users' location from location semantics.*

3.4 Temporal Association Model

When people use POI query service in daily life, they are more inclined to reach the query place. It can be said that the location of a user is related to query content. Most of these associations show different strengths and weaknesses over time. People are more accustomed to querying nearby hotels at a train station at 3:00 am and querying nearby restaurants at the company at 5:00 pm. If attackers use this temporal association to filter dummy locations, it will significantly increase the risk of user location disclosure.

This paper proposes a temporal association model to better measure the temporal association between query content semantics and location. First, the temporal association sequence is constructed by analyzing historical query data. Second, the temporal association digraph is constructed according to temporal association sequence. Finally, the temporal association probability between query content semantics and location semantics is calculated according to out-degree situation. The relevant definitions of the temporal association model are as follows:

Definition 3. Semantic attribute information. *The semantic attribute information of location* $uloc_t$ *in user's POI query is location semantics, denoted as* $Suloc_t$, *and the semantic attribute information of the user's query content is query content semantics, denoted as* qs_t.

Definition 4. Temporal association sequence. *Given an interval [Ta,Tb], two adjacent queries are* $Qa = \langle uid, ta, uloc_{ta}, qs_{ta} \rangle$ *and* $Qb = \langle uid, tb, uloc_{tb}, qs_{tb} \rangle$, *if location semantics* $Suloc_{tb}$ *is same as query content semantics* qs_{ta}, *it is said that there is a temporal association between* qs_{ta} *and* $Suloc_{tb}$ *in time period [Ta,Tb], denoted as* $Suloc_{ta} \rightarrow qs_{ta}$. *We take every hour as a time interval.*

Definition 5. Temporal association frequency. *The temporal association frequency represents the numbers of temporal associations of user's* $Suloc_n \rightarrow qs_n$ *in the time interval [Ta,Tb], denoted as* Nf_n, *where n is the number of temporal association. Since there may be different temporal associations and their corresponding temporal association frequencies in this time period,* $Suloc_n \rightarrow qs_n$ *is only denoted as a certain one temporal association in this time period.*

Definition 6. Temporal association sequence. *The* $CT=\{\langle T, Suloc_1 \rightarrow qs_1, Nf_1 \rangle, \langle T, Suloc_2 \rightarrow qs_2, Nf_2 \rangle, \ldots, \langle T, Suloc_n \rightarrow qs_n, Nf_n \rangle\}$ *is denoted one sequence of temporal association sequences, where T is the number of time interval and* $T \in \{0, 1, \ldots, 23\}$, $Nf_i\,(i = 1, 2, \ldots, n)$ *is the number of times that the users have temporal association* $Suloc_i \rightarrow qs_i\,(i = 1, 2, \ldots, n)$.

Definition 7. Temporal association oriented graph. *The temporal association oriented graph is Gt=(V,E), which contains a set of vertices V and edges E. Each vertex* $v \in V$ *represents semantic type in temporal association sequence, and each edge* $e \in E$ *represents a temporal association. The out-degree of each vertex represents the sum* $\sum_{i=1}^{n} Nf_i$, *which is the number of the temporal association between semantics in the temporal association sequence.*

Definition 8. Temporal association probability. *The temporal association probability Pt between query content semantics and location semantics is calculated as follows: If the out-degree of the vertex is not 0, that is* $\sum_{i=1}^{n} Nf_i > 0$, *the probability of temporal association between location semantics and query content semantics is defined by Equation (1), otherwise* $\sum_{i=1}^{n} Nf_i = 0$, *then* $Pt = 0$.

$$Pt = \frac{Nf_i}{\sum_{i=1}^{n} Nf_i} \tag{1}$$

Where $i \in 1, 2, \ldots, n$.

Definition 9. Association entropy. *Given location set* $L = \{loc_1, loc_2, \ldots, loc_n\}$ *and query content semantics qs, If* $\bar{E}$ *represents temporal association probability between the location semantics in the location set and the query content semantics, then the association probability set is* $LQ = \{\bar{E}_1, \bar{E}_2, \ldots, \bar{E}_n\}$. *The calculation formula for the association entropy is as in Equation (2):*

$$E_{L \rightarrow qs} = -\sum_{i=1}^{n} E_i \cdot \log_2 E_i \tag{2}$$

where E_i *is the normalization of the temporal association probability* $\bar{E}_i$.

To make the location semantics and query content semantics in dummy locations anonymous set closely related and indistinguishable, we define association entropy to describe indistinguishability degree in current query time. The larger entropy value, the less information attackers obtain. Thus, the more difficult it is to filter anonymous set based on the strength and weakness of temporal association.

4 Anonymous Algorithm

To further resist temporal association attack, we design an optimal association table in this section. According to the optimal association table, the users' query requests, and the user's personalized privacy requirements, dummy locations are selected. The algorithm of our scheme mainly includes two parts: the construction algorithm of the optimal association table (COAT), and the dummy locations selection algorithm based on semantic and temporal association (SSTA).

4.1 Constructing the Optimal Association Table

Algorithm 1 describes the construction of a optimal association table at the current moment. The concrete steps are as follows:

1) First, initialize the optimal association table and put the location semantics of a user into the optimal association table;

2) Then, according to the temporal association probability of query content semantics and location semantics at the current moment, the semantic with the maximum association entropy in the optimal association table is selected and put into the optimal association table;

3) Finally, repeat Step 2) until the number of elements in the optimal association table reaches the user-defined optimal association table length threshold $dsem$.

Algorithm 1 COAT

Input: user location semantics $Suloc_t$, optimal association table length $dsem$, semantic type set $ST = \{St_1, St_2, \ldots, St_n\}$
Output: optimal association table BSC

1: $BSC \leftarrow \varnothing$
2: $BSC \leftarrow BSC \cup Suloc_t$
3: **while** $BSC.length < dsem$ **do**
4: Choose a St_i $(i \in 1, 2, \ldots, n)$ with the maximum entropy associated with BSC;
5: **end while**
6: **return** BSC;

4.2 Dummy Locations Selection Method

Algorithm 2 describes a dummy location selection method that can resist temporal association attack and inference attack.

First, initialize the anonymous set AS, dummy locations distributed road segments set ASL, anonymous semantic types set ASS and candidate POI set $POIset$. Add user location to AS, the road segment of the user to ASL, the location semantics of the user to ASS (Lines 1∼2). Then, according to the road segment of the user, the algorithm obtains the set of adjacent road segments NL for network expansion. The algorithm selects the number of road segments and semantic types for each adjacent road segment while judging whether it satisfies the user's privacy requirements (Lines 3∼13). Finally, the algorithm considers whether the user's privacy requirements are divided into three situations.

If the number of road segments satisfies the user's privacy requirements, select the dummy locations from the set of candidate POI with similar historical query frequencies, whose semantic types satisfy the user's privacy requirements in the optimal association table BSC (Lines 14∼21).

If the location semantics type in the dummy location anonymous set already satisfies the user-defined semantic diversity, the road segments will continue to be expanded according to the number of road segments required by privacy (Lines 22∼30).

If the above two situations are satisfied, it belongs to the third situation, or the first two situations have been completed. It is determined whether the dummy locations set satisfies K-anonymity, and if not, the appropriate POI is selected as the dummy locations from the candidate set (Lines 31∼39).

Since we have reordered the POI on the obtained road segments, if a attacker continuously sends query requests to the anonymous server, the dummy locations set obtained by the attacker is random. Thus, the attacker cannot infer the user's private information.

Algorithm 2 SSTA

Input: user query Qu, users' personalized privacy requirements PR, the optimal association table BSC of the current moment, historical frequency similarity threshold δ, the set of adjacent road segments NL
Output: the dummy locations set AS

1: $AS \leftarrow \varnothing, ASL \leftarrow \varnothing, ASS \leftarrow \varnothing, POIset \leftarrow \varnothing$
2: $AS \leftarrow AS \cup Uloc_t, ASL \leftarrow ASL \cup uedge, ASS \leftarrow ASS \cup Suloc_t$
3: **for** each edge in NL **do**
4: Get the $POIs$ from the edge, reorder it and remove the edge from NL;
5: **for** each poi in $POIs$ **do**
6: **if** $Spoi \notin ASS$ and $Spoi \notin BSC$ and $|P(poi) - P(uloc_t)| \leq \delta$ **then**
7: $ASL \leftarrow ASL \cup edge, ASS \leftarrow ASS \cup Spoi, AS \leftarrow AS \cup poi$

Remove poi from $POIs$, put the elements in $POIs$ into $POIset$;
Break;
end if
end for
Satisfy $PR.s$ or $PR.l$, or both, and exit the loop;
end for
if $ASL.length == l$ and $ASS.length \neq s$ **then**
for each poi in $POIset$ **do**
if If $Spoi \notin ASS$ and $Spoi \in BSC$ and $|P(poi) - P(uloc_t)| \leq \delta$ **then**
$ASS \leftarrow ASS \cup Spoi, AS \leftarrow AS \cup poi$;
end if
Until satisfy the number of semantic types, exit the loop;
end for
end if
if $ASS.length == s$ and $ASL.length \neq l$ **then**
for each edge in NL **do**
Get the $POIs$ from the edge, reorder it and remove the edge from NL;
if $Spoi \in BSC$ and $|P(poi) - P(uloc_t)| \leq \delta$ **then**
$ASL \leftarrow ASL \cup edge, AS \leftarrow AS \cup poi$
end if
Until satisfy the number of segments, exit the loop;
end for
end if
if $AS.length < k$ **then**
for each poi in $POIset$ **do**
if $|P(poi) - P(uloc_t)| \leq \delta$ **then**
$AS \leftarrow AS \cup poi$
end if
Until satisfy the degree of privacy protection k, exit the loop;
end for
end if
return AS;

5 Experimental analysis

5.1 Experimental Data and Parameter Setting

In this section, we compare the dummy locations selection algorithm (DLS) proposed by Niu *et al.* [3] with the algorithm SSTA in this paper. Both algorithms use the dummy locations selection method to select a group of dummy locations indistinguishable from users' location. At the same time, we compare our scheme with the location semantics-based road network location privacy protection method (LSBASC) proposed by Chen *et al.* [11] and the improved enhancement algorithm (Enhance-LSBASC) proposed by Lv *et al.* [12]. Same as we considered, in the road network environment, these two algorithms construct anonymous sets through network expansion and consider the location semantics and user's personalized privacy requirements.

The above algorithms are implemented by Python, and the hardware platform is Intel (R) Core (TM) i5-1035G1 CPU @ 1.00GHz, 1.19 GHz, with the operating system Microsoft Windows 10.

The experimental data consists of three parts. The first part is the historical query data. We use the check-in dataset as historical query data, and the California sign-in data set is selected in the Gowalla sign-in data set [7], including 54 semantic types. The second part is California road network data, including 21048 vertices and 21693 edges. The third part is that 1000 users are selected to send POI queries. Through the Brinkhoff Thomas mobile object generator [5] to simulate moving objects in the California road network, 10000 mobile objects are generated. Table 1 shows the experimental parameter settings.

5.2 Analysis of Experimental Results

Many aspects have been considered in the performance evaluation for the above algorithms. It mainly includes include the average anonymous execution time, location entropy, association entropy, anonymous success rate, and so on.

1) Average anonymous execution time. It refers to as the time it takes for an algorithm to anonymize successfully. The less it is, the better the efficiency of the algorithm is. Figure 2 shows the relationship between average anonymous execution time and privacy protection degree.

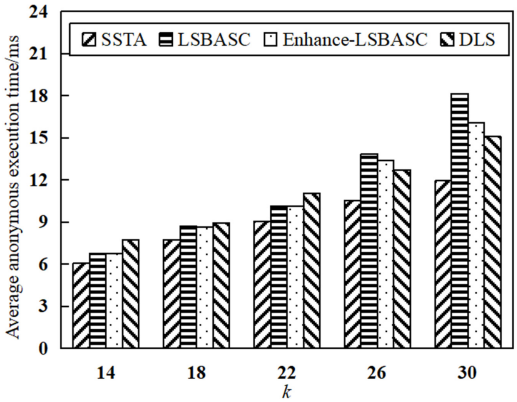


Figure 2: The influence of privacy protection degree k on the average anonymous execution

Obviously, all of the algorithms show an increasing trend. Since the algorithm DLS randomly selects $k-1$ locations with the largest location entropy value as the anonymous set, it needs a lot of calculation to get the optimal set of dummy locations, which takes a relatively long time. The algorithm LSBASC needs to expand road network many times to complete anonymity, like algorithm enhance-LSBASC.

Table 1: Parameter Settings

Parameter	Default values	Range
number of mobile users	10000	
the number of users that request service	1000	
semantic location type	54	
dsem	15	[10,20]
PR.k	14	[2,30]
PR.l	6	[2,15]
PR.s	5	[2,10]
maximum number of cycles	15	

Therefore, the increase in location semantic types and the k value will result in higher time consumption, and the amount of calculation necessary to increase. In our scheme, the proposed algorithm SSTA is less than the other three algorithms. Because we determine whether it satisfies user privacy requirements while selecting network extensions, it reduces the time spent on computing and additional network expansion;

2) Anonymous success rate. It is defined as the proportion of the number of anonymous successes among all anonymous requests. The larger the ratio, the stronger the algorithm's ability to respond to anonymous requests. In Figure 3, the relationship between anonymous success rate and privacy protection degree k is analyzed.

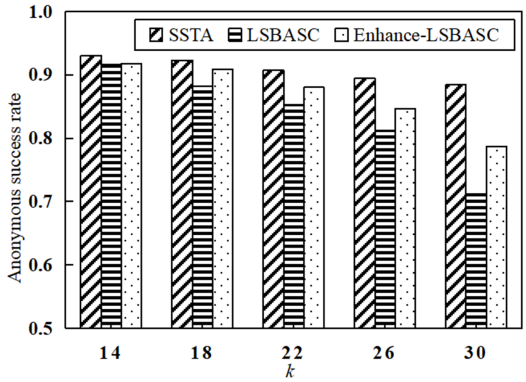


Figure 3: The influence of privacy protection degree k on the anonymous success rate

Obviously, the algorithm SSTA is always better than the others. Because we select and judge whether the current situation satisfies user's privacy requirements. After these two situations are satisfied, the algorithm SSTA randomly selects dummy locations from the extended road segments to achieve k-anonymity. However, the algorithm Enhance-LSBASC adds the current optimal road segment set by calculating the privacy sensitivity on the road to be expanded. As the value k continues to increase, Enhance-LSBASC continuously expands the optimal road section set. The algorithm LSBASC is always lower than the others. Since the optimal road section selected by algorithm LSBASC may exceed the maximum number of cycles in the experiment, the same as the algorithm Enhance-LSBASC, the anonymous protection fails. The maximum number of experimental cycles is set to fail to satisfy its privacy requirements, resulting in a decrease in the anonymous success rate. On the contrary, the algorithm SSTA selects and judges simultaneously. Then, it adjusts according to the lack of privacy protection requirements, so the algorithm SSTA is better than the algorithm Enhance-LSBASC;

3) Location entropy. It refers to the uncertainty of obtaining user's location from dummy locations set. The greater location entropy, the less likely it is for attackers to distinguish users' location from the dummy locations set. Figure 4 shows the relationship between location entropy and privacy protection degree k.

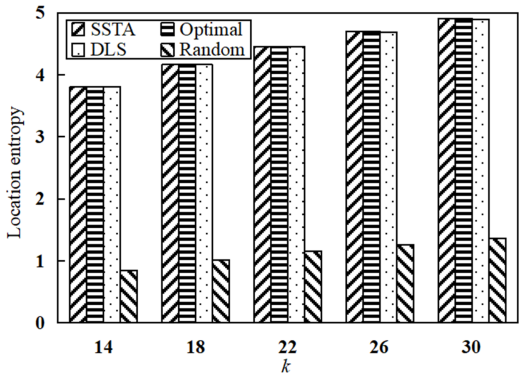


Figure 4: Location entropy under the different anonymity k values

Because the algorithm DLS always selects a set of dummy locations with largest entropy when selecting candidate dummy locations. In contrast, the algorithm SSTA sets a threshold δ when selecting dummy

locations to filter locations with large differences in historical query frequencies, selecting dummy locations with similar historical query frequencies. The location entropy of the algorithm SSTA and DLS tend to the optimal algorithm Optimal as value k changes. In contrast, the algorithm Random does not consider historical query frequency and randomly selects dummy locations, so the location entropy is lower than other algorithms;

4) Association entropy. Because location semantics have a certain relationship with query content semantics in terms of time, in our paper, association entropy is used to measure the indistinguishability of this temporal association. The greater association entropy is, the more difficult it is for attackers to infer the user's location or filter out weakly related locations through the temporal association between dummy locations set and query content. Figure 5 shows the relationship between association entropy and privacy protection degree k. The proposed algorithm SSTA selects dummy locations according to the optimal association table and fully considers the relevance between query content and dummy locations set. Therefore, the association entropy of SSTA tends to the algorithm Optimal. In contrast, the algorithm DLS does not consider this temporal association when selecting dummy locations and tends to the algorithm Random;

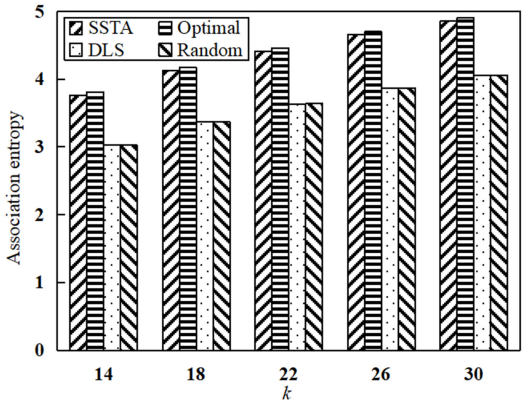


Figure 5: Association entropy under different anonymity value

5) The influence of the number of road segments and semantic types required for privacy protection on the average anonymous execution time and association entropy. Figure 6 shows the influence of road segments required for privacy protection on average anonymous execution time and association entropy. When privacy protection degree k is 14, and the number of semantic types s of privacy protection requirements is 6. As the number of road segments requiring privacy protection increases, the associated entropy does not change significantly. Because the candidate dummy locations are selected according to the optimal association table. With the increase in the number of road segments, the algorithm SSTA needs to constantly expand the number of roads, and the average anonymous execution time will increase slightly.

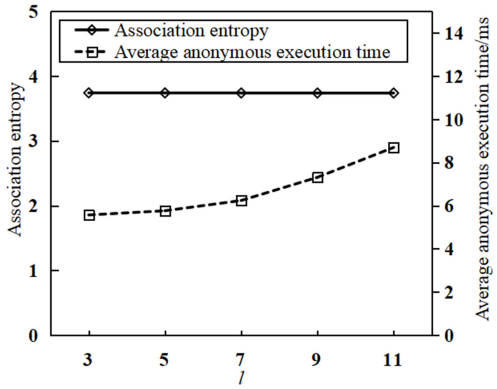


Figure 6: The influence of the number of road segments required on average anonymous execution time and association entropy

Figure 7 shows the influence of the number of semantic types of privacy protection requirements on average anonymous execution time and association entropy. When privacy protection degree k is 14, and the number of segments required for privacy protection is 6. With the increasing number of semantic types required for privacy protection, the associated entropy value does not change significantly. In contrast, the average anonymous execution time increases accordingly. Because the algorithm SSTA needs continuous network expansion to select more candidates to satisfy semantic diversity, the corresponding time will increase.

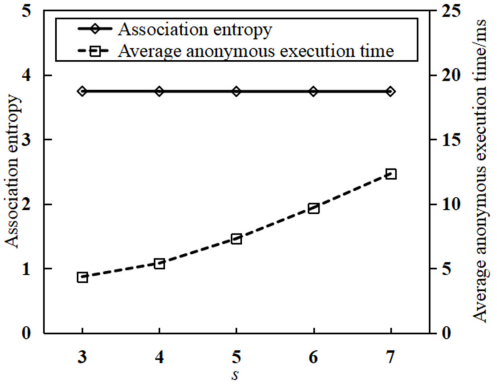


Figure 7: The influence of the number of semantic types required on average anonymous execution time and association entropy

6) System overhead. Since the algorithm SSTA always seaches for $k-1$ dummy locations, the size of anonymous set is always equal to k, reducing unnecessary system overhead. Figure 8 shows the influence of the number of road segment l required on the average anonymous road segments set. Obviously, our proposed algorithm and the others show an upward trend with the increase of value l. On account of adding candidate segments set, the algorithm Enhance-LSBASC system overhead is the largest. While the algorithm LSBASC selects the optimal road segment to add to the anonymous segments set and its system overhead is slightly lower than that of Enhance-LSBASC. However, the number of road segments l for privacy requirements of the algorithm SSTA is always the smallest. Because after the algorithm SSTA satisfies the semantic types of privacy requirements, it considers whether the location on the extended road satisfies the required conditions and then directly joins the anonymous set if it satisfies the requirements. Otherwise, it will continue to expand the road network. This method greatly reduces the system overhead, and the Enhance-LSBASC increases the number of anonymous road segments by expanding the set of adjacent road segments. Therefore, the average anonymous segment set size of the algorithm SSTA is always lower than that of Enhance-LSBASC.

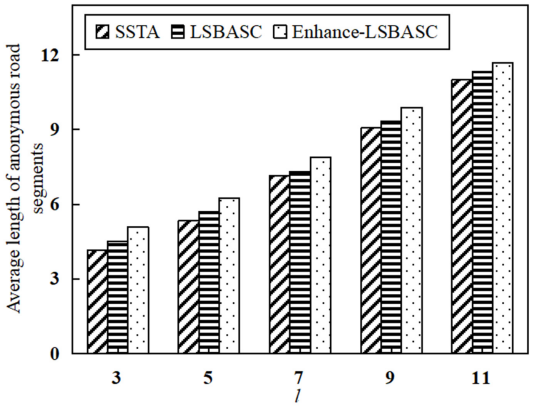


Figure 8: System overhead

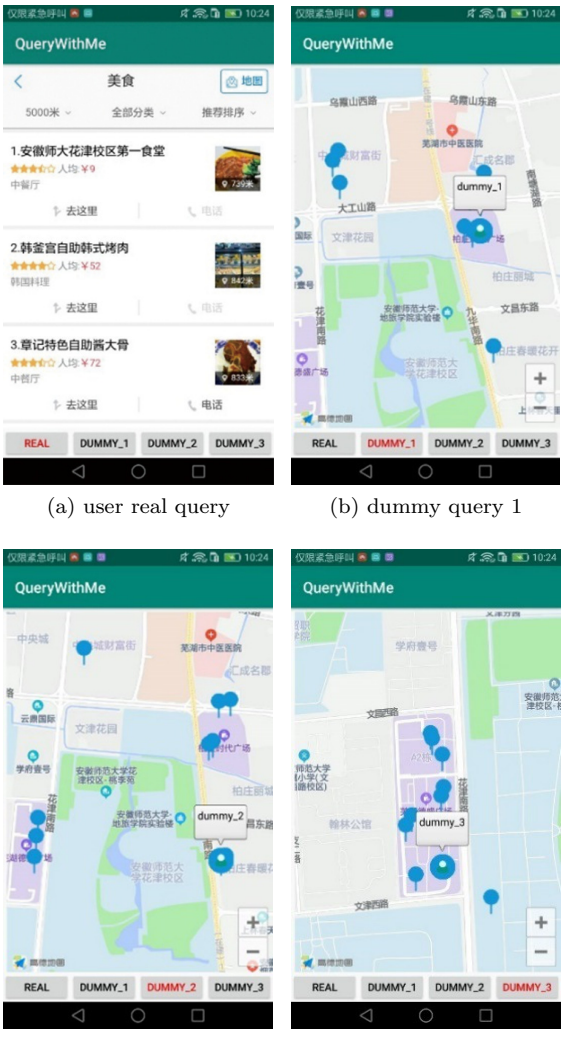
5.3 Practical Application

With the prevalence of smartphones, mobile websites are becoming more and more popular. However, many websites collect location information, which increases the risk of users being accidentally tracked. On the one hand, attackers can intercept user query requests to describe the users' daily habits and preferences. On the other hand, LSP is interested in analyzing the enormous commercial profits behind user query. Based on the proposed algorithm, the next design and implementation assumes that mobile users use a specific network, namely a virtual private network (VPN). The malicious LSP cannot accurately locate the user through IP address requested by the user's location.

Based on Android platform, we use the proposed scheme to develop application software called QueryWithMe, which can confuse user POI query and protects user location privacy.

In the next test, we use the real machine Huawei MT7-TL10 Android 6.0 and Redmi Note 5 Android 9.0 to run our application. For the convenience of testing, the user's privacy requirements are $k = 4, l = 3, s = 2$.

Figure 9 shows the test result of a user querying restaurant. Figure 9(a) shows the feedback of the user's query results. The query results of Amap during the decoy query are shown in Figure 9(b), Figure 9(c), and Figure 9(d). It can be seen that the dummy locations tags and the query results in tags generated by parsing the webpage through the Jsoup library.



(a) user real query (b) dummy query 1

(c) dummy query 2 (d) dummy query 3

Figure 9: Effectiveness Testing on Huawei MT7-TL10 Android 6.0

Figure 10 shows the test result of a user querying hotel. Figure 10(b), Figure 10(c), and Figure 10(d) show the query results, which are not hotels near the query user, but the result that dummy locations request query.

It means that QueryWithMe has successfully provided dummy locations to Amap. Even if the attackers intercept user query and decoy query, it is difficult to obtain the user's location because the dummy locations selected by the SSTA algorithm are truly indistinguishable, and POI query on the dummy location is the same as query in the user's location.

(a) user real query (b) dummy query 1

(c) dummy query 2 (d) dummy query 3

Figure 10: Effectiveness Testing on Redmi Note 5 Android 9.0

6 Conclusions

In this paper, we have proposed a privacy protection scheme for POI query based on semantic and temporal association. Through establishing the temporal association model, this scheme calculates the temporal association probability between query content semantics and location semantics. The generated dummy locations set not only satisfies personalized privacy of users but also effectively resists inference attack and temporal association attack. In addition, based on Android platform, we integrate this scheme to design and implement QueryWithMe, application software that can protect user location privacy. The effectiveness and practicability of our scheme are proved by performance analysis and theoretical analysis. In the future work, we will consider the time division strategy and the spatial distribution characteristics of locations to improve the location privacy protection in POI query.

Acknowledgments

This paper has been awarded by the Natural Science Research Project for Universities in Anhui Province (KJ2021A0090). The authors also gratefully acknowledge the reviewers for their helpful comments and suggestions, which have improved the presentation.

References

[1] M. Alotaibi, M. I. Ibrahem, W. Alasmary, D. Alabri, and M. Mahmoud, "UBLS: User-based location selection scheme for preserving location privacy," in *2021 IEEE International Conference on Communications Workshops*, pp. 1–6, 2021.

[2] N. Amukelani and M. Siyabonga, "Towards enhancing security in android operating systems – android permissions and user unawareness," in *Proceedings of the 2nd International Conference on Computer Applications and Information Security*, pp. 1–6, 2019.

[3] A. Arshad, H. Abbas, W. B. Shahid, and A. Azhar, "Deceiving eavesdroppers by real time persistent spoofing of android users' location coordinates for privacy enhancement," in *the 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 107–112, 2020.

[4] B. Bostanipour and G. Theodorakopoulos, "Joint obfuscation of location and its semantic information for privacy protection," *Computers and Security*, vol. 107, no. 4, pp. 102310–102332, 2021.

[5] T. Brinkhoff, "A framework for generating network-based moving objects," *Geoinformatica*, vol. 6, no. 2, pp. 153–180, 2002.

[6] H. Chen and X. L. Qin, "Location-semantic-based location privacy protection for road network," *Journal on Communications (in Chinese)*, vol. 37, no. 22, pp. 67–76, 2016.

[7] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1082–1090, 2011.

[8] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applica-tions, and Services*, pp. 31–42, 2003.

[9] G. W. Hu, B. Zhang, X. Xiao, W. Z. Zhang, L. Liao, Y. Zhou, and X. Yan, "Samldroid: A static taint analysis and machine learning combined high-accuracy method for identifying android apps with location privacy leakage risks," *Entropy*, vol. 23, no. 11, pp. 1489–1505, 2021.

[10] 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," *Journal of Software (in Chinese)*, vol. 29, no. 3, pp. 703–720, 2018.

[11] R. Liu, K. Z. Zuo, Y. L. Wang, and J. Zhao, "Location privacy-preserving method based on degree of semantic distribution similarity," in *ICPCSEE 2020 Communications in Computer and Information Science*, pp. 118–129, 2020.

[12] X. Lv, H. Shi, A. Wang, T. Zeng, and Z. Wu, "Semantic-based customizable location privacy protection scheme," in *Proceedings of the 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science*, pp. 148–154, 2018.

[13] M. Ma, "Enhancing privacy using location semantics in location based services," in *Proceedings of the 3rd International Conference on Big Data Analysis*, pp. 368–373, 2018.

[14] R. Nieminen and K. Jarvinen, "Practical privacy-preserving indoor localization based on secure two-party computation," *IEEE Transactions on Mobile Computing*, vol. 20, no. 9, pp. 2877–2890, 2021.

[15] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," in *Proceedings of the 33rd Annual IEEE International Conference on Computer Communications*, pp. 754–762, 2014.

[16] Y. L. Wang, K. Z. Zuo, R. Liu, and J. Zhao, "Dynamic pseudonym semantic-location privacy protection based on continuous query for road network," *International Journal of Network Security*, vol. 23, no. 4, pp. 642–649, 2019.

[17] C. Yang and W. Yan, "Location privacy protection scheme based on location services," in *Proceedings of the 9th International Conference on Communication and Network Security*, pp. 30–33, 2019.

[18] P. Zhang, C. Hu, D. Chen, H. Li, and Q. Li, "Shiftroute: Achieving location privacy for map services on smartphones," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4527–4538, 2018.

[19] Q. Y. Zhang, X. Zhang, W. J. Li, and X. H. Li, "Overview of location trajectory privacy protection technology based on lbs system," *Application Research of Computers (in Chinese)*, vol. 37, no. 12, pp. 3534–3544, 2020.

[20] Y. B. Zhang, Q. Y. Zhang, Y. Yan, Y. L. Jiang, and M. Y. Zhang, "A k-anonymous location privacy protection method of polygon based on density distribution," *International Journal of Network Security*, vol. 23, no. 1, pp. 57–66, 2021.

[21] M. Zhao, X. Zhu, J. Niu, and J. F. Ma, "A semantic-based dummy generation strategy for location privacy," in *2019 International Conference on Networking and Network Applications*, pp. 21–26, 2019.

[22] X. H. Zhu and R. L. Qi, "A new fast matching method for dummy k-anonymous location privacy protection in location based services," *International Journal of Network Security*, vol. 23, no. 5, pp. 888–894, 2021.

Biography

KaiZhong Zuo received the Ph.D. degree in computer science from Shanghai University, Shanghai, China, in 2011. He is currently a professor with the School of Computer and Information at Anhui Normal University, Wuhu, China. His research interests include data security, privacy preservation, and machine learning.

Jun Zhao is a Master's student in the School of Computer and Information of Anhui Normal University. His research interests include data security and privacy preservation.

Peng Hu received the Ph.D. degree in computer science from Nanjing University of Science and Technology, Nanjing, China, in 2021. He is currently an assistant professor with the School of Computer and Information at Anhui Normal University, Wuhu, China. His research interests include applied cryptography, information security, and the Internet of Vehicles.

Zhangyi Shen is currently an assistant professor with the School of Computer and Information at Anhui Normal University, Wuhu, China. He completed his Ph.D. at Shanghai University and experienced post-doctoral at Harbin Institute of Technology, Shenzhen. His research interests include artificial intelligence, big data analysis, and decision support.

Xixi Chu is a Master's student in the School of Computer and Information of Anhui Normal University. Her research interests include data security and privacy preservation.