An Improved Privacy Protection Method for Anonymity of Location k Based on Query Probability

Jie LING and Ling-li LI*
Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China
*Corresponding author

Keywords: Privacy protection, Query probability, Anonymous zone, Location sensitivity level.

Abstract. Aiming at the existing location privacy protection methods rarely considering the problem of location semantics, query probability, and that the location anonymous server is semi-trusted, an improved method is proposed to protect the user's location privacy. This method uses the structure based on the cooperation of centralized multiple location anonymous servers to solve the semi-trusted problem. The anonymous server generates an anonymous area according to the user-defined location sensitivity level and the historical query probability of location. And considering that the user location is extremely dense or sparse, we introduced personalized requirements, in case the anonymous area is too small to expose privacy and the area of anonymous zone is too large to reduce the service quality. Experimental results show that this method can protect the location privacy of users and improve service quality under the condition of the anonymous server is semi-trusted.

Introduction

Due to the rapid development of mobile communication, the popularity of smart products such as mobile phones and tablets, LBS (location-based services) has become one of the basic services for social networks. LBS means that the mobile terminal uses positioning technology to obtain its current location, and obtains the user's location-related services. LBS is divided into pulling service and pushing service according to whether it needs direct user interaction or not [1]. The pulling service means the user will send the query request to the location service provider to obtain the corresponding services, such as "The route from Guangzhou University South Metro Station to Guangzhou Tower". Pushing service means the service provider initatively sends relevant information to the user, such as the nearby discount recommendations from Alipay, Meituan and others applications. Location services have brought great convenience to life, subtly changing people's way of living, but also brought some troubles. The LBS server has all the users' information, including the user's location information, which includes the location coordinates, social semantics, physical conditions and some deep-seated personal information, and query content. The attacker can infer the user's personal privacy information with the user's submitted location, query content, combining the background knowledge mastered by the attacker himself. People are increasingly concerned about their own location privacy, and location privacy protection methods have also been widely studied. Many researchers devote themselves to finding a balance between location privacy and service quality. It is significant to design effective privacy protection methods to protect the privacy of users while having users fully enjoyed location service and satisfy their personalized needs [2].

Common k-anonymity Location Privacy Protection Methods

The main idea of the k-anonymity [3] location method is that the other k-1 user locations cannot be distinguished from the real user location, using an anonymous area to replace the user location, so that the attacker cannot obtain the real location of the user. The service quality depends on the size of the anonymous area. The smaller the anonymous area is, the less communication and computing overhead will be and the higher the service quality will be. But if the user-intensive anonymous area is too small, it is very easy to expose the user's real location. And if the anonymous area in the user
sparse environment is too large, LBS service quality will be reduced [4]. These locations may be
locations that do not actually issue inquiries, such as lakes, oceans, or locations that contain multiple
highly sensitive locations. Using such rich location semantic information, some locations with less
popularity in the anonymous area may be excluded, thereby reducing the size of anonymous area [5].

Among some related researches based on of location k anonymity, Huang Y et al. [6] proposed a
privacy-preserving method for location-based privacy protection in user collaboration without
anonymous area. This method forms an anonymous user group of k-users through collaboration
among users, and users in anonymous group use the density center of the group instead of the real
location to send inquiries, achieving the effect of k anonymity without using anonymous area. This
method requires that the client should have high computational and storage performance, and have the
problem of untrustworthy users. Zhuang L et al. [7] proposed a method of using secure multi-party
summation to compute anchor points considering the untrustworthiness and semi-trust of user nodes.

But the true distribution of users was not taken into account. Hu D et al. [8] proposed to adopt
client-server structure and generate k-anonymous zone based on user's privacy preferences and user's
network environment in the user end. It solves the shortcoming that SpaceTwist method does not
ensure k anonymity. However, it's using the client to execute anonymous algorithm cannot use other
user information in the process of anonymity. Zhao D et al. [9] proposed the location privacy
protection method based on the query probability-the ARB method, which realized the intrusion
attack even the attackers obtain the relevant background knowledge. However, there is a problem that
a large anonymous area is formed when the distribution of historical access probability data is not
even enough, and this will greatly affect the service quality.

This paper improves the method of location privacy protection based on query probabilities, which
has the following advantages: a) Generate anonymous area according user-defined location sensitivity
level, a more personalized privacy protection is achieved. b) Adopt system structure of multiple
location anonymous server collaboration to solve semi-trusted problem of anonymous servers and
send the user's real location to form anonymous area containing k position. c) Choosing k locations
instead of k grids to form anonymous areas reduces the anonymous area and improves service quality
with the customization of the minimum anonymous area and the maximum tolerable anonymous area.

**The Improved Method Proposed in This Paper**

An attacker who has enough background information about the anonymous space can deduce more
user data with this background information and disclose the user's personal information [10]. If
merely choose the anonymous area with the largest information entropy, it is vulnerable to the reverse
attack once the attacker obtains historical query probability data, so that the attacker can exclude some
irrelevant user locations in the anonymous area based on the obtained data, failing to ensure effective
k-anonymity [11,12]. The common location k-anonymity algorithm uses a centralized structure. The
user sends his real location to the trusted location anonymous server for anonymity. And the
anonymous server generates an anonymous request which will be sent to the LBS server for
requesting services. Fully trusted location anonymous servers does not exist in reality, once the
location anonymous servers are hacked, the user's location privacy will be leaked. The method uses a
system structure which consists of multiple location anonymous servers to solve the problem that the
location anonymous server is vulnerable. It does not require the client to have powerful computing
ability or large storage capability.

User sends a query request to multiple location anonymous servers in segments, randomly selecting
one of the anonymous servers to combine data segments to form a complete user query request, and
each location anonymous server only has some of the data of the user. Only if the attacker gets all the
data can he or she expose the user's location privacy. Anonymous work is done by that anonymous
servers and it sends an anonymous request to the LBSP (location-based service provider). The
location provider searches for the query results and sends the results to the anonymous server, the
resulting refinement module of which filters out the exact results according to the user's real location and returns them to the query user.

Figure 1 describes the work process of the method in this paper. It is composed of user client, multiple location anonymous servers and location-based service providers.

Step 1: Send the query user's area $G$ to the location server, and the LBS server divides the user's space from top to bottom layer by layer until the space is divided into $m \times m$ grid units with the same size, and each grid cell has at least one point that is less sensitive to the user.

Step 2: Generate user query requests, which is segmented and sent to multiple location anonymous servers. Querying user generates location privacy requirements according to the customized privacy parameters $Q_p = \{k, A_{\min}, A_{\max}\}$, where $k$ represents the threshold of user-defined location $k$ anonymity, that is, another $k-1$ user location is required in the anonymous area, $A_{\min}$ represents the minimum anonymous area, and $A_{\max}$ is the maximum tolerable anonymous area defined by the user. The query request sent to multiple location anonymous servers is as shown in Formula 1.

$$Q_u = \{U_{id}, Loc=(x,y), tp, SEL_u, Q_p, con\}.$$  

The user's identity is $U_{id}$. Loc represents the user's location. $x, y$ stand for the user's latitude and longitude coordinates. $tp$ represents the location type of the user, for example, different types of locations such as hotels, offices, militarized venues, etc. $TP = \{tp_1, tp_2, ..., tp_n\}$ is a collection of $n$ types of location. $SEL_u = \{sel_{tp_1}, sel_{tp_2}, ..., sel_{tp_n}\}$ is the sensitivity level of all location types relative to the user. The location type sensitivity level $sel_{tp_i}$ is used to indicate the level of sensitivity of the location to the user. Location semantic information varies from person to person, and the same type of location has different sensitivities to different people. The higher the sensitivity level, the more sensitive the user is to this type of location.

Step 3: The location anonymous server's database module calculates and stores the historical query probabilities for each location; the calculation is as shown in Formula 2.

$$q_i = \frac{\text{The number of queries submitted by all users in location i}}{\text{The number of queries submitted by all users in all locations}}. $$

It assumed that the greater the probability of location history query is, the higher the probability that the user will send a query request on this location will be. Randomly selects one of the location anonymous servers to combine the data segments into a complete user query request. The location anonymous server first finds all of the original anonymous regions $K_o$-ASR according to the privacy parameters $k$ and $SEL_u$ sent by the user, that is, each original anonymous region contains other $k-1$ positions with lower sensitivity level for the user, and then counts the number of possible anonymous areas $K_p$-ASR that is greater than the user-defined minimum anonymity area. According information entropy equation, we can calculate information entropy of all the possible anonymous areas, taking
the $k$ regions with the largest entropy and the anonymous regions smaller than the user-defined $A_{\text{max}}$. Finally, randomly select one of the $k$ regions as an anonymous area. The resulting anonymous query requests are shown in Formula 3.

$$Q_u = \{U_{id}, k-\text{ASR}, \text{con}\}.$$  

Where $k$-ASR represents the anonymous area generated by the location anonymous server, and $\text{con}$ represents the query content submitted by the user.

Step 4: Anonymous server sends anonymous query requests to the location service provider.

Step 5: The location service provider searches for the query result and sends the result to the anonymous server. The result refinement module of the anonymous server obtains the accurate result according to the real location of the user and returns it to the query user.

Security Analysis

The method proposed in this paper uses collaboration of multiple location anonymous servers based on centralized architecture. And also it generates an anonymous area with user-defined location sensitivity levels to realize a more personalized privacy protection. This paper analyzes the security of this method from the following two points.

Zhao DP et al. [9] did not consider location semantics. If there are multiple sensitive locations contained in the anonymous area, once attacked, it is considered that users are most likely to make service queries in these sensitive locations, while users do not want to be misjudged as being in sensitive position. Because users have different sensitivity levels for each type of location, users can customize the level of sensitivity for each type of location. And the method proposed in this paper is combining the user-defined location type sensitivity level to generate anonymous area, that is, the user sends a customized set of location-type sensitivity level to the location anonymous server, and selects $k-1$ location users with low sensitivity level to the user to generate an anonymous area. As users need to set sensitivity level for each type of location, the amount of computing at the client is increased, and the user-defined location type sensitivity level needs to be taken into consideration when selecting $k-1$ locations, so the amount of computation and execution time anonymous algorithms are also increased.

The common $k$-anonymity method uses a centralized architecture, assuming that a location anonymous server is fully trusted, but this condition does not exist in the network actually. Zhao DP et al. [9] solved the semi-trusted problem of the anonymous server by sending the smallest grid area where the user locates to the location anonymous server. The anonymous server returns the accurate result according to the area submitted by the user, leading to inaccurate results. In this paper, multiple location anonymous servers cooperate with each other to solve the semi-trusted issues of location anonymous server, and the result set is filtered according to the actual location of the user. The probability that an anonymous server is half trustworthy. This paper uses multiple anonymous server collaboration, and randomly selects one of the anonymous servers to rebuild the query data segments into a complete query request. The attacker can obtain the real location information of the user only after obtaining all the data, while the semi-trusted probability of anonymous server in the whole system structure is $(1/2)^n$, where $n$ is the number of anonymous server. This paper reduces the probability of unreliability of the location anonymous server of the whole system structure and improves the security.

Experimental Results Analysis

Our method is implemented in Java language and runs on a Windows operating system, with processor at 2.6 GHz, and 4 GB memory. Experiments uses Thomas Brinkhoff road network data generator to generate simulation data, selects the 3.2km × 3.2km space, the target space is divided into $128 \times 128$ grid space by the 25m of edge threshold, and generating 1000 mobile users. Using 5 location
anonymous servers, the user set the privacy requirement $k$, the minimum anonymous area $0.25 \text{km}^2$, and the maximum anonymous area $1.3 \text{km}^2$. In this paper, the method is mainly evaluated with the size of the anonymous area and the response time.

Figure 2 shows the relationship between the anonymous area and the user-defined $k$. It can be seen that the anonymous area is positively related to $k$ and the method we proposed is smaller than the anonymous area formed by the ARB method. The smallest anonymous area in this paper is $0.25 \text{ km}^2$, the largest anonymous area is $1.3 \text{km}^2$. This is because the ARB method takes a grid as a minimum unit, selects $k$ grids as an anonymous region, but at least one eligible user position is included in a grid. Compared with ARB, this paper uses every position point as the smallest unit and chooses $k-1$ positions to generate anonymous area, greatly reducing the anonymous area, assuring the security of privacy protection for location-based service seeks and making the anonymous area as small as possible.

![Anonymous area sizes](image)

**Figure 2. Anonymous area sizes.**

As shown in Figure 3, it can be seen that with the ARB method, as the value of $k$ increases, the anonymous area is getting larger, more and more search points are returned, and the response time becomes longer. Compared with the ARB method, this method adds the mutual cooperation between multiple anonymous servers, and uses user-defined location sensitivity level to generate anonymous area. The response time is longer than that when the ARB method adopts a small $k$ value, but when the $k$ value increases, the anonymous area generated by this method is smaller, the number of returned search points is reduced and the response time is shortened.

![Response time](image)

**Figure 3. Response time.**
Conclusion

Compared with the ARB method, this paper generates anonymous areas combining with the user-defined location sensitive level, and realizes more personalized privacy protection. The problem that the location anonymous server is semi-trusted is solved by sending the user's real location to multiple anonymous servers, rather than sending the user's grid area to an anonymous server. The ARB method uses $k$-grid to generate anonymous area, while we use $k$-location user to generate it and make it smaller. And we use user-defined minimum anonymous area and maximum tolerable anonymous area to solve the problem that the information is easily leaked caused by the over centralized location of users and the problem that the over sparse location of users will lead to an over large anonymous area which affects the service quality. Since we do not consider the privacy protection of the query content in this paper, the next research direction is to search the content privacy protection.

Acknowledgement

This work is supported by the science and technology project of Guangdong Province (No.2015B090906016, 2016B010107002, 2016B090918058), and the project of Guangzhou Science and Technology (No.201604016003, 201604016067).

References

[1] Zhang X J, Gui X L, Wu Z D. Privacy preservation for location-based services: a survey, J. Journal of Software, 2015, 26(9):2373-2395.
[2] Min L I, Qin Z G. Survey of location privacy protection over road networks, J. Application Research of Computers, 2014, 16(16):1-10.
[3] Niu B, Li Q, Zhu X, et al. Achieving k-anonymity in privacy-aware location-based services[C]// IEEE INFOCOM. IEEE, 2014:754-762.
[4] Wang W, Chen M, Zhao Y. An improved location-cloaking based on grid-divided space, J. Computer Simulation, 2016, 33(09):208-210.
[5] Chen H, Qin X. Location-semantic-based location privacy protection for road network, J. Journal on Communications, 2016, 37(08):67-76.
[6] Huang Y, Huo Z, Meng X. CoPrivacy: A collaborative location privacy preserving method without cloaking origin, J. Chinese Journal of Computers, 2011, 34(10):1976-1985.
[7] L Zhuang OU Yuyi, J Ling. Location privacy protection technology based on KHINN, J. Telecommunications Science, 2016, 32(09):89-94.
[8] Hu D, Zheng X. Space Twist-based k-anonymity incremental nearest neighbor query algorithm for location privacy protecting, J. Application Research of Computers, 2016, 33(08):2402-2404.
[9] D Zhao, G Song, Y Jin, X Wang. Query probability-based location privacy protection approach, J. Journal of Computer Applications, 2017, 37(02):347-351.
[10] Liao D, Huang X, Anand V, et al. k-DLCA: An efficient approach for location privacy preservation in location-based services[C]// IEEE International Conference on Communications. IEEE, 2016:1-6.
[11] Zhou C, Ma C, Yang S, et al. A Location Privacy Preserving Method Based on Sensitive Diversity for LBS [M]// Network and Parallel Computing. Springer Berlin Heidelberg, 2014:409-422.
[12] Yu R, Bai Z, Yang L, et al. A Location Cloaking Algorithm Based on Combinatorial Optimization for Location-Based Services in 5G Networks, J. IEEE Access, 2017, 4:6515-6527.