A Novel Spatial Cloaking Scheme Using Hierarchical Hilbert Curve for Location-Based Services

Ningning Cui(✉), Xiaochun Yang, and Bin Wang

School of Computer Science and Engineering, Northeastern University,
Liaoning 110819, China
willber1988@163.com,
{yangxc,binwang}@mail.neu.edu.cn

Abstract. With the rapid development of positioning and wireless technologies, Location-based Services (LBS) appear anywhere in our daily life. Though LBS brings a great convience to users, the location privacy is vulnerable in many ways (e.g., untrusted LBS Server). To address privacy issues, we propose *HHScloak*, a novel Hierarchical Hilbert Curve Spatial Cloaking algorithm to effectively achieve k-anonymity for mobile users in LBS. Different from existing methods, we take Average Query Density (AQD) into consideration, and generate the Anonymity Set (AS) which satisfies reciprocity and uniformity. Based on the hierarchical method and optimal splitting bucket strategy, our scheme provides a larger cloaking region which guarantees privacy level. Security analysis and experimental evaluation prove that *HHScloak* scheme can perform more efficiently and effectively than other methods.

1 Introduction

With the fast development of the wireless positioning technologies, and the appearing of the mobile devices (e.g., smart phone, panel PC, etc.), more and more location-based services are widely used by people. But the user information may be divulged by untrusted Location-based Provider (LSP) and other compromised entities. In this case, the user suffers from a threat to privacy (e.g., LSP may get the information and location of the user, and then the LSP can infer users hobbies, religious belief and political activities, etc.). At present, Gruteser et al. [1] first introduce k-anonymity into location-based privacy. Cloaking-based location privacy preserving mechanisms are mainly divided into two aspects [2]. One is Trusted Third-party Server (TTP), and the other is Peer to Peer (P2P). But TTP is easy to become a single point of failure and bottleneck of system performance and P2P is vulnerable by colluding attack and is difficult to satisfy privacy level for generating a too small Cloaking Region (CR).

The work is partially supported by the NSF of China for Key Program (No. 61532021), the NSF of China for Outstanding Young Scholars (No. 61322208), and the NSF of China (Nos. 61173031, 61572122, 61272178).

B. Cui et al. (Eds.): WAIM 2016, Part II, LNCS 9659, pp. 15–27, 2016.
DOI: 10.1007/978-3-319-39958-4_2

According to the aforementioned problems, we propose a novel hierarchical Hilbert Curve spatial k-anonymity algorithm that called *HHScloak* algorithm. Different from existing methods, *HHScloak* takes relevant features (e.g., reciprocity, uniformity) of Anonymity Set into consideration, and assigns equal degree of anonymity to each subregion to resist against positions clustering in a single subregion in candidate set. What is more important is the *HHScloak* scheme can generate a more larger Cloaking Region to realize k-anonymity well. The main contributions are as follows.

a. We conduct a novel system architecture which is based on a cooperative group containing an Auxiliary Server and mobile users.
b. We describe an attack model: Subregion Probability Attack Model. Proposed properties (e.g., reciprocity, uniformity) of Anonymity Set can resist to subregion probability attack well.
c. We propose a novel Hierarchical Hilbert Spatial k-anonymity algorithm. We take Average Query Density into consideration, and construct hierarchical index to fill each layer with Hilbert Curve respectively.

The rest of the paper is organized as follows. We give an overview of related work in Sect. 2. We present preliminaries in Sect. 3. *HHScloak* algorithm is proposed in Sect. 4. The security analysis and experimental evaluation results are shown in Sects. 5 and 6 respectively. Section 7 represents conclusion of our paper.

2 Related Work

In order to protect users location privacy, researchers have proposed a great deal of methods which are based on k-anonymity in LBS. These methods are mainly classified into TTP [3–6,11] and P2P [7–10,12] model.

On the one hand, based on TTP, Bamba et al. [3,4] proposed a *Privacy grid*, according to privacy preference profile (P3P) model, which can satisfy users personal measures (e.g., location k-anonymity and location l-diversity). Through a dynamic bottom-up or top-down spatial grid cloaking algorithm, *Privacy grid* makes sure it has a higher anonymization success rate than Quad Grid Cloaking algorithm. Kalnis et al. [5] proposed Hilbert Cloak (*HC*) algorithm who is the first person to introduce the Hilbert Curve to achieve k-anonymity. *HC* yields a mininal Anonymization Spitial Region (ASR) to achieve k-anonymity which satisfies reciprocity. Lee et al. [6] proposed a Cloaking Region generating algorithm by storing adjacent grid information which is not considered by Hilbert Curve. This method may discard the extention of Cloaking Region, but it may lead to all selected locations clustering in a small region to reveal users real location.

On the other hand, based on P2P, Ghinita et al. [7] proposed a *Prive* algorithm. It creates a distributed B^+ tree based on user information, and generates a k-anonymity region by the order of Hilbert Curve in B^+ tree. Lu et al. [8] proposed the Privacy-Area Dummy (*PAD*) location generating approach, which is generated by a virtual grid or a virtual circle. *PAD* algorithm generates a bigger CR which can be controlled for location k-anonymity. Based on side information and entropy metric, Niu et al. [9,10] proposed a privacy-aware location k-anonymity called Dummy Location Selection (*DLS*) and *enhanced DLS*, and also proposed a Fine-Grained Spatial Cloaking Scheme (*FGcloak*). Based on query distribution, *FGcloak* modifies the standard Hilbert Curve to finer grains which has higher query distribution.

3 Preliminaries

3.1 System Architecture

In this paper, a novel system architecture is proposed, as shown in Fig. 1. System contains an Auxiliary Server and mobile users form a cooperative group through WiFi/Bluetooth/3G/4G Ad Hoc network. Mobile users communicate with LBS Server through similar network. Each mobile user exchanges information with Auxiliary Server periodically, and only sends delayed positions to Auxiliary Server without anything else. Auxiliary Server maintains a location table, which stores user positions and query quantities in each grid (e.g., dividing the entire map into n × n grids). Mobile users send a query $Q = (Q_{id}, Q_{POI}$, AS, CR, Others) to LBS Server, where Q_{id} represents query *id*, which can uniquely identify a query; Q_{POI} represents query interest content; AS represents Anonymity Set; CR represents cloaking region which contains a real user and k-1 dummy users; Others represent time, etc. LBS Server receives queries, deals with them, and returns query results to users.

3.2 Basic Concept

Definition 1 (Average Query Density). *Given a map G, dividing G into* $n \times n$ *grids, for any grid* $G_{ij}, 1 \leq i \leq n, 1 \leq j \leq n, P_{ij}, U_{ij}$, *and* D_{ij} *respectively represents Query Probability, User Density and Average Query Density of* G_{ij}, *and*

$$P_{ij} = \frac{Count_{ij}(Q_u)}{\sum_{i=1}^{n} \sum_{j=1}^{n} Count_{ij}(Q_u)} \tag{1}$$

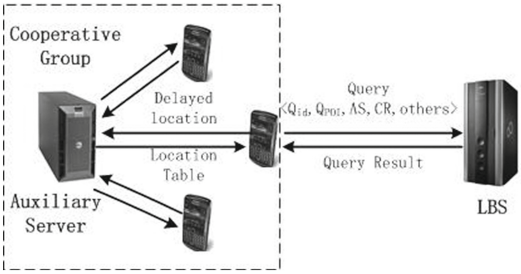


Fig. 1. System architecture

$$U_{ij} = \frac{Count_{ij}(U)}{\sum_{i=1}^{n}\sum_{j=1}^{n} Count_{ij}(U)} \tag{2}$$

where $Count_{ij}(Q_u)$ and $Count_{ij}(U)$ respectively represents the number of user querys and the number of users in G_{ij}

$$D_{ij} = \frac{P_{ij}}{U_{ij}} \tag{3}$$

Definition 2 (Subregion Probability). *Given a region grid C_i, which is the one of i-th layers of the map division. The real user locates in the grid whose AQD is D_u. In C_i, n_i represents the number of grids which has the same AQD with D_u, and SP_i represents subregion probability,*

$$SP_i = \frac{n_i}{N} \tag{4}$$

where N represents all grids which has the same AQD with D_u.

3.3 Attack Model

Generally, different attackers have different background knowledge, and have different abilities to attack. At present, the common attack models including colluding attack model and inference attack model (see Sect. 5). In this paper, we propose a *Subregion Probability Attack* (SPA) model.

Subregion Probability Attack Model. Based on positions distributing in Anonymity Set, for any layer of map division, if existing a subregion probability is much more bigger than others, then attackers can obtain the region of clustering positions and filter out positions in other subregions. What is more, attackers can infer Minimum Containing Anonymity Region (MCAR) of AS, and this disgrades the level of privacy preserving. MCAR represents minimum region which contains maximum subregion probability of AS, and then attackers can further infer the real position.

3.4 Basic Idea and Motivation

Property 1 (Reciprocity) [5]. *Given a mobile user U, U sends a query $Q = (Q_{id}, Q_{POI}, AS, CR, Others)$ to LBS Server, with a cooperative group, and U generates an AS. If AS satisfies: (i) AS contains U and other k-1 mobile users; (ii) with the same anonymity degree k, each user in AS generates the same AS. Then we call AS satisfies reciprocity.*

For example, *Hilbert Cloak* [5] satisfies reciprocity, but *Interval Cloak* [2] and *Casper* [13] algorithms do not satisfy reciprocity. However, *Hilbert Cloak* still exists a few problems, (i) Kalnis et al. [5] indicated that two adjacent points in the transformed space are likely to be close in the original space. Therefore it may lead that the CR is too small to reveal position information. (ii) It does not take AQD into consideration. As shown in Fig. 2, the map is divided into 8×8 grids, and different grid types represent different AQD, generating two anonymity sets AS_1 (e.g., yellow region) and AS_2 (e.g., red region). In the case of the similar size of CR, with the background knowledge of AQD, attackers may filter out the positions which has low AQD, therefore this way disgrades privacy preserving degree. However, different positions have the same AQD in AS_2, and attackers can not distinguish the real position from dummy positions according to background knowledge. Therefore it realizes location-privacy preserving.

Based on aforementioned issues, we propose a *Basic* method. As shown in Fig. 3. The real user u_8 locates in subregion II where the AQD of the grid is D_u. *Basic* method selects the grids which contains red point having the same AQD with u_8 and randomly selects a position in these grids as dummy position. *Basic* regards the selected positions $\{u_1, u_2, \cdots, u_{19}, u_{20}\}$ as a candidate set. According to anonymity degree k, Basic divides positions in candidate set into buckets, and based on relative location of the real user in bucket, *Basic* selects dummy locations in each bucket which corresponds to relative location of the real user. The anonymity set satisfies reciprocity and generates a more larger CR. For $k = 10$, generating an AS is $\{u_2, u_4, u_6, u_8, u_{10}, u_{12}, u_{14}, u_{16}, u_{18}, u_{20}\}$. Though *Basic* realizes k-anonymity well, it still exists a certain problem.

Property 2 (Uniformity). *Given a region grid C_i, which is the one of i-th layers of the map division (see Sect. 4.1). C_i is divided into four subregions C_{i1}, C_{i2}, C_{i3}, and C_{i4}. As the anonymity degree $k^{'}$ of C_i, if each division of the map satisfies, (i) $\mid C_i \mid = \sum_{j=1}^{4} \mid C_{ij} \mid = k^{'}, 1 \leqslant j \leqslant 4, \mid C_i \mid$ represents anonymity degree of the corresponding region; $\mid C_{ij} \mid$ represents anonymity degree of j-th subregion of C_i. (ii) $0 \leqslant \mid\mid C_{ij} \mid - \mid C_{ij'} \mid\mid \leqslant \sigma, 1 \leqslant j^{'} \leqslant 4$. Then we call AS satisfies uniformity.*

As shown in Fig. 3, the anonymity set of subregion I is u_2, and the anonymity set of subregion II is $\{u_4, u_6, u_8, u_{10}, u_{12}\}$. Note that, the subregion probability of AS_{II} is 5 times bigger than AS_I, that is AS does not satisfy uniformity. Attackers can do subregion probability attack through background knowledge, filter out

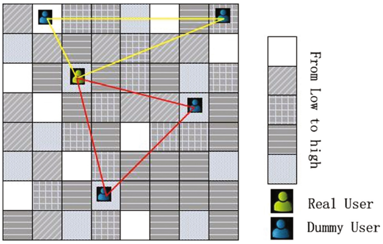


Fig. 2. Problem of existing approach (Color figure online)

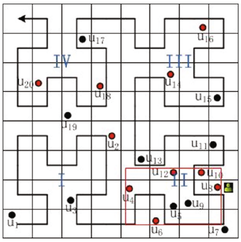


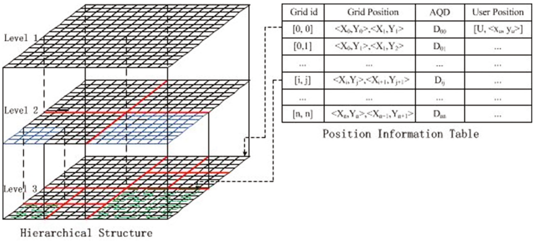
Fig. 3. Basic method (Color figure online)

positions in other regions and generage MCAR (e.g., red rectangle region) where have a higher probability containing the real position. To address this problem, a hierarchical spatial cloaking algorithm is proposed.

4 Hierarchical Hilbert Curve Spatial Cloaking Scheme

4.1 Hierarchical Algorithm Based on Grid

In hierarchical algorithm, the entire map is divided into grids of size $2^n \times 2^n$, n is the system-difined parameter, as shown in Fig. 4 (e.g., $n = 4$). In the system, Auxiliary Server collects users position information, calculates statistic information in the grid, and updates the location table. Location table contains many attributes like grid id, grid position, grid AQD and user position, which is the form of $([i,j], [\langle X_s, Y_s\rangle, \langle X_t, Y_t\rangle], D_{ij}, [u_k, \langle x_k, y_k\rangle])$. Grid position denotes coordinates of points of left-bottom and right-up. User position denotes a list of k positions which have the most queries.



Grid id	Grid Position	AQD	User Position
[0, 0]	$\langle X_0,Y_0\rangle,\langle X_1,Y_1\rangle$	D_{00}	$[U, \langle x_u, y_u\rangle]$
[0,1]	$\langle X_0,Y_1\rangle,\langle X_1,Y_2\rangle$	D_{01}	...
...	...	...	...
[i, j]	$\langle X_i,Y_j\rangle,\langle X_{i+1},Y_{j+1}\rangle$	D_{ij}	...
...	...	...	...
[n, n]	$\langle X_n,Y_n\rangle,\langle X_{n-1},Y_{n+1}\rangle$	D_{nn}	...

Fig. 4. Hierarchical structure

The hierarchical algorithm is as Algorithm 1. First, through the mapping methods, the algorithm maps the real location into a grid, matchs grid id where the real user located in with grid id in location table, and obtains AQD D_u which the grid id is matchable (Line 1). The first layer division (Line 3), the algorithm divides the map region G of $2^n \times 2^n$ into four subregions G_i of $2^{(n-1)} \times 2^{(n-1)}, 1 \leqslant i \leqslant 4$, and calculates quantity of grids n_i which has the same D_u with the grid where the real user located in and corresponding subregion probability SP_i (Line 4). If the results satisfy $0 \leqslant \mid SP_i - SP_{i'} \mid \leqslant \delta_1, i, i^{'} \in [1, 4]$, and $i \neq i', \delta_1 \in [0, 1]$, then the algorithm fills the map region of $2^n \times 2^n$ in Hilbert Curve directly (Lines 6–7). Otherwise the algorithm does the second layer division, and calculates corresponding n_{ij} and SP_{ij} in each subregion $G_{ij}, 1 \leqslant j \leqslant 4$ of region G_i in sequence. If $0 \leqslant \mid SP_{ij} - SP_{ij'} \mid \leqslant \delta_2, j, j^{'} \in [1, 4]$, and $j \neq j', \delta_2 \in [0, 1]$, then the algorithm fills the corresponding subregion of $2^{(n-1)} \times 2^{(n-1)}$ in Hilbert Curve directly. Otherwise the algorithm continues to judge subregion of region G_{ij}. If the number of divided layers h is equal with h_{max}, then the hierarchical algorithm terminates (Lines 9–12).

Algorithm 1. Hierarchical Algorithm

```
Data: Map $G$, Location Table $T$, Threshold Set $\{\delta_i\}$, $h_{max}$
Result: Hierarchical Index
Calculate grid id and AQD $D_u$ where the real user located in ;
Hierarchical_algorithm $(G, T, \delta_i, h, D_u)$;
divide the map region $G$ of $2^n \times 2^n$ into four subregions $G_i$ of
$2^{(n-1)} \times 2^{(n-1)}$ ;
calculate the corresponding subregion probability $SP_i$ ;
for judge subregion grid in sequence do
    if difference of arbitrary two subregion probabilities is no more than $\delta_i$
    then
        fill the region in Hilbert curve directly;
    else
        if $h$ ¡ $h_{max}$ then
            Hierarchical_algorithm $(G_i, T, \delta_{(i+1)}, h + 1, D_u)$;
        else
            break;
return Hierarchical Index
```

4.2 Hierarchical Hilbert Filling Curve Anonymization Algorithm

According to hierarchical algorithm (as shown in Fig. 5), we can easily obtain Hilbert filling Curve of subregion I, II, III and IV (called H_1, H_2, H_3 and H_4). Dummy positions selection algorithm is described as follows. Specifically, we first divide the degree of anonymity k into each subregion, that is subregion degree of anonymity is $k_{sub} = \lfloor k/4 \rfloor$. If k can not be divided exactly, as subregion probability order from big to small, we make $k\%4$ subregion degree of anonymity plus 1, that is $k_{sub} = \lfloor k/4 \rfloor + 1$. As shown in Fig. 5, the subregion degree of anonymity of I, II, III and IV is 2, 3, 2 and 3, the anonymity set satisfies uniformity. Then, as Basic approach, we split the candidate set of each subregion into corresponding k_{sub}-bucket on Hilbert Curve of H_1, H_2, H_3 and H_4. Based on relative location of the real user in bucket, we select dummy locations in each bucket which corresponds with relative location of the real user. As shown in Fig. 6, $k = 5$, respectively, subregion degree of anonymity is 1, 2, 1 and 1, the relative bucket position of the real user u_3 is 3, then Basic selects dummy positions in each bucket which has relative position of 3.

However, if predefined k is 10, as Basic approach, subregion degree of anonymity is 2, 3, 2 and 3, and the relative position of the real user u_3 in bucket is 3. Note that we can not find relative position of 3 in each bucket of subregions of I and III, so the anonymity set can not satisfy reciprocity.

Based on aforementioned issue, we propose an optimal splitting bucket strategy. As Definition 1, we can know the density in each grid is uniform distribution,

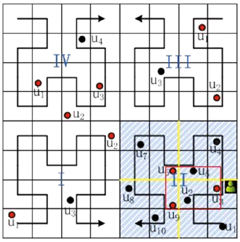


Fig. 5. Hierarchical Hilbert Curve

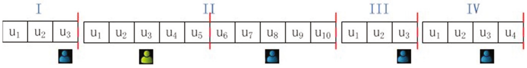


Fig. 6. Spliting of Hierarchical Hilbert Curve

and after dividing a grid into four sub-grids, each sub-grid has the same AQD with the grid. Specifically, we first calculate the maximum number of grids n_{max} in all candidate sets of subregion dummy positions. Then, as the order of Hilbert filling Curve, in each candidate set of subregion dummy positions except the one which contains n_{max}, we subdivide each grid until the number of grids in each candidate set of subregion dummy position is similar to n_{max}. As shown in Fig. 7, subregion II contains the maximum number of grids n_{max} (e.g., $n_{max} = 10$), and subregion I contains the number of candidate grid n (e.g., $n = 3$). From u_1 in subregion I, we subdivide the grid which contains u_1 into u_{11}, u_{12}, u_{13} and u_{14}, then subdivide the grid which contains u_2, until subregion I containing the number of candidate grids n (e.g., $n = 9$) is similar to n_{max}. Subregion III and IV continue to subdivide grids in sequence like I. After subdivision, the number of candidate grids in each subregion is 9, 10, 9 and 10, based on the formula of k_{sub}, we can obtain that each subregion degree of anonymity is 2, 3, 2 and 3. As shown in Fig. 7, We split each subregion positions of candidate set into buckets. The relative position of the real user u_3 in bucket is 3, then we select dummy positions in each bucket which has relative position of 3.

In reality, selecting dummy positions in a grid is random, but it is probable to select a position where there is no user or no LBS. In this paper, we provide a replacement scheme using a structure of cooperative group. Each user cooperates with other users who are in the same cooperative group to maintain a location table, and user position denotes a list of k positions which have the most queries in a period. If the grid in candidate set has not been subdivided, then the scheme

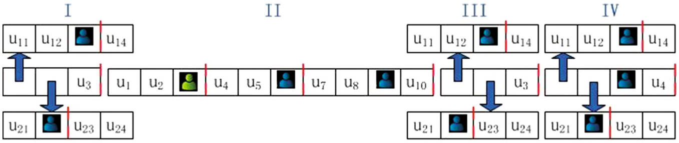


Fig. 7. Optimal splitting bucket scheme

Algorithm 2. Optimal Spliting Bucket Dummy Selecting Algorithm

```
Data: Hierarchical Index; Anonymity degree $k$; AQD $D_u$
Result: Anonymity Set
Dummy_Position Selection (Hierarchical Index,$k$, $D_u$);
Calculate the maximum number of grids $n_{max}$ in all subregion dummy
position candidate sets;
for each subregion except containing $n_{max}$ do
    for candidate grids as Hilbert Curve order in subregion do
        if the number of grids in subregion $n \cong n_{max}$ then
            break;
        else
            subdivide the grid, $n = n + 3$;
Calculate $k_{sub}$ in each subregion;
Split each subregion positions of candidate set into buckets;
Calculate dummy positions, and obtain Anonymity Set;
return Anonymity Set
```

directly selects the position which has the most queries as dummy position in the grid. Otherwise, the scheme selects the position from top-k positions which has the most queries in corresponding sub-grid as dummy position. Though this approach costs more storage space and communication, it guarantees the selected dummy position is a real query position.

5 Security Analysis

In this section, we analyse the security of the proposed *HHScloak* algorithm.

Resistance Against Colluding Attack. Generally, colluding attack means that attackers compromise with other collaborative users or Auxiliary Server, even compromise with LBS Server, and obtain the real user location information.

Theorem 1. *HHScloak is resistant against colluding attack.*

Proof. We prove this theorem in two aspects. First, we analyse the system architecture. Our paper organizes a novel cooperative group structure. Each user does not exchange information with each other, even though attackers compromising with users can not get the real user location information. Moreover, Auxiliary Server only stores users delayed positions periodically, and does not know current user position. Second, aspect of selecting dummy positions. Our proposed method is based on a location table, and attackers can compromise with Auxiliary Server to get the table. But the positions in AS have the same AQD, and attackers can not filter out dummy positions from background knowledge, that is the probability of recognizing real position is no more than $1/k$.

Resistance Against Inference Attack. Inference attack means attackers possess certain knowledge (e.g., historical position, query content, anonymity algorithm, etc.). Based on the submitted AS, attackers infer users real position. Generally, attackers compromise with LBS Server to get background knowledge.

Theorem 2. *HHScloak is resistant against inference attack.*

Proof. Based on background knowledge, attackers obtain submitted AS, and execute anonymity algorithm at each position of AS in sequence to distinguish different positions. Our proposed *HHScloak* k-anonymity algorithm satisfies reciprocity, that is any position in AS generates the same AS. Attackers can not distinguish the real position from other k-1 dummy positions.

Resistance Against Subregion Probability Attack. (*SPA* see Sect. 3.3)

Theorem 3. *HHScloak is resistant against subregion probability attack.*

Proof. The reason of being vulnerable to *subregion probability attack* is positions in AS clustering in a small subregion, which results in the *subregion probability* is much more bigger than other subregions. Attackers may filter out positions in other subregions based on MCAR. Our proposed *HHScloak* algorithm satisfies uniformity, that is each *subregion probability* is similar, and will not appear that existing a *subregion probability* in AS is much more than others. Therefore, attackers can not distinguish the real position through *subregion probability attack*.

6 Performance Evaluations

In our experiments, we use the map of city Aalborg which contains 129680 mobile users in the size of 60 km × 60 km. We divide the map into 64 × 64 grids that the size of each grid is 937.5 m × 937.5 m. In every minute, we randomly choose 10 % mobile users to send location-based queries to LBS Server and at the same time the position table T stores every delayed query record. After 2 h simulation, the Auxiliary Server records all queries which only contains users delayed location information from the selected users, and exchanges the table T with every user (we assume the position information in table T does not change any more in the period).

We set several parameters to evaluate the performance of anonymity approachs. Specifically, k denotes anonymity degree (from 3 to 30 in evaluation) and h denotes the maximum of hierarchy, which are determined by users. Cloaking Region shows the minimum area contains all k-1 dummy positions and the real position. Moreover we use Entropy and Variance to evaluate the uncertainty of AS. Entropy and Variance are defined respectively as follows.

$$Entropy = -\sum_{i=1}^{k} \frac{D_i}{\sum D_i} \times \log_2 \frac{D_i}{\sum D_i} \tag{5}$$

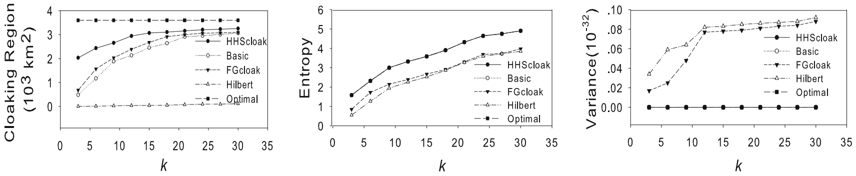
$$Variance = \sum_{i=1}^{k}(D_i - \mu)^2 \tag{6}$$

where D_i denotes the AQD of the grid where the selected user located in. μ denotes mean value of AQD in AS.

We analyse the performance of *HHScloak* from two aspects. First, we compare our proposed *HHScloak* with several recently proposed methods. *Hilbert* denotes Kalnis et al. [5] proposed Hilbert Cloak to achieve k-anonymity. *FGcloak* represents Niu B et al. [9] proposed a fine-gained modified Hilbert Cloak to achieve k-anonymity. The *Optimal* represents ideal condition to complete the k-anonymity. Second, we analyse the relationship between parameter h and the other parameters.

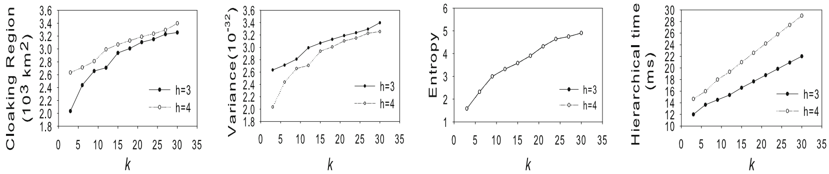
6.1 K vs Cloaking Region, Entropy and Variance

The results are respectively shown in Fig. 8(a) represents k vs Cloaking Region, we can see that the *optimal* scheme achieves ideal result no matter what the value of k is, its Cloaking Region reachs the entire map (size of $3.6 \times 10^9 m^2$). *Hilbert* scheme can not reach a large region, and its maximum is still very small because the approach splits candidate users into buckets in a restricted region (adjacent region). Based on fine-gained Hilber Curve, *FGcloak* can reach a larger region than aforementioned two schemes, but it does not take subregion information of every layer into consideration and just considers the side information of individual grids. Observe that no matter what the value of k is, *HHScloak* can achieve a more larger Cloaking Region, even $k = 3$. Figure 8(b) represents k vs Entropy, Entropy represents uncertainty of AS, the larger value of Entropy is, the higher of uncertainty of AS is, and AS can achieve a better k-anonymity. We know that the same with *Optimal* scheme, *HHScloak* and *Basic* have the equal value of AQD in each grid in AS. So the value of Entropy in these three schemes is $\log_2 k$. However, in *Hilbert* scheme and *FGcloak* scheme, the value of AQD in each grid in AS is uncertain and has lower probability to achieve optimal condition. Figure 8(c) represents k vs Variance, similar to Entropy, Variance also represents uncertainty of AS by the way of statistics. But contrast to Entropy, the lower value of Variance is, the higher of uncertainty of AS is. Note that the mean value of AQD of *HHScloak* and *Basic* scheme in AS is equal to *Optimal*



(a) k vs CR (b) k vs Entropy (c) k vs Variance

Fig. 8. The effect of HHScloak algorithm, h = 3



(a) h vs CR (b) h vs Variance (c) h vs Entropy (d) h vs Time

Fig. 9. The effect of h in HHScloak algorithm

condition in a large extent, so the value of Variance tends to be zero, but in *Hilbert* and *FGcloak* scheme, the difference between mean value and them is obvious. Along with the increasing value of k, the uncertainty of different grids become more distinct and the curve unexpectedly rise up.

6.2 H Vs Cloaking Region, Entropy, Variance and Hierarchical Time

The relationship between these several parameters in *HHScloak* is shown in Fig. 9. We respectively set $h = 3$ and $h = 4$ to evaluate the effect of different number of layers. Figure 9 represents when h is 3 or 4, the system creates a hierarchical structure of 3 layers or 4 layers and h is 3 which the minimum subregion is 16×16 and h is 4 which the minimum subregion is 8×8. Then we can note that when h is 4, it has a higher resolution of Hilbert Curve traversal. As show in Fig. 9(a), when h is 4, the Cloaking Region is larger than h is 3, because h is 4 has more exact division. But along with the increasing value of h, the cost becomes higher. Figure 9(d) shows that h is 3 has a lower cost of creating hierarchical time than h is 4. Figure 9(b) and (c) shows that the hierarchical influence on Variance and Entropy is very unconspicuous. Though the curve of h is 3 and h is 4 in Fig. 9(b) is different, the magnitude is 10^{-32} that we can ignore the influence.

7 Conclusion

In this paper, we studied the problem of *subregion probability attack*. Based on this problem, we proposed a novel Hierarchical Hilbert Curve Spatial Cloaking Scheme which satisfies reciprocity and uniformity. Meanwhile we also proposed a optimal splitting bucket strategy during selecting dummy positions. Aforementioned works are based on a system structure of cooperative group which overcomes the demerits of TTP and P2P. The security analysis and experiment results have proved our proposed method can achieve an effective performance.

Acknowledgments. The work is partially supported by the NSF of China for Key Program (No. 61532021), the NSF of China for Outstanding Young Scholars (No. 61322208), and the NSF of China (Nos. 61173031, 61572122, 61272178).

References

1. Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 31–42. CA (2003)
2. Rao, U.P., Girme, H.: A novel framework for privacy preserving in location based services. In: Fifth International Conference on Advanced Computing and Communication Technologies (ACCT), pp. 272–277. IEEE (2015)
3. Bamba, B., Liu, L.: PRIVACYGRID: Supporting Anonymous Location Queries in Mobile Environments. J. GEORGIA INST OF TECH ATLANTA COLL OF, COMPUTING, pp. 7–17 (2007)
4. Bamba, B., Liu, L., Pesti, P., et al.: Supporting anonymous location queries in mobile environments with privacygrid. In: International Conference on World Wide Web, pp. 237–246. ACM (2008)
5. Kalnis, P., Ghinita, G., Mouratidis, K., et al.: Preventing location-based identity inference in anonymous spatial queries. IEEE Trans. J. Knowl. Data Eng. **19**, 1719–1733 (2007)
6. Lee, H.J., Hong, S.T., Yoon, M., et al.: A new cloaking algorithm using Hilbert curves for privacy protection. In: ACM Sigspatial International Workshop on Security and Privacy in GIS and LBS, pp. 42–46. ACM (2010)
7. Ghinita, G., Kalnis, P., Skiadopoulos, S.: PRIVE: anonymous location-based queries in distributed mobile systems. In: WWW 2007: International Conference on World Wide Web, pp. 371–380. IEEE (2007)
8. Lu, H., Jensen, C.S., Yiu, M.L.: PAD: privacy-area aware, dummy-based location privacy in mobile services. In: The Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access, pp. 16–23. ACM (2008)
9. Niu, B., Li, Q., Zhu, X., et al.: A fine-grained spatial cloaking scheme for privacy-aware users in Location-Based Services. In: Computer Communication and Networks (ICCCN), pp. 1–8. IEEE (2014)
10. Niu, B., Li, Q., Zhu, X., et al.: Achieving k-anonymity in privacy-aware location-based services. In: INFOCOM, pp. 754–762. IEEE (2014)
11. Gkoulalas-Divanis, A., Kalnis, P., Verykios, V.S.: Providing K-Anonymity in location based services. J. ACM SIGKDD Explor. Newsl. **12**, 3–10 (2010)
12. Hossain, A., Hossain, A.A., Jang, S.J., et al.: Privacy-aware cloaking technique in location-based services. In: IEEE First International Conference on Mobile Services, pp. 9–16. IEEE (2012)
13. Mokbel, M.F., Chow, C.Y., Aref, W.G.: The new Casper: query processing for location services without compromising privacy. In: The 32nd International Conference on Very Large Data Bases, pp. 763–774. VLDB (2006)

Springer

http://www.springer.com/978-3-319-39957-7

Web-Age Information Management
17th International Conference, WAIM 2016, Nanchang,
China, June 3-5, 2016, Proceedings, Part II
Cui, B.; Zhang, N.; Xu, J.; Lian, X.; Liu, D. (Eds.)
2016, XXII, 552 p. 203 illus., Softcover
ISBN: 978-3-319-39957-7