A Fine-Grained Spatial Cloaking Scheme for Privacy-Aware Users in Location-Based Services

Ben Niu*, Qinghua Li†, Xiaoyan Zhu* and Hui Li*
*National Key Laboratory of Integrated Networks Services, Xidian University, China
†Department of Computer Science and Computer Engineering, University of Arkansas, AR, USA
*xd.niuben@gmail.com, †qinghual@uark.edu, *{xyzhu, lihui}@mail.xidian.edu.cn

***Abstract*—In Location-Based Services (LBSs) mobile users submit location-related queries to the untrusted LBS server to get service. However, such queries increasingly induce privacy concerns from mobile users. To address this problem, we propose FGcloak, a novel fine-grained spatial cloaking scheme for privacy-aware mobile users in LBSs. Based on a novel use of modified Hilbert Curve in a particular area, our scheme effectively guarantees k-anonymity and at the same time provides larger cloaking region. It also uses a parameter σ for users to make fine-grained control on the system overhead based on the resource constraints of mobile devices. Security analysis and empirical evaluation results verify the effectiveness and efficiency of our scheme.**

I. Introduction

Location-Based Services (LBSs) have been popular in recent years. The widely used modern mobile devices such as smartphones and tablets provide mobile users with more opportunities of communications and better awareness of their surroundings. Through Apple Store or Google Play Store, users can download and install location-based applications into their smartphones, submit queries to LBS servers, and obtain location-related service data about Point of Interests (POIs) in vicinity. For example, users can look for the clinics or banks nearby, and check the price information of the nearest Red Lobster restaurant.

Normally, the LBS servers serve a user based on its submitted LBS query (e.g., *show me the clinic information within 1 mile*), which typically includes a $\langle location, query\ interest\rangle$ pair and possibly some other information such as the user's ID, query radius, etc. However, these submitted information may be abused by the untrusted LBS servers (and other parties that compromise the servers). Hence the LBS servers may know where the users are, what kind of queries they submit, what they are doing, etc. They may track users or release their personal information to third parties such as advertisers. We thus need to pay more attention to protecting privacy.

To address the privacy issue, many approaches have been proposed over recent years in the literature. The state-of-the-art approaches can be roughly divided into two main categorizes [1]: trusted anonymization server-based schemes [2], [3], [4], [5], [6], [7] and mobile devices-based schemes [8], [9], [10], [11], [12], [13], [14]. Most of them achieve *k-anonymity* [15] using location perturbation and obfuscation, temporal and spatial cloaking or dummies. Among these schemes, the temporal and spatial cloaking [2], [8], [5], [9], [10], [7], [14] technique is very popular and can be deployed to real smartphones easily. Such schemes either minimize the cloaking region [8], [7] to reduce the system overhead, or maximize the cloaking region [9], [14], [16] to provide better privacy. In trusted anonymization server-based schemes, a query is submitted to the LBS server via a trusted third-party server (e.g., *location anonymizer* [4], [6]), which enlarges the queried location into a bigger cloaking region covering $k-1$ other users to achieve *k-anonymity*. In this way, the untrusted LBS server cannot identify the user's real location. These schemes rely on a trusted server, which becomes the weak point of the system and also a single point of failure. Mobile device-based approaches remove the trusted server by constructing the cloaking region based on exchanged location information from other encountered mobile users. However, both approaches have limitations. First, existing solutions either provide users with the minimum cloaking region or the maximum, but lack a balanced consideration between user's required privacy level and the constrained resources of their mobile devices. Second, sometimes it is difficult to find enough users in a reasonable cloaking region.

To address the aforementioned problems, in this paper, we propose a Fine-Grained Spatial Cloaking scheme, called *FGcloak*, which achieves *k-anonymity* for users in LBSs and provides fine-grained control on the system overhead. Different from existing approaches, *FGcloak* uses a set of algorithms to do fine-grained spatial cloaking. First, *FGcloak* uses a Modified Hilbert Curve Constructing (MHCA) algorithm to fully fill the considered map area based on users' query probability. Then, to provide *k-anonymity* and guarantee bigger cloaking region, it uses a Privacy-Aware Dummy Selection (PADS) algorithm to carefully separate the modified Hilbert curve into k segments. Finally, it uses a Fine-Grained Local Replacement (FGLR) algorithm to reduce the system overhead according to users' personalized requirements.

The major contributions of this paper are as follows.

- We construct a modified Hilbert Curve considering users' query distribution, and design a spatial cloaking scheme based on it to protect user's location privacy in LBSs. This scheme protects privacy through *k-anonymity* and large cloaking regions. Due to the dimension reduction property of Hilbert Curve, the system overhead can also be reduced.
- Through the Fine-Grained Local Replacement (FGLR) algorithm which combines dummy-based and encounter-based

approaches, we provide users with fine-grained controls on system overhead.

- We provide thorough security analysis and extensive evaluation results to show the effectiveness and efficiency of *FGcloak*.

The rest of the paper is organized as follows. We discuss related work in Section II. Section III presents some preliminaries of this paper. We present our *FGcloak* scheme in Section IV. The security analysis and the evaluation results are shown in Section V and VI, respectively. We conclude this paper in Section VII.

II. Related Work

To protect user's location privacy in LBSs, many research solutions have been proposed over recent years [17], [18], [1]. Most of them try to provide anonymity on users' real locations, such as *k-anonymity* [15], which tries to hide the user's real information into $k-1$ other users. Entropy-based metrics [19], [20], [21], [13] and other metrics [22], [23] have also been widely adopted. Generally speaking, location privacy can be preserved through location perturbation and obfuscation, temporal and spatial cloaking or using dummies. *k-anonymity* was introduced into location privacy by Gruteser *et al.* [2], which hides user's real location to protect location privacy. Their cloaking algorithm constructs spatio-temporal cloaking boxes which contain at least k users and these boxes are sent to the LBS server as locations. *CliqueCloak* [5] is a personalized *k-anonymity* model which allows users to adjust their levels of anonymity. However, most of these work use a *location anonymizer* to generate the cloaking region, rendering the *anonymizer* a central point of failure and the performance bottleneck. To avoid the anonymizer, Kido *et al.* [3] proposed using dummy locations to achieve anonymity. However, they focus on reducing the communication overhead. Also, they use a random walk model to generate dummy locations which cannot ensure privacy when *side information* [24], [14], [16] is available to the adversary. Niu *et al.* [14] proposed a privacy-preserving cloaking scheme which can effectively achieve *k-anonymity* while providing bigger cloaking region. However, the scheme needs a warm-up phase and thus cannot provide protection all the time. Similar problems also happen in many encounter-based solutions such as *SMILE* [11] and *EPS* [21]. Besides the aforementioned approaches, cryptography-based schemes [26], [27] and policy-based solutions [25] are also proposed to protect user's privacy.

Hilbert Curve has also been used in several other schemes (e.g., MobiHide [28]), but most of them use the standard Hilbert Curve which is different from our modified Hilbert Curve. Our modified Hilbert Curve is similar to the Various-grid-length Hilbert Curve (VHC) used in *CAP* [10], but there are several key differences. First of all, VHC is constructed from road density, but our curve is built upon query distribution. Second, VHC is used to perturb a single location, but our curve is used to select dummy locations. Moreover, our scheme provides fine-grained control on system overhead but CAP does not.

III. Preliminaries

In this section, we first present a basic concept and the adversary model used in this paper, and then, we present the motivation and the basic idea of our scheme.

A. Basic Concept

Background Information: the background information in our work is limited to the user's query probability information in the local map. Specifically, suppose the local map is divided into a set of cells (i.e., $n \times n$ cells). The query probability in a particular cell can be represented as the probability that users submit location-based queries from the cell.

B. Adversary Model

Based on the different abilities the adversary has, we consider two types of adversaries in our work, *passive adversary* and *active adversary*.

Passive adversary: any entity can be a *passive adversary* if he can eavesdrop the wireless channels or compromise users to obtain the sensitive information of other users.

Active adversary: any entity can be an *active adversary* if he can compromise the LBS server and obtain all the information known to the server to perform attacks such as *inference attack*. In this paper, the LBS server is considered as the *active adversary*. As the result, he can obtain each user's information and monitor the queries sent from users. Also, the historic data of a particular user as well as the current situation can be captured. In addition, he has knowledge about the location privacy protection algorithm of the system.

C. Motivation and Basic Idea

Mobile users in existing LBSs applications need to submit queries to the LBS server to obtain service data. A typical query includes user's identifier, exact location, the query interest as well as the query range, etc. However, these data may release user's sensitive information to either adversaries or the public. To protect user's privacy, *k-anonymity* is a widely used technique but with several drawbacks. Our work is thus motivated by these drawbacks of existing *k-anonymity*-based solutions. First, the most important problem is caused by the third party server employed in existing approaches (e.g., *location anonymizer* in [4], [6]). Obviously, it becomes the bottleneck of both system performance and privacy concern. Second, few existing solutions provide fine-grained control for mobile users to tune the tradeoff between system overhead and privacy based on the restricted resources of their smartphones. Last but not least, the size of cloaking region cannot always be guaranteed, especially when the system is at the warming up phase in [14] or a small number of users are encountered in [11], [20], [21]. Hence, our work is to design a fine-grained cloaking algorithm for privacy-aware mobile users without relying on any third part servers.

To achieve fine-grained cloaking for privacy-aware users, our main idea is to employ a modification of Hilbert Curve to provide effective *k-anonymity* protection. Our approach works in several steps, which is illustrated in Fig. 1 and Fig. 2.

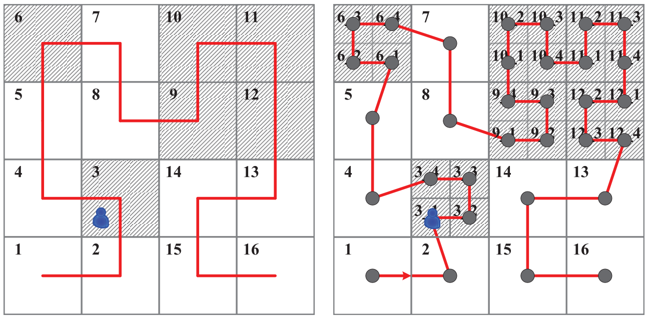
First, we modify the standard Hilbert Curve according to the query distribution. Specifically, Fig. 1(a) shows the standard Hilbert Curve which covers the whole local map. Since the distribution of all the queries from users in the local map may not be an uniform distribution, we modify the standard Hilbert Curve considering the query probability. Normally, higher query probability leads to finer grains. We perform the standard Hilbert Curve with finer grains in the region with higher query distribution. Based on the modified Hilbert Curve, we can obtain a set of points shown in Fig. 1(b). The corresponding Hilbert Value of the standard Hilbert Curve and our modified Hilbert Curve can be found in Fig. 2(a) and Fig. 2(b), respectively.

Second, due to a well-known property of Hilbert curve that two adjacent points in the projected space are likely to be close in the original space [10], given a particular point, we can easily find out the adjacent points around. This property is a normal use of the Hilbert Curve in constructing the cloaking region in LBSs. While in our work, we prefer to construct a bigger cloaking region as some existing research approaches [14] do. Therefore, the adjacent points in the Hilbert Curve should be avoided. To achieve this goal, we separate the modified Hilbert Value into k segments, where k is a user-defined value (e.g., $k = 4$ in Fig. 2(c)). Note that the user's real location is in one segment. Then we choose $k - 1$ candidates from the other $k - 1$ segments. Through this way, the chosen candidates can cover an as big area of the local map as possible under the limitation k. Now we have $k-1$ candidates in hand to achieve *k-anonymity* while guaranteeing a desired cloaking region.

Although we can generate dummy locations at each chosen candidate easily, it is hard to guarantee the effectiveness for *k-anonymity* since some locations are unlikely to be real (such as lakes, swamps, and rugged mountains) and can be easily filtered out by the adversary with *background information*. We can collect some history locations from encountered users to make a more realistic anonymous set; however, the cost may be high in terms of communication, computation and storage. To solve this problem, we combine dummy-based and encounter-based solutions together through a Fine-Grained Local Replacement (FGLR) algorithm. Specifically, in our scheme, each user shares some history locations with encountered users. When the LBSs are needed, the $k - 1$ candidates can be determined by performing the aforementioned processes. For each candidate in the chosen set, if there is a proper location within comfortable offset in the buffer, our FGLR algorithm makes replacement; otherwise, our FGLR algorithm uses a carefully perturbed dummy location to achieve *k-anonymity*.

IV. Our Fine-Grained Cloaking Scheme

In this section, we present the system architecture of our proposed *FGcloak*, then, we introduce the three algorithms in our scheme in details.



(a) Standard Hilbert Curve (b) Modified Hilbert Curve

Fig. 1. Our Modified Hilbert Curve based on query distribution

A. System Architecture

Our proposed *FGcloak* is a pure P2P-based scheme. Mobile users of *FGcloak* communicate with each other within a collaborative group through WiFi/Bluetooth/Ad Hoc network, and connect to the LBS server through cellular networks, such as 3G/4G. Each mobile user in our system keeps a buffer to record and maintain the information exchanged from encountered users. When a user *Alice* needs LBSs, she runs the Modified Hilbert Curve Constructing (MHCC) algorithm, the Privacy-Aware Dummy Selecting (PADS) algorithm, and the Fine-Grained Local Replacement (FGLR) algorithm to construct the cloaking region, and then sends the query as well as the cloaking region to the LBS server. Fig. 3 shows the data flow of our proposed *FGcloak*.

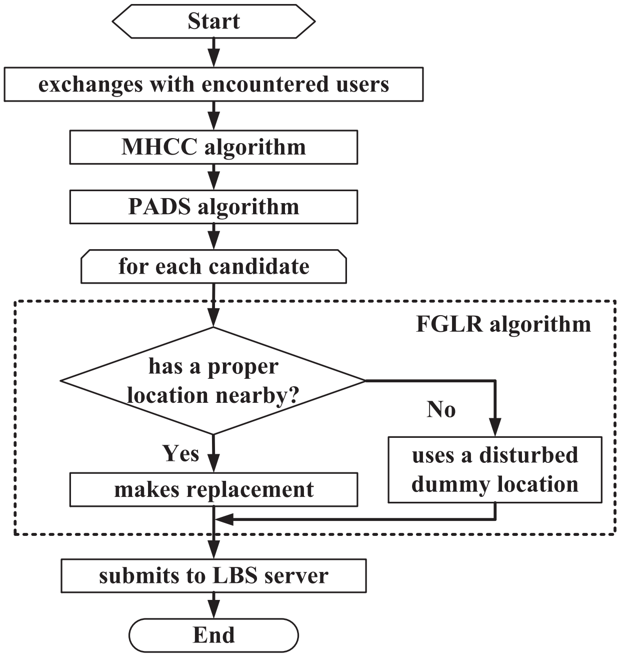
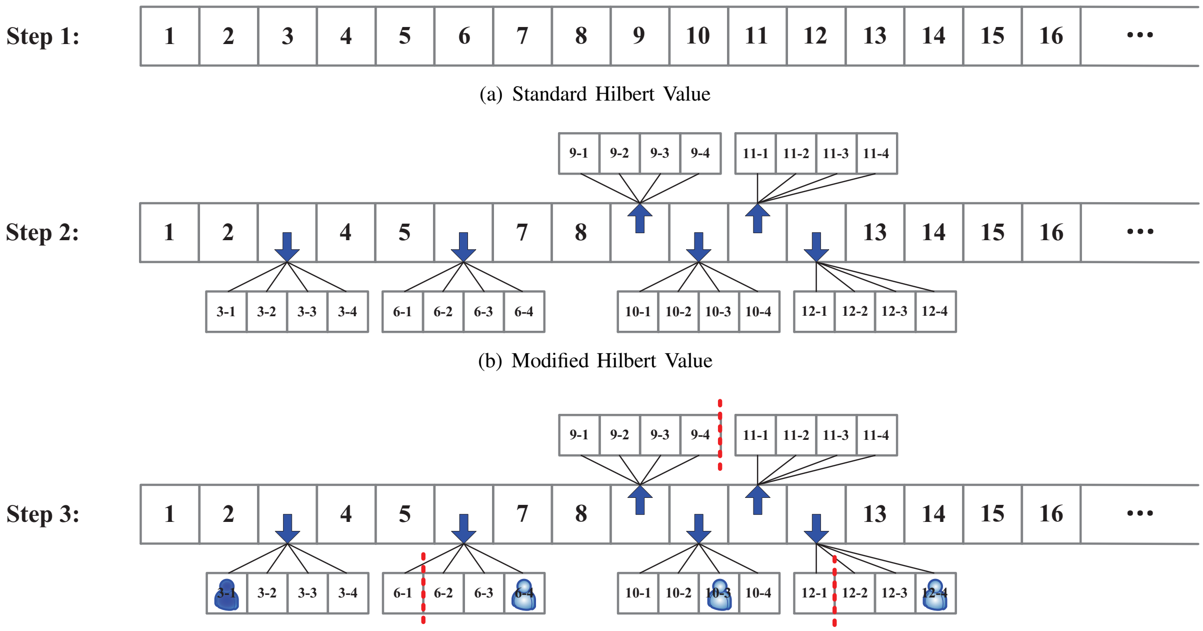


Fig. 3. Data Flow in our Scheme

B. Modified Hilbert Curve Constructing Algorithm

Generally, the query probability in local map can be obtained from some third parties easily, e.g., the Internet or some



(a) Standard Hilbert Value

(b) Modified Hilbert Value

(c) Separation and Candidates Chosen

Fig. 2. Our Solution

social network applications such as Yelp! and Foursquare. With these information, the mobile user can construct a modified Hilbert Curve, which covers the whole map with different grains. Normally, bigger query probability leads to denser Hilbert Curve. Let's recall the aforementioned example shown in Fig. 1. Specifically, Fig. 1(a) shows the standard Hilbert Curve when the queries are uniformly distributed in the area. In Fig. 1(b), suppose the query probability in the 1^{st} region is a benchmark, marked as 1. The 2^{nd} region's query probability is similar to the 1^{st} region, hence marked as 1 too. The query probability in the 3^{rd} region is much higher, say, 4 times of the benchmark. Thus, we divide the 3^{rd} region into finer grains (e.g., through quadrant technique), and mark them as 3-1, 3-2, 3-3 and 3-4 respectively. Similarly, the 6^{th}, 9^{th}, 10^{th}, 11^{th} and 12^{th} regions are also divided into finer grains. In this way, the modified Hilbert Curve is constructed. We can see that the modified Hilbert Curve covers the regions of higher query probability with higher density. Note that, since the query probability in an area usually does not change very frequently, this algorithm can be accomplished offline.

C. Privacy-Aware Dummy Selecting Algorithm

Based on the standard Hilbert Value (shown in Fig. 2(a)) of the standard Hilbert Curve, we can easily compute the modified Hilbert Value (shown in Fig. 2(b)) by region quadrants. Next, we describe the PADS algorithm using the example in Fig. 2. Specifically, we first count the total number (N_{total}) of the cells irrespective of their sizes (see Fig. 2(b)), and then evenly divide them into k segments (see Fig. 2(c)). The average number ($N_{average}$) of cells in each segment can be computed by

$$N_{average} = \lceil \frac{N_{total}}{k} \rceil. \qquad (1)$$

According to the rank (say r) of the real user (c_{real}) in her own segment (e.g., $r = 1$ if she is the first element of the segment), we choose the r^{th} element of each of the remaining $k-1$ segments, and use the chosen elements as the $k-1$ candidates. For example, suppose we try to achieve *4-anonymity*, in Step 2 of the example in Fig. 2, the total number of the cells $N_{total} = 34$. We can compute $N_{average} = \lceil \frac{N_{total}}{k} \rceil = \lceil \frac{34}{4} \rceil = \lceil 8.5 \rceil = 9$. In Step 3 of the example in Fig. 2, we divide the modified Hilbert Value into 4 segments as 1 ~ 6-1, 6-2 ~ 9-4, 10-1 ~ 12-1 and 12-2 ~ 16. Then, based on the rank of the real user c_{real} in her segment which is 3, we choose the third cell (6-4, 10-3 and 12-4, respectively) in each segment as the other 3 candidates. The formal description of our PADS algorithm can be found in Alg. 1.

```
Algorithm 1: Privacy-Aware Dummy Selecting Algorithm
Input  : standard Hilbert Value, current cell $c_{real}$ and $k$
Output: an anonymous set of candidates
constructs the modified Hilbert Value;
counts $N_{total}$;
computes $N_{average} = \lceil \frac{N_{total}}{k} \rceil$;
divides the modified Hilbert Value;
chooses the other k-1 candidates from each segment;
outputs the anonymous set.
```

D. Fine-Grained Local Replacement Algorithm

We provide a fine-grained local replacement algorithm by combining dummy-based and encounter-based solutions. Generally speaking, dummy-based approach has low communication and computation cost but cannot guarantee the effectiveness of generated dummy locations since some dummy locations may appear in unlikely places in reality and can be easily filtered out by the adversary. Encounter-based solution can guarantee that the obtained information (e.g., locations) from encountered users are solid and effective to achieve *k-anonymity*, but it has higher overhead in terms of communication, computation and storage. Our idea is to combine the two and achieve a tradeoff between them.

In our scheme, each mobile user uses encounter-based algorithm to share and obtain information from encountered users. Specifically, each user records the history locations periodically (e.g., every 5 minutes), and shares part of collected history data (i.e., a randomly chosen collected location) when an encounter happens. To control the cost of information exchange, we use a parameter *exchange ratio* (denoted by σ) to adjust the amount of exchange. It measures the fraction of time during which a user exchanges information with other encountered users. Each user can set her own *exchange ratio* based on the available resources of her mobile device. We use a simple example shown in Fig. 4 to show the definition of σ. Specifically, each user divides time into segments such as (t_0, t_2), (t_2, t_4) and (t_4, t_6), etc. In the first segment (t_0, t_2), the user turns off encounter-based exchange during (t_0, t_1) but turns on encounter-based exchange during (t_1, t_2). Formally, the exchange ratio σ can be computed as

$$\sigma = \frac{t_2 - t_1}{t_2 - t_0}, 0 < \sigma \leq 1. \tag{2}$$

Generally, bigger σ indicates more chances of communicating with encountered users and higher communication cost as well.

After a user obtains $k-1$ candidates by performing the MHCC and PADS algorithms, she constructs the cloaking region locally with both the obtained information from encountered users and carefully generated dummy users. Specifically, for each candidate, the user searches the exchanged information in the local buffer to find out the nearest location. If the distance from this candidate to the nearest location in the local buffer is close enough (i.e., within the same cell), the original location of this candidate is replaced with the nearest location in the buffer. Otherwise, the original location of the candidate should be perturbed by a random comfortable offset (e.g., 50 or 100 meters) set by the user. Normally, this offset can guarantee that the perturbed location is still within a same cell as the original candidate. In Alg. 2, we formally describe our proposed FGLR algorithm.

V. Security Analysis

We provide security analysis in this section. Some attacks such as *eavesdropping attack* on the wireless channel between users and other entities can be easily avoided by cryptography techniques. Thus, our analysis focuses on other attacks, such as *colluding attacks* and *inference attacks*, which may cause serious privacy problems.

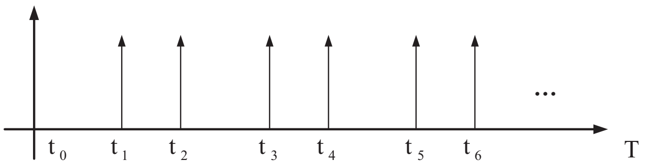


Fig. 4. Exchange ratio in our Scheme

Algorithm 2: Fine-Grained Local Replacement Algorithm

```
Input : an anonymous set obtained from Alg. 1, current
        cell $c_{real}$ and $k$
Output: the cloaking region
for (each candidate in the anonymous set except the
$c_{real}$) do
    searches the nearest location within the local buffer;
    computes the distance;
    if the distance is close enough then
        replaces the candidate with the nearest location;
    else
        disturbs the candidate's location with a randomly
        chosen offset;
    end
end
constructs the cloaking region with the proceeded
anonymous set;
outputs the cloaking region.
```

A. Resistance to Colluding Attacks

Adversaries are likely to collude with some users or the LBS server to obtain other user's private information.

Theorem 1. *Our scheme is collusion resistant.*

Proof: We consider the colluding behavior within a set of users. We prove this theorem from two aspects. First, we analyze the privacy issue when information exchange happens in the encountering phase. In our scheme, each user is independent with others. He maintains a buffer and randomly chooses one location-related record from the buffer to share with the encountered user each time. The chosen record may belong to either the user himself or other users encountered before. Since the buffer size (e.g., 100 or 200) is much bigger than k, the successful guessing probability cannot be bigger than $\frac{1}{k}$. Second, we analyze the privacy issue when generating dummy locations. In our scheme, the FGLR algorithm guarantees that all the processes are executing locally, not dependent on other entities at all. That is to say, it is helpless for the adversaries to compromise and collude with the nearby users; the adversaries can only guess randomly. ■

The best case to this kind of adversaries is that he can get the global information by compromising the LBS server and all the users, but in this case he becomes an *active adversary* as discussed below.

B. Resistance to Inference Attack

We directly consider the untrusted LBS server as the *active adversary* to perform the *inference attack*. It can obtain knowledge by monitoring all the users in the system, including their interests, history queries as well as the current queries, etc. Its aim is to match an observed location within the cloaking region to the real user.

Theorem 2. *Our scheme is inference attack resistant.*

Proof: For the *active adversary*, beside some basic knowledge, it knows the proposed scheme and the related algorithms exactly. We first recall our PADS algorithm. Each candidate is chosen from the cell with same rank in each segment. This technique can confuse the powerful *active adversary*. Due to the *active adversary*'s knowledge, it can perform our scheme for k tests. The best result is that he cannot distinguish the real user from others based on the testing results; i.e., all the test results should be the same or totally different no matter what the input is. In our scheme, for a submitted cloaking region which covers k locations, the *active adversary* can choose any one as the observed real user and perform our scheme. Obviously, it can get the same set of locations for constructing cloaking region. Therefore, it is hard for the adversary to reverse the algorithm.

■

VI. Performance Evaluations

In this section, we present the simulation setup and our evaluation results in turn.

A. Simulation Setup

To evaluate the performance of our proposed *FGcloak*, 10000 mobile users are deployed into a central part ($8km \times 8km$) of the Borlange Data Set[1], which was collected from 1999 to 2001. This dataset is part of an experiment on traffic congestion that happened in Borlange (see [29] for more details). We implement the Levy walk model [30], which has been proven to better describe the mobility patterns of human being [31], to generate synthetic encounters between users. Query distribution is also generated based on the mobility model. Specifically, we separate the map into 32×32 cells similar to the one shown in Fig. 1(a). In every minute, we randomly choose 10% of users, and send a location-based query from these users. After 4 hours simulation, we get 240,000 queries from different cells, and their distribution constitutes the *background information*.

Several parameters are employed in our evaluation. k is related to *k-anonymity*. t represents the simulation time. We compare our proposed *FGcloak* scheme with some recently proposed schemes. The *enhanced-DLS* represents the enhanced dummy location selection algorithm in [14], which protects mobile user's location privacy considering query distribution and the size of cloaking region. *SMILE* [11] is a privacy preserving algorithm trying to provide *k-anonymity* for users in mobile P2P networks. The random scheme means the solution which achieves *k-anonymity* by randomly choosing dummy locations. The optimal scheme shows the optimal results of *k-anonymity* in theory.

B. Evaluation Results

For fairness, we assume that all the aforementioned schemes have no *background information* at the beginning. They need to communicate with others to obtain these information gradually. Based on the limited information, we evaluate the system performance of different schemes.

1) Cloaking region vs. k: We first evaluate the effect of k on the cloaking region. The simulation time is 60 minutes. Generally, the cloaking region increases with increasing k. The cloaking regions of all the mentioned schemes are smaller than the ideal value achieved in the optimal scheme. The reason is that in those schemes the real user may be located at any place of the map, which affects the size of cloaking region significantly. In Fig. 5, the cloaking region stays at 64 km^2 all the time in the optimal solution. The random scheme outperforms the *enhanced-DLS* [14] and *SMILE* [11] schemes in most cases, since both of the *enhanced-DLS* [14] scheme and the *SMILE* [11] scheme need a warm-up phase to learn the information of users far away. While in our *FGcloak*, the modified Hilbert Curve can be constructed to guarantee the full coverage on the map, and with the help of our PADS algorithm, we can effectively avoid the dummies located very close to each other. As the result, the cloaking region in our scheme is much larger than other existing solutions and is close to the ideal results in the optimal scheme when k goes large.

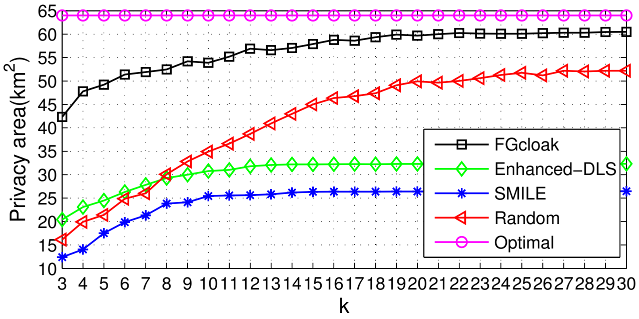


Fig. 5. Cloaking region vs. k

2) Cloaking region vs. t: Fig. 6 shows the evaluation results of the cloaking region with varying simulation time t. Here we set $k = 10$. We can see that the optimal results are fixed at 64 km^2, and the results in the random scheme are around 35 km^2, which indicates that the dummy selection phase has no relationship with the simulation time. Comparing our *FGcloak* with other schemes, the performance of our *FGcloak* is much better than other schemes in all the cases. It stays at a higher

[1] The data set is available at http://icapeople.epfl.ch/freudiger/borlange.zip in Jan. 2012

level 54 km^2 on average. We can also see that the cloaking region of the random scheme outperforms the schemes of *enhanced-DLS* [14] and *SMILE* [11] when the simulation time is short (e.g., $t \leq 70$ and $t \leq 150$ minutes, respectively). As an example, when $t = 30$ minutes, the cloaking region of *enhanced-DLS* and *SMILE* are 18.81 km^2 and 17.37 km^2. However, when $t = 120$ minutes, these values are 39.15 km^2 and 32.59 km^2, while the cloaking region can achieve 54.00 km^2 in our *FGcloak*.

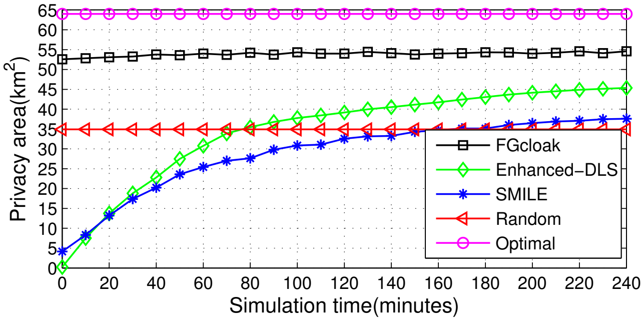


Fig. 6. Cloaking region vs. t, k = 10

3) Entropy vs. k: In this experiment, we run all the simulation for 240 minutes, and then evaluate the effect of k (number of dummies) on entropy. Entropy denotes the uncertainty of determining a user's location from all the candidates [32]. Let p_i denote the probability of the i^{th} possible location being queried in the past. The sum of all probabilities p_i is 1. Then, the entropy H of identifying the real location is given by $H = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$. From Fig. 7, we can see the entropy increases with k. The optimal scheme obviously has the highest entropy in theory. The reason is that all the k candidate locations have the same probability of being queried. The random scheme has the lowest entropy, because it ignores that background information may be exploited by adversaries. The performance of the *enhanced-DLS* [14] scheme is close to the optimal scheme, since it carefully chooses dummy locations to achieve high entropy. The *SMILE* [11] scheme performs much worse since the dummy locations are randomly chosen from the user's buffer. While in the *FGcloak*, the entropy is at a higher level due to use of the FGLR algorithm, which puts each candidate into a proper position with either the encountered information or the dummy information.

4) The effect of exchange ratio σ: As to the fine-grained control on system overhead, we evaluate the effect of the *exchange ratio* σ, which is set by each user. In the following two experiments, we show the distribution of the 10000 users on different "number of encountered users". In Fig. 8, we change the exchange ratio from 0.1 to 1.0, and evaluate the relationship between the number of users in our scheme and the corresponding number of encountered users, which is closely related with the exchanging times. Since more exchanges with encountered users always lead to higher resource consumption on mobile devices, we need to reduce the number of encountered users for each mobile user for the overhead reason. We can see that about 93.15% of users in our simulation have limited number of encounters (less than 10) when $\sigma = 0.1$. We recall the definition shown in Fig. 4. A smaller σ can save more resources for user's mobile device, and it can be achieved in simple ways such as turning off exchange periodically. When we set $\sigma = 0.5$, which means that the exchanging time is a half of the total simulation time, more than a half of users (5837) exchange with others for 20-30 times. Comparing these two set of results ($\sigma = 0.1$ and $\sigma = 0.5$), we find that the total number of exchanges decreases with the decreasing σ. Therefore, we need to turn down this parameter on resources-restricted mobile devices to reduce the system overhead.

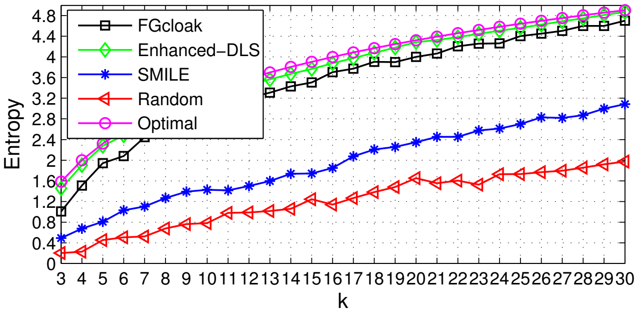


Fig. 7. Entropy vs. k

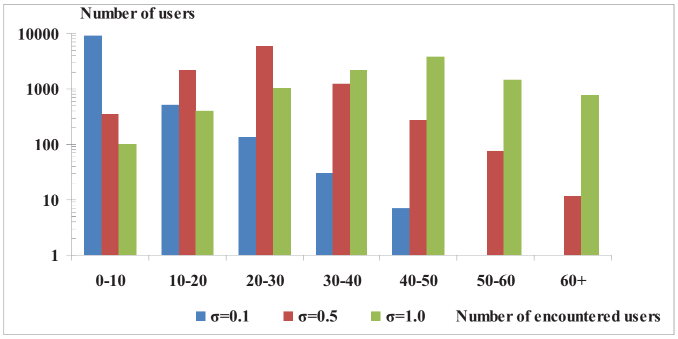


Fig. 8. The effect of exchange ratio σ, $t = 60$ minutes

Similar results can be found in Fig. 9, which shows the relationship between the number of users and the corresponding number of encountered users under different simulation time. Generally, the total number of the exchanges increases significantly with the simulation time t. For example, when the simulation runs for 10 minutes under condition of $\sigma = 1.0$, more than 97% of users only exchange with a limited number of encountered users (less than 10). When the simulation runs for 120 minutes, this value decreases dramatically to 4, and more seriously, about 86.07% of users exchange with others for more than 60 times, which is a huge cost for mobile users in terms of communication, computation and storage.

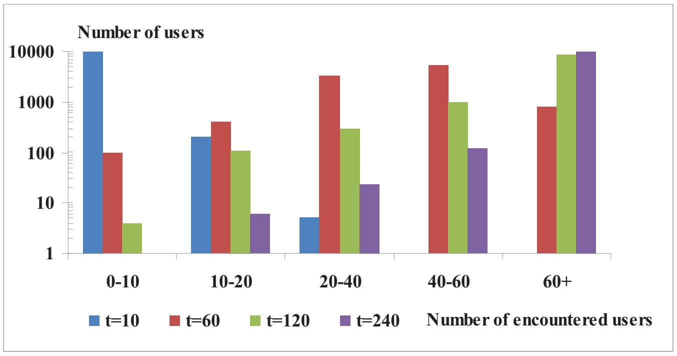


Fig. 9. The effect of the simulation time t, $\sigma = 1.0$

VII. Conclusions

In this paper, we proposed a novel and fine-grained spatial cloaking algorithm, *FGcloak*, which guarantees *k-anonymity* for users while providing fine-grained control on the system overhead. With a set of algorithms including Modified Hilbert Curve Constructing (MHCA) algorithm, Privacy-Aware Dummy Selection (PADS) algorithm and Fine-Grained Local Replacement (FGLR) algorithm, *FGcloak* makes a novel use of modified Hilbert Curve based on the query probability to protect privacy. The fine-grained property is guaranteed by the parameter exchange ratio σ. Security analysis and evaluation results indicated that our scheme is effective and efficient.

Acknowledgment

This work was supported by National Natural Science Foundation of China under Grant 61003300, Fundamental Research Funds for the Central Universities under Grant K5051201041, and China 111 Project under Grant B08038. The work of Dr. Hui Li was supported by the National Project 2012ZX03002003-002, 863 Project 2012AA013102, IRT1078 and NSFC 61170251.

References

[1] K. Shin, X. Ju, Z. Chen, and X. Hu, "Privacy protection for users of location-based services," *Wireless Communications, IEEE*, vol. 19, no. 1, pp. 30–39, 2012.

[2] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. of ACM MobiSys 2003*.

[3] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. of IEEE ICPS 2005*, 2005, pp. 88 – 97.

[4] M. F. Mokbel, C.-Y. Chow, and W. G. Aref, "The new casper: query processing for location services without compromising privacy," in *Proc. of ACM VLDB 2006*.

[5] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, Jan. 2008.

[6] 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.

[7] H. Lee, B.-S. Oh, H.-i. Kim, and J. Chang, "Grid-based cloaking area creation scheme supporting continuous location-based services," in *Proc. of ACM SAC 2012*.

[8] C.-Y. Chow, M. F. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," in *Proc. of ACM GIS 2006*.

[9] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: privacy-area aware, dummy-based location privacy in mobile services," in *Proc. of ACM MobiDE 2008*.

[10] A. Pingley, W. Yu, N. Zhang, X. Fu, and W. Zhao, "Cap: A context-aware privacy protection system for location-based services." in *Proc. of IEEE ICDCS 2009*.

[11] J. Manweiler, R. Scudellari, and L. P. Cox, "Smile: Encounter-based trust for mobile social services," in *Proc. of ACM CCS 2009*.

[12] C.-Y. Chow, M. F. Mokbel, and X. Liu, "Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments," *Geoinformatica*, vol. 15, no. 2, pp. 351–380, Apr. 2011.

[13] X. Zhu, H. Chi, B. Niu, W. Zhang, Z. Li, and H. Li, "Mobicache: When k-anonymity meets cache," in *Proc. of IEEE GLOBECOM 2013*.

[14] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," in *Proc. of IEEE INFOCOM 2014*.

[15] L. Sweeney, "k-anonymity: a model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[16] B. Niu, Z. Zhang, X. Li, and H. Li, "Privacy-area aware dummy generation algorithms for location-based services," in *Proc. of IEEE ICC 2014*.

[17] J. Krumm, "A survey of computational location privacy," *Personal Ubiquitous Comput.*, vol. 13, no. 6, pp. 391–399, Aug. 2009.

[18] Z. Zhu and G. Cao, "Applaus: A privacy-preserving location proof updating system for location-based services," in *Proc. of IEEE INFOCOM 2011*.

[19] J. Meyerowitz and R. Roy Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *Proc. of ACM MobiCom 2009*.

[20] B. Niu, X. Zhu, H. Chi, and H. Li, "3plus: Privacy-preserving pseudo-location updating system in location-based services," in *Proc. of IEEE WCNC 2013*.

[21] B. Niu, X. Zhu, X. Lei, W. Zhang, and H. Li, "Eps: Encounter-based privacy-preserving scheme for location-based services," in *Proc. of IEEE GLOBECOM 2013*.

[22] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *Proc. of IEEE SECURECOMM 2005*.

[23] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. of IEEE Security and Privacy 2011*.

[24] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in location based services," in *Proc. of IEEE INFOCOM 2013*.

[25] W3C. (2011, Apr.) Platform for privacy preferences (p3p) project. [Online]. Available: http://www.w3.org/P3P/

[26] I. Bilogrevic, M. Jadliwala, K. Kalkan, J.-P. Hubaux, and I. Aad, "Privacy in mobile computing for location-sharing-based services," in *Proc. of ACM PETS 2011*.

[27] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proc. of IEEE ICDCS 2014*.

[28] G. Ghinita, P. Kalnis, and S. Skiadopoulos, "Mobihide: A mobilea peer-to-peer system for anonymous location-based queries," in *Proc. of ACM SSTD 2007*.

[29] E. Frejinger, "Route choice analysis: data, models, algorithms and applications," Ph.D. dissertation, Lausanne, 2008.

[30] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the levy-walk nature of human mobility," in *Proc. of IEEE INFOCOM 2008*.

[31] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A new mobility model for human walks," in *Proc. of IEEE INFOCOM 2009*.

[32] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Proc. of ACM PETS 2003*.