This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

A K-Anonymity Based Schema for Location Privacy Preservation

Fan Fei, Student Member, IEEE, Shu Li, Student Member, IEEE,
Haipeng Dai, Member, IEEE, Chunhua Hu, Member, IEEE,
Wanchun Dou, Member, IEEE and Qiang Ni, Senior Member, IEEE

Abstract—In recent years, with the development of mobile devices, the location based services (LBSs) have become more and more prevailing and most applications installed on these devices call for location information. Yet, the untrusted LBS provider can collect these location information, which may potentially threaten users' location privacy. In view of this challenge, we propose a two-tier schema for the privacy preservation based on $k-$anonymity principle meanwhile reduce the cost for privacy protection. Concretely, we divide the users into groups in order to maximize the privacy level and in each group one proxy is selected to generate dummy locations and share the returned results from LBS provider; then, on each group, an auction mechanism is proposed to determine the payment of each user to the proxy as the compensation, which satisfies budget balance and incentive compatibility. To evalue the performance of the proposed schema, a simulated experiment is conducted.

Index Terms—LBS, privacy preservation, k-anonymity, auction

1 Introduction

Recently, smart phones and other portable network accessible devices have gained great popularity for its convenience. LBSs have been more and more prevailing and became a new service paradigm. Generally, users with smart phones download the applications from Google Play or Apple Store and install these applications on their smart phones to enjoy a variety of location-based services. Generally, this procedure is performed that users submit a query with their location information to the LBS provider and get their point of interests within a range. For instance, they can query the hotel price in the neighborhood, the evaluation towards a restaurant, the nearby hospital, etc.

Although users benefit a lot from the convenience of LBSs , users are subject to untrusted location based service provider. These attackers can mine the collected location information of users, together with other side information from public website such as twitter, facebook, yelp, etc, and infer the true identity of the user. Thus, LBS users are faced with severe privacy exposure.

To protect the location privacy, many methods have been proposed to make it difficult for the attacker to accurately infer the true location of users. These methods mainly include data transformation [1], [2], k-anonymity [3], [4], mix zone [5] [6], dummy location [7], differential privacy [8] [9], etc.

- *Fan Fei, Shu Li, Haipeng Dai and Wanchun Dou are with the State Key Laboratory for Novel Software Technology and the Department of Computer Science and Technology , Nanjing University, China E-mail: ffwilliam1992@gmail.com, shuli@smail.nju.edu.cn, haipengdai@nju.edu.cn, douwc@nju.edu.cn. Wanchun Dou is the corresponding author.*
- *C. Hu is with the School of Computer and Information Engineering, Hunan University of Commerce, Changsha 410205, China (e-mail: huchunhua777@ 163.com).*
- *Q. Ni is with School of Computing and Communications, Lancaster University, Lancaster, UK. (e-mail: q.ni@lancaster.ac.uk)*

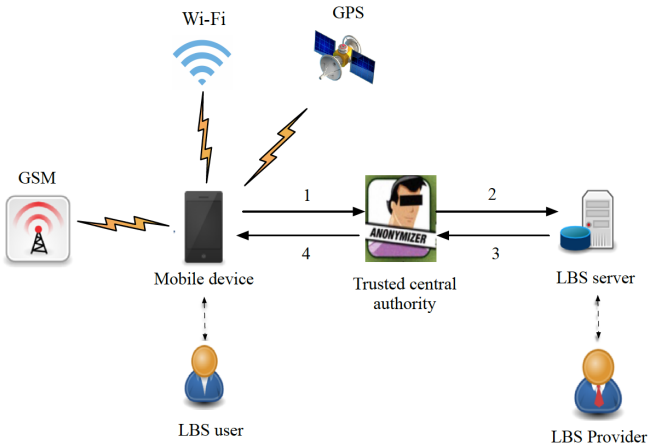


Fig. 1. The procedure of the LBS with k-anonymity: 1. users submit LBS requests with pseudonyms and CA collects these locations from users; 2. CA obtains a minimum bounding region and the region is submitted to LBS server; 3. LBS server returns the search results to CA; 4. CA removes the false results and returns the final results to the user.

Among these methods, k-anonymity method is extensively deployed in the existing researcher and has been shown its effectiveness [3], [4]. The main idea behind k-anonymity based methods is that the true location is screened by the other k-1 dummy locations and thus add uncertainty to the inference of the attacker. Generally, k-anonymity based method is based on the assumption that there exists a high-performance trusted Central Authority (CA). Fig.1 shows, in the procedure of LBSs, the CA acts as the anonymizer who collects the locations of different k users and utilize spatial-temporal cloaking or location obfuscation techniques to obtain a minimum bounding region [10]. Next, the CA submit this pack containing the

obfuscated location information to the LBS server. Finally, the service provider processes the input query information and returns to the CA a collection of candidate results. After removing the false result, the CA sends the final results back to the customer.

To some degree, these methods surely protect location privacy. However, these methods would also leads to the location information coarse-grained, which will more or less degrades the quality of service. Moreover, these methods are inevitably faced with the single-point failure problem. If the central authority breaks down or it cannot process all the incoming requests, the privacy of users cannot be guaranteed. Attackers can also obtain side information, such as the history query probability of each PoI. With the help of these side information, the attacker can easily eliminate locations with little query probability and make ineffective the traditional k-anonymity based privacy protection methods.

To protect the location privacy of the customer and meanwhile guarantee the quality of service, R.Lu et al. [11] proposed *PLAM* framework which did't involve the central authority. X.Zhu *et al* [12] proposed a collaborative system "MobiCache", which improved the hit ratio and reduced the cost of privacy protection.

Generally, a user sends information such as his own id, his location, RoI (region of interest) and PoI (point of interest) to the LBS provider. To save his privacy, he generate dummy locations and dummy PoIs and enlarged RoI. Our work is based on the following observations: i) the enlarged query scope make useless the majority of the returned results, which is a great waste of resources; ii) the user in the vicinity may well require the similar LBS and a content sharing can decrease the cost on the privacy protection; iii) when a user acts as proxy and shares the data with other users, it is not reasonable that the user undertakes all the cost of generating the dummy location. To stimulate user to undertake the role as a proxy, his cost of dummy location generation should be compensated and too much compensation would increase the cost of privacy protection.

In this paper, we propose a two-tier cost sharing mechanism to preserve location privacy. The contributions that we make in this paper are listed as follows:

- We propose a method to divide users into group in order to share data with users in vicinity and meanwhile maximize the entropy. The proposed group division method is immune to the inference attack from attachers with side information.
- To motivate people to be the proxy, generate dummy information and share data with others, we propose a cost sharing mechanism based on auction.
- We prove that the proposed auction mechanism satisfies budget balance and incentive compatibility.
- We conduct an experiment evaluate the performance of our proposed schema.

The remainder of our paper is organized as follows: Section 2 presents the relate work; Section 3 models the scenario of our problem and gives a brief introduction of the preliminary knowledge ; Section 4 presents the basic idea of our proposed schema and formally defines the problem; Section 5 and Section 6 presents the algorithm and gives a theoretical analysis about the solution to the problem and the evaluation of the proposed method is presented in Section 7; Section 8 concludes this paper.

2 Relate work

In this section, we review the existing researches on privacy protection of LBSs and the related incentive mechanism.

2.1 Privacy Metrics

Privacy metric selection is vital in the design of privacy protection algorithm in that the metric is generally the objective function to optimize. It more or less reflects the design principle. One major principle is that how accurately the attacker can infer the user's true location. A better algorithm should make it harder for the attacker to determine the true location. Entropy is widely utilized [13] as a measure of the uncertainty to determine the true location of a user. Other metrics, such as k-anonymity [14], t-closeness [15], [16], and other variations, measure the privacy level according to the size of anonymity set.

The above metrics are uncertainty based. B.Niu *et al.* used another metric, CR (Cloaking Region), which is based on the intuition that a larger cloaking region can make it more difficult for the attacker to determine the true location of a user. However, in our work, we concern primarily on the privacy level when dividing users into groups, we adopt entropy to measure the location privacy.

2.2 Privacy Protection Methods

Most of the exsisting researches on promoting location privacy level is to add uncertainty to the inference of attackers. These methods includes pseudomyzation [17]. data transformation [1], [2], k-anonymity [3], [4], [18], mix zone [5] [6], dummy location [7], differential privacy [8] [9].

Among these methods, k-anonymity is the mainstream in location privacy preservation and has enjoyed great success. [3], [4] proved that k-anonymity method can guarantee the security of LBS users. These methods, generally assume that there exists a trusted and high-performance central authority who acts as the broker between the service provider and users. LBS provider can also find a third party server such as cloud server and store the LBS data there by a data as a service (DaaS) model [19] [20]. This service approach avoid the necessity of a single central authority and therefore is free from the single point failure. Generally an attribute-based encryption [21] [22] scheme is explored in some LBS system to achieve fine-grained access control on the LBS data.

X. Zhu et al [12] proposed a concept *MobiCache* considering the spatio-temporal relation among users and designed a collaborative system to increase the hit ratio. A. K. Tyagi et al. [23] concerned about the location privacy problem in vehicular ad hoc networks (VANETs) and proposed a solution based on mix zone. Qiu et al. [24] and K.Vu et al. [25] studied the privacy problem in the participatory sensing scenario.

Other researches concern the privacy protection when the attackers have the side information and launch an inference attack towards users. B.Niu et al. [26] proposed a dummy location selection method, which can overcome

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

the single-point failure and does not rely on the centralized trusted anonymizer. In their work, the final dummy locations are selected according to entropy and area of cloaking area. D.Liao et al. [27] proposed a greedy-based method, $k-$DLCA, to select dummy locations, taking into consideration the semantic information of these locations.

2.3 Incentive Mechanism

An incentive mechanism is crucial in the location privacy preservation, especially in the scenario that the protection method goes in a collaborative way. Feng et al. [16] designed a truthful auction for location-aware collaborative sensing in mobile crowdsourcing. Yang et al. [28] proposed auction-based motive mechanisms to incentive users to help others achieve k-anonymity. [29] proposed an auction based mechanism which stimulates those who are indifferent to privacy protection to participate in the anonymity set. X.Liu et al. [30] model the dummy generation process using game theory and the existence of Bayesian Nash Equilibria (BNE) was analyzed. In our work, we hope to stimulate users to undertake the task of dummy location generation meanwhile reduce the cost. A mechanism therefore should satisfy incentive compatibility. The incentive compatibility [31] can be categorized into dominant strategy incentive compatibility (DISC) and Bayesian incentive compatibility (BIC).

3 System Modeling

3.1 Basic Knowledge

In this section, we introduce some basic knowledge about location privacy and several important concepts used in our paper and the metric for privacy preservation.

An untrusted location-based service provider may well utilize the position information from the query which is generally in a *JSON* format. In this paper, we emphasize the situation where the untrusted location-based service provider proposes an *inference attack* towards users by a probabilistic model which utilizes **side information**. There are a variety of side information that an attacker can collect and utilize, such as posts of twitter, a registration to certain location, query information and so forth, which users post voluntarily. In this paper, to avoid ambiguity, we give a explicit definition of side information.

Definition 1. *The **side information S** refers to a user's query probability in an area, which is counted according to history records. It reflects the popularity degree of a certain location in history.*

In this paper, an **entropy-based privacy metric** [32] is adopted to measure the degree of anonymity for LBSs. Entropy reflects how certain attackers can recognize the true location of the user from the anonymity candidate set.

$$q_i = \frac{\#\ of\ queries\ in\ location\ i}{\#\ of\ queries\ in\ all\ locations} \tag{1}$$

To calculate the entropy, the query probability of each dummy location is needed and can be calculated according to equation (1). This process calls for the history record data of all the interested locations, which reflects semantic meanings. In the dummy location set $D = \{l_1, l_2, \ldots, l_k\}$, the entropy of each user u therefore is defined as :

$$H(u) = -\sum_{i=1}^{k} p_i log_2 p_i, \tag{2}$$

where p_i is the normalized location probability so that $\sum_{i=1}^{k} p_i = 1$ and therefore p_i can be calculated by

$$p_i = \frac{q_i}{\sum_{i=1}^{k} q_i}. \tag{3}$$

The entropy reaches its maximum when the p_i is uniformly distributed, formally $\forall i, j, i \neq j, i, j = 1, 2, \ldots, k$ we have $p_i = p_j$. In this situation, the optimal entropy is identically $H(u) = klog_2 k$. Another perspective supporting this idea is that the real location of a user is perfectly protected by the dummy locations with the same distribution probability.

3.2 System Model of LBS

A typical LBS system usually consists of two main components, i.e., LBS users and LBS providers. In detail, LBS users utilize sensors on mobile devices to obtain the accurate location information and then submit it to the LBS providers for location based service. Next, LBS providers search the location related information and recommend the corresponding results to LBS users. During this process, in order to protect location privacy, LBS users usually adopt the pseudonyms to hide their true identities. Unfortunately, this simple measure cannot fully prevent the leakage of location privacy of LBS users, since untrusted LBS provider may make full use of the side information to infer and confirm the location information with the true identity.

For a user who asks for LBS , the query mainly consists of the following elements [27]:

- **identity**, which is the unique identification of a user and denoted by u_{id};
- **position**, which is generally represented by location coordinate $l = (x, y)$;
- **range of RoI (Region of Interest)**, which is the area a user concerns about and is denoted by R;
- **PoI (Point of interest)**, which is the concrete point category a user cares about, such as a restaurant, a hospital, a movie theater, etc and denoted by P.

Therefore, a query Q can be viewed as a 4-tuple, formally $Q = \{u_{id}, (x, y), R, P\}$. To implement the $k-$anonymity of the location, $k-1$ dummy locations are generated and denoted by $\{l_1, l_2, \ldots, l_{k-1}\}$ and k is determined by the privacy level that a user needs. A larger k can ensure a higher privacy level. When the user's real RoI is R, he can enlarge his region of interest to $\bar{R}$ to hide his real needs and therefore protect his own privacy. The location of a user is readily subject to some sensitive location query, because of the limited number of these places such as hospitals, gas stations, etc. In this situation, the user can expand his PoI into a larger set $\{P_1, P_2, \ldots, P_m\}$. Therefore, the final query is $\bar{Q} = \{u_{id}, \{l_1, l_2, \ldots, l_k\}, \bar{R}, \{P_1, P_2, \ldots, P_m\}\}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

3.3 Adversary Modeling

Suppose that there is an attacker who attempts to connect the anonymous location with the true identity of a LBS user. In general, the attacker firstly collects side information and then launches an inference attack. In detail, we restrict the side information as location query probability. And the attacker may also have the knowledge about the mechanism how users protect their privacy.

Meanwhile, we suppose that the untrusted LBS provider won't actively track the users directly and accumulate accurate side information of certain users. The LBS provider can query its log database and know the history probability of each positions which users ask for LBS requests. The provider can also monitor the current service requests with location information.

Let G_i denotes the dummy locations a user u_i sends to the LBS server. An inference attack is to utilize the probability $Pr(l_i|G_i)$ to measure the similarity between each potential position l_i and side information S. We assume that the priori knowledge is derived from the uniform distribution, i.e., $Pr(l_i) = \frac{1}{k}$. Besides, the side information is independent on user i. Therefore, according to the Bayesian theorem, we have

$$Pr(l_i|G_i) = \frac{P(G_i|l_i))Pr(l_i)}{Pr(G_i)}. \tag{4}$$

The adversary meanwhile knows exactly about the mechanism how the users protect their privacy. The users' location privacy are leaked when the adversary is able to infer the users' real location from the location information. Moreover, when the LBS server is compromised by the adversary, the adversary can get all private information stored in the LBS server. **Therefore, in this paper we suppose that the LBS provider is the adversary**.

4 Problem Statement

4.1 Motivation

A typical LBS procedure goes as following: a user sends a query to the LBS provider, with information such as identity, PoI, RoI and the GPS position, etc. A $k-$anonymity method is adopted by collecting the information of other $k-1$ users nearby. These information are first sent to a trusted anonymizer and the anonymizer replaces these exact locations by a cloaking region. There are mainly three weaknesses of this method. First, this privacy protection method is subject to the single-point failure problem. If the anonymizer breaks down, the privacy of users cannot be guaranteed. Second, the anonymizer may well be the bottleneck of the system especially when there are a large quantity of incoming service requests. Last but not the least, the quality of service is degraded by the lower accuracy of provided position.

Fig.2 shows, in an area paved by hexagons, the gray-colored hexagons represent popular locations while the hallow hexagons represents locations with lower query possibility. To protect the true location, the user in location 1 generates 2 other dummy locations. However, the dummy position 2 and 3 have a very low query probability and in reality, this position may be lakes, mountain, etc. that a user can never reach. The attacker can easily filter this kind of

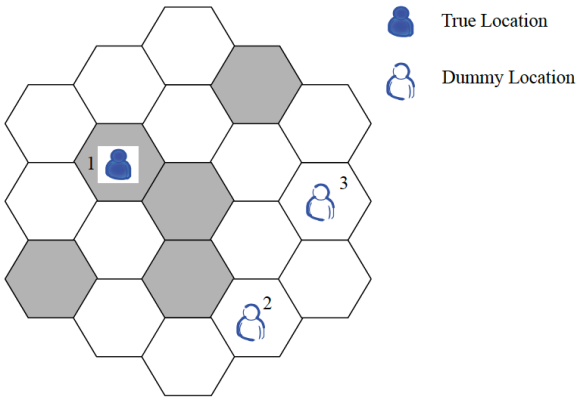


Fig. 2. A motivation example

dummy locations with the help of side information. In our case, the $3-$anonymity is reduced into a $1-$anonymity and therefore the user's true location is exposed to the attacker. When a user asks for very sensitive information such as *where is the nearest gas station*, the user may well add other dummy information such as dummy PoI, dummy RoI into the query to hide his true intention. The user can get many returned results that are useless.

Therefore, it is valid that the locations with similar query probability are sorted into the same group and camouflage each other. We also assume that, users in the same hexagon with certain range have the same returned results when their requests information contain the same PoIs and RoI. Once a user in such a group ask for a service request, he can act as a proxy and store the returned results in his own buffer for a while. Other users in the same group can search the buffer and fetch the matched data within hops. To better depict this problem, we solicit to the following definition.

Definition 2. *A **group G** is all active LBS sers on a certain number of hexagons, while these hexagons represent a subset of an area Ψ . In the following discussion, we use $G(\pi)$ denotes users on hexagons π and use $\pi(G)$ denotes the hexagons these users takes up.*

According to the definition of the group, it is easy to see that all users in the area Ψ is a group and a group G can be divided into n sub-groups $\Lambda = \{G_1, G_2, \ldots, G_d\}$ satisfying the two following property:

$$G = \bigcup_{i}^{n} G_i \tag{5}$$

and $\forall i, j \in \{1, 2, \ldots d\}$ and $i \neq j$,

$$G_i \bigcap G_j = \emptyset, \tag{6}$$

where Λ represents a division of the group G.

A group is the basic unit in our paper to preserve privacy and share returned results from the LBS provider. The definition of group also indicates that the division of users is accompanied with the division of area, which on the hexagon as a unit. The population information is important in the cost sharing phase. The population of each

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

hexagon is counted according to the location information the user sends. And as the population of users within an area changes slowly rather than drastically, we update these information every few minutes.

Based on these observations, a two-tier scheme is proposed in order to maximize the privacy level in a given area with the least cost.

4.2 Problem Formulation

The generation of $k-1$ dummy locations and sharing of the returned results with other users in the same group are inevitably accompanied with some cost (e.g. the communication, energy consumption, etc). The cost for the user i is denoted by c_i. In a group, it is sufficient that only one user acts as a proxy to generate the $k-1$ dummy locations and shares the returned results with others during a certain time interval. The cost of privacy preservation for group G_i is denoted by C_i. In our paper, we adopt an auction-based motive mechanisms to determine which user undertake the role as the proxy. Therefore, if u_j of group G_i is selected as a proxy, the cost of group G_i is equivalent to the cost of u_j, namely $C_i = c_j$.

We suppose that the number of hexagons which a group G takes up is no larger than a constant Π_{max}. This is reasonable because of the limited computation, communication and storage capacity of the selected user in the group.

With the rigorous definition and system modeling, here, we formally define our problem and reduce it to an optimization problem with several constraints.

The proposition for the problem are listed as following:

- an area Ψ, which is seamlessly paved by N same-sized and same-shaped hexagons,
- M users $\{u_1, u_2, \ldots, u_m\}$ who are active and call for LBSs,
- N hexagons $\{\pi_1, \pi_2, \ldots, \pi_n\}$, each hexagon π_i has α_i active users and $\sum_{i=1}^{N} \alpha_i = M$.
- time period T for result sharing,
- cost of each user,
- side information S.

The optimization problem is to maximize the cost effectiveness and therefore the objective function of our problem is

$$\max_{C^*, G^*} \quad \frac{\sum_{i=1}^{M} H_i}{\sum_{i=1}^{d} C_i}, \tag{7}$$

where $\sum_{i=1}^{d} C_i$ represents the total cost for all users and the numerator represents the sum of privacy level.

It is hard to satisfy the objective function (7) directly. The primary goal is to maximize the entropy and protect our users from the inference attack of adversaries. Therefore, to solve this problem, we propose a two-tier schema: first, we propose a method to divide the group into d sub-groups in order to maximize the total entropy; second, we propose an auction-based mechanism to minimize the total cost.

4.3 Notations

Table 1 introduces the main symbols and notations used in our paper.

TABLE 1
Table of Key Notations

Notation	Description
Ψ	An area
π_i	The ith hexagon
u_i	The ith user
Π_{max}	The largest number of hexagon a group can take up
S	Side information
Q	A query containing basic information
$\bar{Q}$	A query with dummy information
T	Time period for sharing result
R	Region of interests
P	Categories of point of interests (e.g. restaurant, hospital, etc.)
G	A group, which is made up of users on several π
Λ	A division of group G.
α_i	The population of users taking up the ith hexagon
l	Location information, which is made up of two coordinate
H	The entropy, which represents the privacy level
q_i	The query probability of the ith PoI
p_i	The normalized query probability of the ith PoI
c_i	cost of generating $K-1$ dummy locations for u_i
C_i	cost of generating dummy locations for group G_i

5 A Group division Algorithm

In this section, we propose an group division method to maximize the privacy level as a whole. Meanwhile, we demonstrate how the proposed method is immune to the inference attack from the untrusted LBS provider.

5.1 Proposed Algorithm

In this subsection, we propose a greedy-based algorithm to maximize the total entropy. Our objective function is to maximize the numerator of equation (7), and since the users in the same group share the same privacy level, therefore, the objective function can be rewritten as following:

$$\max_{\Lambda^*} \quad \sum_{i=1}^{d} |G_i| H_{G_i} \tag{8}$$

$$s.t. \quad \sum_{i=1}^{d} |G_i| = M,$$

$$|\pi(G_i)| < \Pi_{max}, i = 1, 2, \ldots, d,$$

where $|G_i|$ represents the number of users in the subgroup G_i and $|\pi(G_i)|$ represents the number of hexagons these users take up. The entropy for the group G_i can be readily computed by the equation (2), (3).

When partitioning the users, we set the hexagon with more population a higher priority. We first sort the hexagons according to the number of users that the hexagon takes up. Formally, the original ordered set $\{\pi_1, \pi_2, \ldots, \pi_N\}$ is mapped to the set $\{\pi_1^{'}, \pi_2^{'}, \ldots, \pi_N^{'}\}$, and $\forall i, j \in \{1, 2, \ldots, N\}, i > j$, we have $G(\pi_i^{'}) < G(\pi_j^{'})$.

Algorithm 1 demonstrates the concrete procedure to generate the final group division scheme Λ. Because our partitioning method is designed to protect the privacy of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

Algorithm 1 Partitioning the hexagons into groups

Require: The side information S, the hexagons set $\{\pi_1, \pi_2, \ldots, \pi_N\}$, the population of each hexagon $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$
Ensure: final group set Λ
1: $L \leftarrow \emptyset$; //initialize a list.
2: $\Lambda \leftarrow \emptyset$; // initialize the final group set.
3: **for** $i \leftarrow 1$ to N **do**
4: **if** $\alpha_i > 0$ **then**
5: $L.append(\pi_i)$;
6: **end if**
7: **end for**
8: $n \leftarrow L.length; V \leftarrow L$; // V represents the candidate sets
9: $sort(L)$; //sort the list according to the population with descending order.
10: $mark \leftarrow \{True\}_n$; //initialize an array with Boolean type.
11: **for** $i : 0 \rightarrow n-1$ **do**
12: **if** $mark[i] == True$ **then**
13: /* call for algorithm to determine group G containing the current hexagon $_*$/
14: $G \leftarrow select(L[i], V)$;
15: **for** π in G **do**
16: $mark[\pi] = False$; //mark all the selected hexagon False
17: **end for**
18: $V \leftarrow (V - G)$;
19: $\Lambda.append(G)$; //add group G to the division set.
20: **else**
21: continue;
22: **end if**
23: **end for**
24: **return** Λ;

users, these hexagons with no population is not taken into consideration. In lines $3-7$, we add the nonempty hexagons to a List. Then, Lines $11-23$ shows that the hexagon with a higher priority begins to select hexagons which are not selected by other users. In Line 9, the hexagons are sorted in descending order according to the population and thus, the hexagon with the highest population select hexagons first. In line 14, a hexagon selects these ungrouped hexagons and form a new group. This process is implemented in a greedy algorithm.

Algorithm 2 shows the concrete process to determine the final group that best matches an input hexagon π. In lines $4-9$, we get the hexagon candidates first. In the sorted list containing the input hexagon, we derive the Π_{max} hexagons left the hexagon and Π_{max} hexagons right the hexagon as the candidates. This is based on the observation that the entropy in equation (2) reaches its maximum when each of the probability are equal, which implies that hexagons with similar query probability can better protect each other. This process reduces the search space and reduces the complexity of the algorithm

On line 10, we randomly select $\beta\Lambda_{max}$ candidates where β is the control parameter ranging from 1 to 2. This process is designed to add some randomness to the result for fear of these attackers who know the privacy protection mechanism and infer the true location of the user with the generated dummy locations.

In line $11-22$, the selection of hexagons goes in a greedy-based manner. The hexagon is added to a group one by one and these which can better augment the entropy of the group are added to the group first.

Algorithm 2 The Greedy Algorithm to select hexagon

Require: The input hexagon π,the query probability for each hexagon π is $p(\pi)$, the left ordered hexagon list $V = \{\pi_1, \pi_2, \ldots, \pi_k\}$
Ensure: the final group G
1: $G \leftarrow \emptyset$; //initialize a list
2: $G.append(\pi)$;
3: $count \leftarrow 1$;// count the number of hexagon in G
4: $V^{'} \leftarrow \emptyset$; // initialize the candidates set
5: index $\leftarrow$GetIndex(π, V);
6: **for** i : $1 \rightarrow \Pi_{max}$ **do**
7: $V^{'}.append(\pi_{index+i})$;
8: $V^{'}.append(\pi_{index-i})$;
9: **end for**
10: $V' \leftarrow V^{'}.RandomSelect(\beta\Lambda_{max})$;
11: **while** $count < \Pi_{max}$ **do**
12: bestcandidate $\leftarrow 0$; maxentro$\leftarrow 0$;
13: **for** π_i in $V^{'} - G$ **do**
14: entro $\leftarrow H(G \cup \pi_i) - H(G)$;
15: **if** entro>maxentro **then**
16: maxentro$\leftarrow$ entro;
17: bestcandidate$\leftarrow \pi_i$;
18: **end if**
19: **end for**
20: $G \leftarrow G \cup$bestcandidate;
21: $count \leftarrow count + 1$;
22: **end while**
23: **return** G;

5.2 Security Analysis

In this subsection, we prove that our proposed method is immune to the inference attack from the untrusted LBS provider. In the following discussion, the attacker is restricted to the untrusted LBS provider with side information. The attacker knows the query probability of the given area and also the mechanism of how the dummy locations are generated and how the groups are formed. The true location is free from the inference attack when it is not distinguishable from the dummy locations. The proposed method divides the area by hexagons and when we infer the hexagon that a user on, we get the true location of the user. Formally, to prove our proposed method is reasonable, we turn to the following definition.

Definition 3. *A privacy protection method is* ***inference attack resistant*** *from the untrusted LBS provider if* $\forall i, j, i \neq j$ *and* $i, i = 1, 2, \ldots, k$, *we have*

$$Pr(\pi_i \in \Lambda | \Lambda \subset \Psi) = Pr(\pi_j \in \Lambda | \Lambda \subset \Psi) \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

To prove the proposed method is free from the inference attack from the untrusted LBS provider, according to the definition (3), we have

$$Pr(\pi_i \in \Lambda | \Lambda \subset \Psi) = \frac{Pr(\pi_i \in \Lambda \cap \Lambda \subset \Psi)}{Pr(\Lambda \subset \Psi)} = \frac{q_i}{Pr(\Lambda \subset \Psi)}.$$

Since $\pi_i \in \Lambda$ and $\pi_j \in \Lambda$, the method is effective when $q_i = q_j$ holds. As our proposed method is aimed at maximize the entropy and this guarantee the equality between q_i and q_j.

In our method, the user get the information shared from the proxy and thus is free from the inference attack from the LBS provider. The attacker also understands the location privacy protection mechanism and knows how the groups are formed and thus infer the true location of the proxy. However, in algorithm 2, we add randomness to the dummy locations generating process and thus ensure the security of the proxy.

6 Cost Sharing mechanism between users

In this section, we start to solve the problem which user in the partitioned group undertakes the task to generate dummy locations and share the returned results with other users. A cost sharing mechanism is proposed to minimize the cost of generating dummy locations and meanwhile stimulate the users in the area to participate in the coalition. The proposed mechanism satisfies both incentive compatibility and budget balance.

6.1 Mechanism Design

For k-anonymity based methods, in our method, we select a user u_i from the group $G = \{u_1, u_2, \ldots, u_{|G|}\}$ to undertake the task of generating the dummy locations and sharing returned results. This process is based on the assumption that the selected user is trusted. Moreover, the task is inevitable accompanied with a certain amount of cost such as energy consumption, communication and computation. We use c_i denote these cost for u_i. In fact, it is not reasonable to let a user undertake all the cost. We propose a mechanism to determine which user undertake the task and receive payment from others in the group as the compensation.

In this paper, an auction mechanism is explored to determine which user in the group G is to undertake the task and how to allocate the cost. In our auction, LBS users are both bidders and auctioneers. The user who wins the auction can act as the proxy and undertakes the task. The mechanism is mainly composed of the following three steps:

- every participant in the auction reports his bids;
- the one with the least bid is selected as the winner;
- other users in the group give a certain amount of payment to the winner as the compensation.

The valuation of different users varies and this information is private in the auction. We use v_i denotes the private valuation of u_i. Therefore, we have $v_i = c_i$. Meanwhile, each user u_i has a reported bid b_i. We use vector $\vec{b} = \{b_1, b_2, \ldots, b_{|G|}\}$ denotes the reported bids of the group G and use b_{-i} denotes the profile of all the users except u_i, namely $b_{-i} = \{b_1, b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_{|G|}\}$. Hence, $\vec{b} = (b_i, b_{-i})$. In the auction, the winner is selected as the proxy and get payments from others.

Definition 4. *A **allocation function** $f_i : \Theta \rightarrow \{0, 1\}$ is a indicator function which determines whether or not user u_i is the winner, where Θ stands for all the possible combination of reported bid b_i, formally $\Theta = \Theta_1 \times \Theta_2 \times \ldots \times \Theta_{|G|}$.*

Definition 5. *A **payment function** $g_i : \Theta \rightarrow \mathbb{R}^+$ denotes the payment the winner will receive from user i as a compensation.*

Let v^* denotes the final selling price of the winner. Since the target of our mechanism is to maximize equation (7), the final selling price of each group should be as little as possible. In order to achieve this goal, each user should report the true cost as the bid. To stimulate the user to report the real cost, the user should obtain the optimal utility when the user report his true preference. The following definition formally describes this property that the mechanism should meet.

Definition 6. *A mechanism satisfies **incentive compatibility** [31] if for any other reported bid $b_i^{'}$ and other LBS users strategy profile b_{-i}, there exists $b_i \neq b_i^{'}$ so that*

$$u_i(b_i, b_{-i}) \geq u_i(b_i^{'}, b_{-i}), \tag{10}$$

where $u_i(b_i, b_{-i})$ represents the utility u_i obtains when the reported bid of the group is (b_i, b_{-i}).

When we determine the compensation paid to the winner, on the one hand the compensation should be greater than the cost; on the other hand, we hope the cost is minimized. To describe this target, we turn to the following definition.

Definition 7. *A mechanism satisfies **budget balance** [31] if the sum of the net payments of all the LBS users is equal to zero. That is,*

$$\sum_{i=1}^{|G|} g_i(b) = 0. \tag{11}$$

According to this definition, a mechanism satisfying the budget balance means that it never takes a loss nor makes a profit.

6.2 Proposed Mechanism

To stimulate the users in a group to generate dummy locations and share data, a cost sharing mechanism satisfies incentive compatibility and budget balance is needed. the Arrow-d Aspremont-Gerard-Varet (AGV) mechanism [33] can satisfies the above two property. Therefore, we design a cost sharing mechanism based on AGV mechanism. The auction is divided into two key steps.

Fig.3 demonstrates the auction model and we suppose the auctioneer is the trusted independent third party [34]. First, all the users act as bidders and participate in the auction by submitting their reported bids $\vec{b} = \{b_1, b_2, \ldots, b_{|G|}\}$ to the auctioneer. The first step is to determine the winner who undertake the task and the auctioneer is the trusted

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

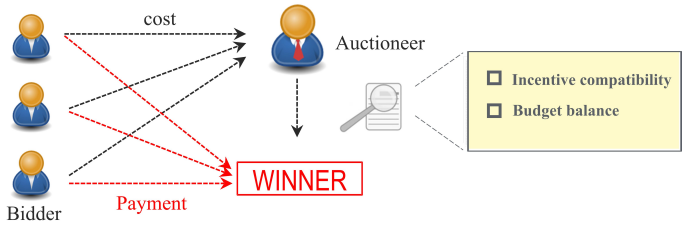


Fig. 3. Auction model with special property

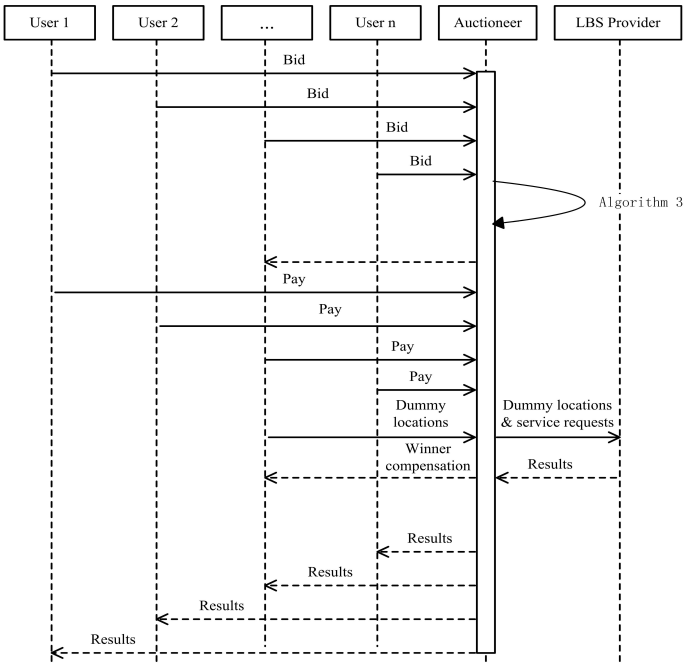


Fig. 4. The execution procedure of the cost sharing mechanism

independent third party who compares the bids and decides the winner. Since we just need to select one user to undertake the privacy protection task and any user in the group share the same privacy level, the user with least bids is preferred. Therefore, the allocation function $f_i(b)$ can be calculated by the following formula:

$$f_i(\vec{b}) = \begin{cases} 0 & \text{if } b_i > \min_{j \neq i}\{b_j\}, \\ 1 & \text{if } b_i \leq \min_{j \neq i}\{b_j\}. \end{cases} \tag{12}$$

Equation (12) illustrates that u_i is the winner when $f_i(\vec{b}) = 1$, otherwise not. Thus, the user with least reported bid is selected when every user can provide the same privacy level. When there exists two user submit the same reported bid, we just select one randomly as the final unique winner.

The second key step is to determine the payment for each user in the group as the compensation for the cost of the winner, who generates dummy locations and acts as the proxy to share returned results with other users. One the one hand, these payments should compensate for the cost of the winner; one the other hand, to minimize the cost of privacy protection, the mechanism should satisfies the budget balance. As is mentioned in [35], an incentive mechanism satisfies incentive compatibility does't necessarily satisfies the budget balance, especially for the strong degree of incentive compatibility. However, other mechanisms, such as AVG, satisfies these two property at the same time. In our situation, the v_i is a random variable and follows some distribution D_i. Moreover, we refers to a term, the expected social welfare, which is given as

$$ESW_{-i}(b_i) = E_{b_{-i}}[\sum_{j \neq i}^{|G|} v_j f_j(\vec{b})]. \tag{13}$$

Then, the payment for each user u_i can be calculated by the following reverse formula:

$$g_i(\vec{b}) = ESW_{-i}(b_i) - (\frac{1}{|G|-1}\sum_{j \neq i}^{G} ESW_{-i}(b_j)) \tag{14}$$

For the winner, $g_i > 0$, otherwise, the user should pay a certain amount of money as the compensation. As a result, the proposed auction mechanism balance the budget and ensure the interests for those who generate dummy locations. The concrete execution process of the proposed mechanism is demonstrated in the Fig.4. Algorithm 3 presents the concrete process about the cost sharing. The algorithm mainly contains two section: in line $4-9$, the winner is selected; in line $10-13$, the payment of each user in the group is computed.

Algorithm 3 The cost sharing mechanism

Require: Set of users $G = \{u_1, u_2, \ldots, u_{|G|}\}$, set of reported bids $\vec{b}$
Ensure: The winner w and the cost allocation vector $\vec{g}$
1: $w \leftarrow 0$;
2: $\vec{g} \leftarrow \{0\}_{|G|}$;
3: $\alpha \leftarrow +\infty$;
4: **for** $i : 1 \rightarrow |G|$ **do**
5: **if** $b_i < \alpha$ **then**
6: $w \leftarrow i$;
7: $\alpha \leftarrow b_i$;
8: **end if**
9: **end for**
10: **for** $i : 1 \rightarrow |G|$ **do**
11: $ESW_{-i}(b_i) \leftarrow E_{b_{-i}}[\sum_{j \neq i}^{|G|} v_j f_j(\vec{b})]$;
12: $g_i \leftarrow ESW_{-i}(b_i) - (\frac{1}{|G|-1}\sum_{j \neq i}^{G} ESW_{-i}(b_j))$;
13: **end for**
14: **return** $w, \vec{g}$;

6.3 Analysis

The proposed mechanism satisfies both incentive compatibility and budget balance, which is given by the following theorems, respectively.

Theorem 1. *Our proposed cost sharing mechanism satisfies the Bayesian incentive compatibility.*

Theorem 2. *Our proposed cost sharing mechanism satisfies the budget balance.*

The concrete proof the the Theorem (1) (2) is omitted. The theorem 1 demonstrates that our proposed mechanism can ensure that the winner who undertakes the task won't lose his interest. It is important to stimulate the users in the group to participate in the auction. Theorem (2) indicates that the auction won't bring in extra income for the winner, which minimize the average cost for the privacy protection.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

7 Experiment Evaluations

In this section, we conduct an experiment and evaluate the performance of the proposed method.

7.1 Dataset Description

To simulate the mobility, we use the dataset *GeoLife* [36], which is shown in Fig.5. This dataset is composed of the GPS trajectory data ,which was collected for over five years by MSRA (Microsoft Research Asia) with 182 participants. These trajectories are sorted according to the user and for each user, a trajectory is saved in a file. The trajectory of each user is organized as a series of 4-tuple, namely the latitude, longitude, altitude and the time stamp, which is ordered by time stamp.

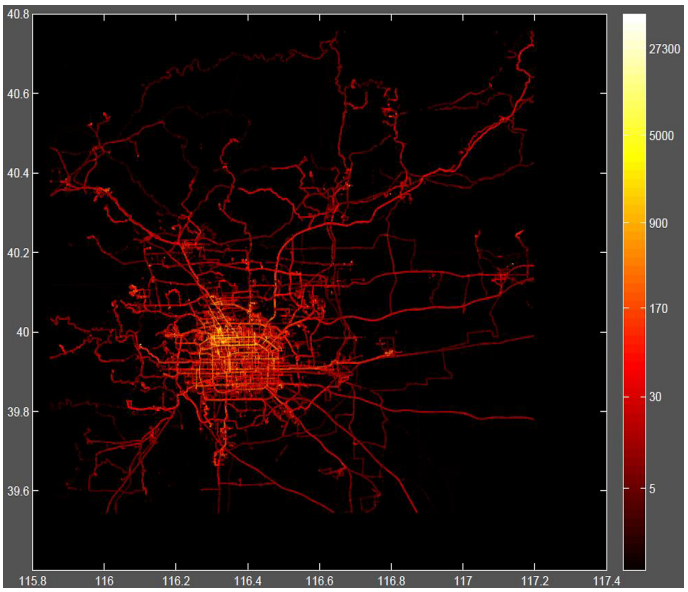


Fig. 5. Dataset overview in Beijing [36]

To approximate the history query to the LBS server, we use the dataset from Dianping [37]. Without loss of generality, we use the total number of comments to the PoIs in certain area to approximate the query probability of LBS users. Fig.6 shows the PoIs in Beijing is not uniformly distributed, the areas in the circles have higher density of the PoIs.

7.2 Simulation Setup

In the experiments, we select users and PoIs within a rectangle area of which the latitude ranges from 39.949 to 39.994 and the longitude ranges from 116.304 to 116.384. For the side information of inference attack, we use the total comments to PoI in a certain area as the query probability. The query probability of these areas without comment record is set as 0 by default.

The parameter k is related with the privacy level in the $k-$anonymity method and in our experiment the value of k ranges from 2 to 10. The cost of each user to generate dummy locations follows a uniform distribution, ranging from 10 to 20.

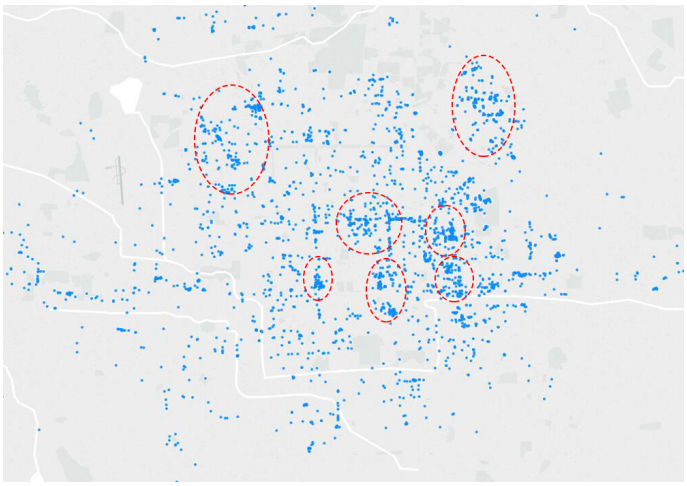


Fig. 6. Distribution of PoIs in Beijing

Technically, we conducted our experiment in a HANA cluster environment [38] . The proposed services are distributed in the cluster. As a result of making full potential of the memory database, our method can be implemented in a short response time.

TABLE 2
The experiment settings

	Client	HANA Cluster
Hardware	Lenovo ThinkpadT430 machine with Intel i5-3210M 2.50GHz processor, 8GB RAM and 250GB Hard Disk.	**Master** (1 node): HP Z800 Workstation Intel(R)Multi-Core X5690 Xeon(R), 3.47GHz/12M Cache, 6cores, 2 CPUs, 128GB (8x8GB+4X16GB) DDR3 1066MHz ECC Reg RAM, 2TB 7.2K RPM SATA Hard Drive **Slave** (1 node): HP Z800 Workstation Intel(R) Multi-Core X5690 Xeon(R), 3.47GHz/12M Cache, 6cores, 2 CPUs, 1066MHz ECC Reg RAM, 128GB (8x8GB+4X16GB) DDR3 2TB 7.2K RPM SATA Hard Drive
Software	Windows 7 Professional 64bit OS and HANA Studio.	SUSE Enterprise Linux Server 11 SP3 and SAP HANA Platform SP07

7.3 Evaluation Result

First, we focus on the privacy protection performance of the different dummy locations generation methods. We use a certain selected location as reference. We use the method proposed in [39], which randomly selects dummy locations. As as shown in Fig.7, our proposed method approximates the optimal condition and outperforms the baseline. The optimal condition comes when each of the dummy locations has the same query probability and the probability that a user exposes to privacy leaks is identically $\frac{1}{k}$. Our method approximates the baseline because hexagons with similar

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

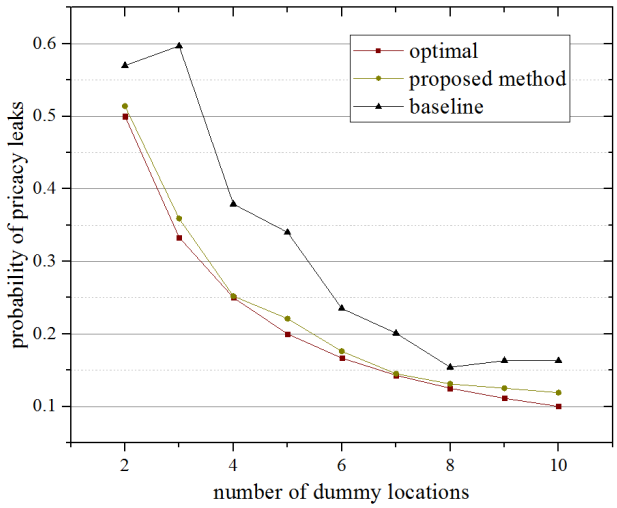


Fig. 7. The probability of exposing location privacy leaks with different number of generated dummy locations

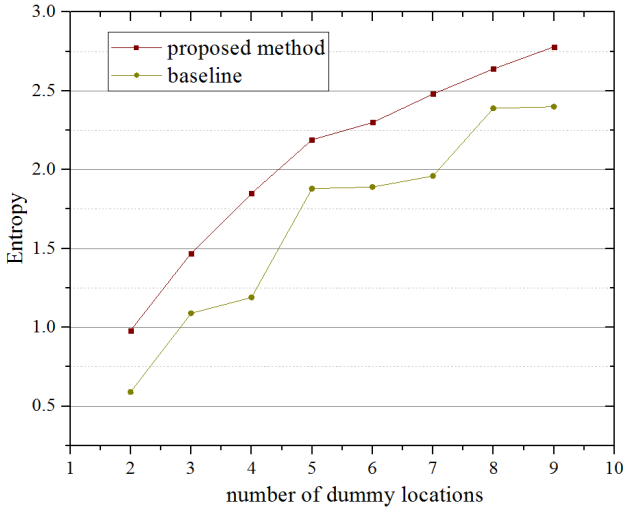


Fig. 9. The average entropy as the size of each group rises

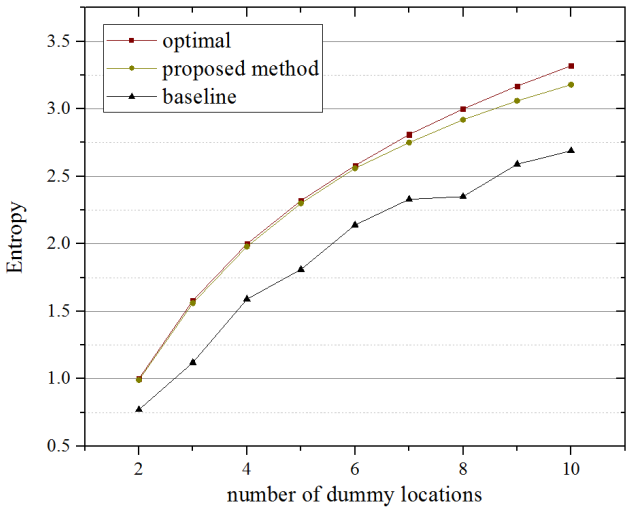


Fig. 8. The entropy with different number of generated dummy locations

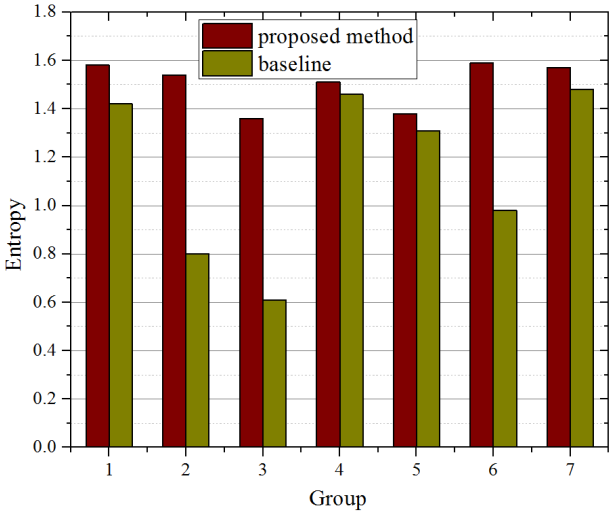


Fig. 10. Entropy reached of each group using $3-$anonymity

probability are selected as the dummy locations to each other. Moreover, as more dummy locations are generated, all of the three methods can effectively alleviate the probability of exposing to privacy leaks. Fig.8 demonstrates that our proposed method can achieve a higher privacy protection level than that of baseline when a user generates the same number of dummy locations. Moreover, Fig.7 and Fig.8 also indicate that the proposed method is more stable when the number of dummy locations rises.

Second, we focus on the privacy protection performance of our division schema. The baseline divides the users into groups in a random way. Fig.9 shows that our proposed scheme can reach a higher privacy level when compared with baseline. Moreover, our greedy-based method can have a comparably more smooth improvement of entropy as the group size rises. The Fig.9 also indicates that whatever the group division method, a user on average can better obtain a higher privacy level when the user is in a group of bigger size. The high privacy level on average does't ensure the privacy level of each individual and a privacy protection method should not sacrifice the interests of each individual. We also concern the privacy level of each group. Fig.10 and Fig.11 show our proposed schema provide a privacy level to each group. The distribution of the entropy reached by our proposed method is more concentrated than that of baseline method. This is achieved by rationally partitioning hexagons according to the history query probability.

At last, we concern about the cost of the privacy protection. Given entropy, Fig.12 indicates that the cost decreases as the number of users increases. The probability of user with lower cost arises as the number of users increases. To measure the privacy protection performance for given budget, we use the P-C (Privacy level-Cost) ratio as the metric, which is defined as:

$$\frac{\sum_{i=1}^{M} H_i}{\sum_{i=1}^{M} cost_i}, \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733018, IEEE Transactions on Sustainable Computing

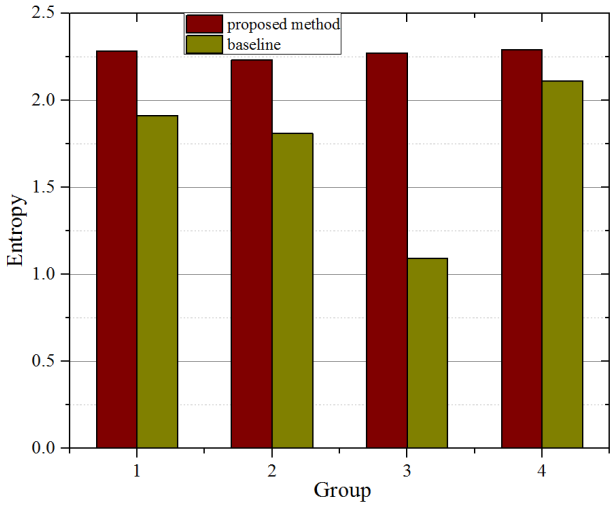


Fig. 11. Entropy reached of each group using $5-$anonymity

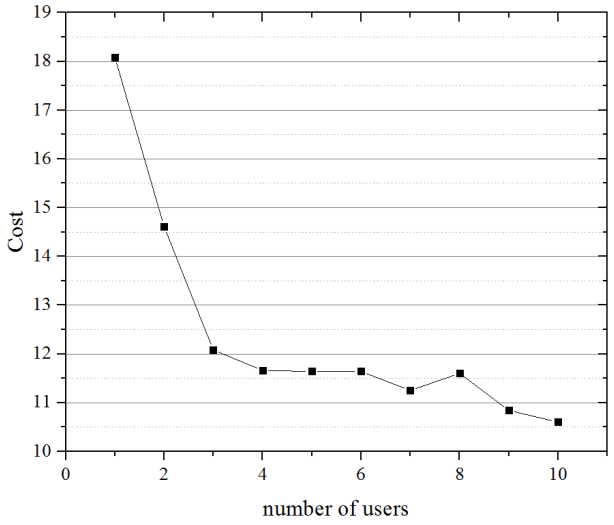


Fig. 12. Cost as the number of users increases

where H_i denotes the entropy-measured privacy level for user i and $cost_i$ is the expenses of privacy protection for user i. Fig.13 shows the P-C ratio of our proposed method outperforms the baseline method among different groups, which means that our method reaches a higher privacy level when given the cost is the same. We also find that the P-C ratio varies among different group, which is caused by the difference of population. Users in the groups which have a higher population can enjoy a higher privacy level with a lower cost.

8 Conclusion and Future Work

In this paper, we have studied the privacy protection problem for LBS using $k-$anonymity. For privacy protection, the level is often measured by entropy and generally the query probability varies among different places. With this side information, the attacker can readily narrow down the search space when launching the inference attack.

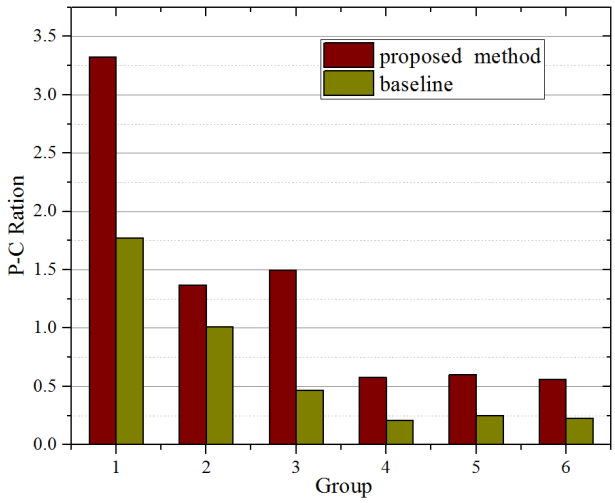


Fig. 13. P-C ratio among different groups

Faced with this challenge, in this paper, we propose a cost sharing schema. We first divide the area into small groups by maximizing the total entropy in order to maximize the users' privacy. Moreover we have studied the cost sharing problem to determine which LBS user to generate $k-1$ dummy trajectories and receive the payments from the others. At first, we have constructed an auction based model, in which each LBS user as a bidder, reports his cost and dummy trajectories. Then, we have proposed a cost sharing mechanism for location privacy preservation in LBS. Next, we have demonstrated that our mechanism satisfies incentive compatibility and budget balance.

To verify that the proposed method is effective, we carry out experiments on real dataset $GeoLife$ and dianping. The experiments proves that our method can achieve a higher privacy level when the cost is the same. Meanwhile, individuals of different groups have a similar privacy level.

We also notice the value of k is important when balancing the privacy level and the cost for dummy locations generation. A larger value of k would certainly enhance the privacy level. However, it will also inevitably incur some extra cost. Hence, we are going to have an research on the determination of k in our future work.

Acknowledgments

This paper is partially supported by the EU FP7 CROWN project under grant number PIRSES-GA-2013-610524, the Security Lancaster Mini Projects, the National Science Foundation of China under Grant No. 61672276, No. 61273232, the Key Research and Development Project of Jiangsu Province under Grant No. BE2015154, BE2016120 and the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University.

References

[1] B. Yao, F. Li, and X. Xiao, "Secure nearest neighbor revisited," pp. 733–744, 2013.

[2] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing private queries over untrusted data cloud through privacy homomorphism," vol. 6493, no. 4, pp. 601–612, 2011.

[3] 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, 2008.

[4] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," in *2008 IEEE 24th International Conference on Data Engineering*. Ieee, 2008, pp. 376–385.

[5] J. Freudiger, R. Shokri, and J. P. Hubaux, "On the optimal placement of mix zones," in *International Symposium on Privacy Enhancing Technologies*, 2009, pp. 216–234.

[6] X. Liu, H. Zhao, M. Pan, H. Yue, X. Li, and Y. Fang, "Traffic-aware multiple mix zone placement for protecting location privacy," *Proceedings - IEEE INFOCOM*, vol. 131, no. 5, pp. 972–980, 2012.

[7] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Pervasive Services, 2005. ICPS '05. Proceedings. International Conference on*, 2005, pp. 88–97.

[8] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *The ACM Sigsac Conference*, 2015, pp. 1298–1309.

[9] E. Elsalamouny and S. Gambs, "Differential privacy models for location-based services," *Transactions on Data Privacy*, vol. 9, no. 1, pp. 15–48, 2016.

[10] X. Yi, R. Paulet, E. Bertino, and V. Varadharajan, "Practical approximate k nearest neighbor queries with location and query privacy," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 6, pp. 1546–1559, 2016.

[11] R. Lu, X. Lin, Z. Shi, and J. Shao, "Plam: A privacy-preserving framework for local-area mobile social networks," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 763–771.

[12] X. Zhu, H. Chi, B. Niu, W. Zhang, Z. Li, and H. Li, "Mobicache: When k-anonymity meets cache," in *GLOBECOM 2013 - 2013 IEEE Global Communications Conference*, 2013, pp. 820–825.

[13] B. Niu, Z. Zhang, X. Li, and H. Li, "Privacy-area aware dummy generation algorithms for location-based services," in *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014, pp. 957–962.

[14] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[15] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 2007, pp. 106–115.

[16] J. Soria-Comas, J. Domingo-Ferrer, D. Sanchez, and S. Martinez, "t-closeness through microaggregation: Strict privacy with enhanced utility preservation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3098–3110, 2015.

[17] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive computing*, vol. 2, no. 1, pp. 46–55, 2003.

[18] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis, "Privacy-preserving indoor localization on smartphones," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3042–3055, 2015.

[19] H. Hu, Q. Chen, and J. Xu, "Verdict: Privacy-preserving authentication of range queries in location-based services," in *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. IEEE, 2013, pp. 1312–1315.

[20] J. Shao, R. Lu, and X. Lin, "Fine: A fine-grained privacy-preserving location-based service framework for mobile devices," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 244–252.

[21] J. Li, X. Huang, J. Li, X. Chen, and Y. Xiang, "Securely outsourcing attribute-based encryption with checkability," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2201–2210, 2014.

[22] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in *Theory of Cryptography Conference*. Springer, 2007, pp. 535–554.

[23] A. K. Tyagi and N. Sreenath, "Location privacy preserving techniques for location based services over road networks," in *Communications and Signal Processing (ICCSP), 2015 International Conference on*. IEEE, 2015, pp. 1319–1326.

[24] F. Qiu, F. Wu, and G. Chen, "Privacy and quality preserving multimedia data aggregation for participatory sensing systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 6, pp. 1287–1300, 2015.

[25] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 2399–2407.

[26] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," *Proceedings - IEEE INFOCOM*, pp. 754–762, 2014.

[27] D. Liao, X. Huang, V. Anand, G. Sun, and H. Yu, "k-dlca: An efficient approach for location privacy preservation in location-based services," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[28] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k-anonymity location privacy," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2994–3002.

[29] Y. Zhang, W. Tong, and S. Zhong, "On designing satisfaction-ratio-aware truthful incentive mechanisms for k-anonymity location privacy," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2528–2541, 2016.

[30] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in location based services," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2985–2993.

[31] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007, vol. 1.

[32] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *International Workshop on Privacy Enhancing Technologies*. Springer, 2002, pp. 41–53.

[33] H. Prakash, R. Narayanam, D. Garg, and Y. Narahari, "Game theoretic problems in network economics and mechanism design solutions," 2009.

[34] A. S. Prasad and S. Rao, "A mechanism design approach to resource procurement in cloud computing," *IEEE Transactions on Computers*, vol. 63, no. 1, pp. 17–30, 2014.

[35] Y. Shoham and K. Leyton-Brown, *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[36] Y. Zheng, X. Xie, and W.-Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory." *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.

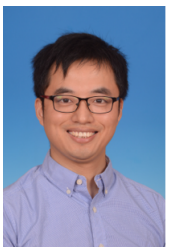
[37] Dianping. [Online]. Available:http://www.dianping.com/.

[38] Hana. [Online]. Available:https://www.sap.com/products/hana.html.

[39] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *ICPS'05. Proceedings. International Conference on Pervasive Services, 2005.* IEEE, 2005, pp. 88–97.

Fan Fei is currently working towards the Master degree at the Department of computer Science and Technology, Nanjing University, China. He has received his Bachelors degree in Information Engineering School from Zhengzhou University.

His research interests include Cloud Computing, Intelligent Transportation Systems.



Shu Li received his Master degree in Software Engineering from Nanjing University, in 2012. Now, he is a PhD student of Prof. Dou Wanchun in the Department of Computer Science and Technology of Nanjing University. He is also the director of Computing Olympiad Team in Nanjing Foreign Language School. He studied the application of In-Memory Database in Hasso-Plattner Institute from October 1st, 2014 to April 1st, 2015.

Haipeng Dai received the B.S. degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2010, and the Ph.D. degree from the Department of Computer Science and Technology, Nanjing University, Nanjing, China, in 2014. He is currently a Research Assistant with the Department of Computer Science and Technology, Nanjing University, Nanjing, China. His research interests are mainly in the areas of wireless sensor networks, wireless power transfer, and smartphone.

Chunhua Hu received the Ph.D. degree in computer science from Central South University, Changsha, China, in 2007. He is currently a Professor from the School of Computer and Information Engineering, Hunan University of Commerce, Changsha. Up to now, he has chaired two National Natural Science Foundation of China projects and published more than 20 research papers in international journals and international conferences. In 2012, he has been selected into the Program of New Century Excellent Talents in University. His research interests include cloud computing, service computing, and dependability computing.

Wanchun Dou received his PhD degree in Mechanical and Electronic Engineering from Nanjing University of Science and Technology, China, in 2001.

From Apr. 2001 to Dec. 2002, he did his postdoctoral research in the Department of Computer Science and Technology, Nanjing University, China. Now, he is a full professor of the State Key Laboratory for Novel Software Technology, Nanjing University, China. From Apr. 2005 to Jun. 2005 and from Nov. 2008 to Feb. 2009, he respectively visited the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, as a visiting scholar. Up to now, he has chaired three NSFC projects and published more than 60 research papers in international journals and international conferences. His research interests include workflow, cloud computing and service computing.

Qiang Ni received his Ph.D. degree in Engineering from Huazhong University of Science and Technology, Wuhan, in 1999. He is a Full Professor and the Head of Communication Systems Group, School of Computing and Communications, Lancaster University, UK. He is with Data Science Institute and Security Lancaster Centre. His research interests include future generation communications and networking systems, big data analytics, mobile and cloud networks, 5G, SDN, security and privacy, etc. Up to now, he had published more than 160 research papers in international journals and conferences.