Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

ASA: Against statistical attacks for privacy-aware users in Location Based Service

Yanming Sun [a], Min Chen [a], Long Hu [a,*], Yongfeng Qian [a], Mohammad Mehedi Hassan [b,c]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
[b] Computer Engineering Department, King Saud University, P.O. Box 51178, Riyadh, Saudi Arabia
[c] Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

HIGHLIGHTS

- According to the rule of activity of most users, we introduce LSA by using the historical data. For the attack, we give out two methods to preserve user's privacy.
- We divide the regions in the map into different PLs according to the privacy requirement. We design algorithm to make the regions of high level to be dummies at a high rate and the regions of low level at a low rate. The problem that the attacker can violate the privacy of a particular region by analyzing the historical data is solved.
- We analyze the ability to preserve user's privacy by entropy. The internal relation among the frequency of user's LBS query, the division of regions in the map, and the length of the interval of historical information collected is discussed.

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form
14 March 2016
Accepted 19 June 2016
Available online xxxx

Keywords:
Location privacy
Privacy preserving
k-anonymity
Information entropy
Location based service (LBS)

ABSTRACT

The fusion of mobile devices and social networks is stimulating a wider use of Location Based Service (LBS) and makes it become an important part in our daily life. However, the problem of privacy leakage has become a main factor that hinders the further development of LBS. When a LBS user sends queries to the LBS server, the user's personal privacy in terms of identity and location may be leaked to the attacker. To protect user's privacy, Niu et al. proposed an algorithm named enhanced-Dummy Location Selection (en-DLS). In this paper, we introduce two attacks to en-DLS, namely long-term statistical attack (LSA) and regional statistical attack (RSA). In the proposed attacks, an attacker can obtain the privacy contents of a user by analyzing LBS historical data, which causes en-DLS to be invalid for user's privacy protection. Furthermore, this paper proposes a set of privacy protection schemes against both LSA and RSA. For LSA, we propose two protection methods named multiple user name (MNAME) and same user name (SNAME). To solve the regional privacy issue, we divide the map into various regions with different requirements on privacy protection. For this purpose, four levels of protection requirements (PLs) are defined, and true location is protected by allocating a certain number of positions from the dummies according to the location's PL. Performance analysis and simulation results show that our proposed methods can completely avoid the vulnerabilities of en-DLS to both LSA and RSA, and incur marginal increase of communication overhead and computational cost.

1. Introduction

With the development of mobile computing and network technology, mobile phone has become a necessity in people's life. Besides satisfying the need of daily communications, mobile phone also provides many convenient services for human being [1]. With the development of smartphones, Global Positioning System (GPS) has been solidified in the most smartphones and provides function for mobile service provider to position the smartphone. America E-911 document has pointed out that mobile service providers should provide location recognition service in 125 m in order that the owner of mobile phone can obtain timely rescue and help when she is in danger, such as fire or hijacked [2].

[*] Corresponding author.
E-mail addresses: yanming.epic@gmail.com (Y. Sun), minchen2012@hust.edu.cn (M. Chen), longhu.cs@gmail.com (L. Hu), yongfengqian.epic@qq.com (Y. Qian), mmhassan@ksu.edu.sa (M.M. Hassan).

http://dx.doi.org/10.1016/j.future.2016.06.017
0167-739X/© 2016 Elsevier B.V. All rights reserved.

Recently, many Location Based Service (LBS) applications come into being. In spite of various benefits brought by LBS, the intrinsic privacy leakage problem cannot be ignored due to the openness of wireless networks [3]. At present, privacy leakage issues become the main obstacle to the wide application of LBS services.

As stated in [4], location privacy is ''*the ability to prevent others from obtaining the current or past location of the user*''. For privacy preserving in LBS service, there are several challenges:

- High Precision: The user's identity and location should be protected. Meanwhile, the precision of LBS service should be ensured.
- Low Overhead: The communication, computation and storage ability of the user terminal is limited. Thus, the communication overhead, the computational cost, and the storage overhead should be low in preserving the user's privacy.
- Privacy: LBS server itself may be an attacker. It can obtain the user's real location and historical data directly.

The main solutions to protect user's privacy can be divided into obfuscation and anonymity according to the technology used. In anonymity, using dummy is efficient since it need not a trusted third party to preserve privacy and it attracts many scholars' attention. Among them, Niu et al. proposed enhanced-Dummy Location Selection (en-DLS) based on the probability of sending LBS queries from a location in the history by users [5]. It solves the problem of privacy leakage in a single LBS query. In [6], Niu et al. proposed Caching-aware Dummy Selection Algorithm (CaDSA) and enhanced-CaDSA which use caching to improve the privacy of user. En-DLS has the following characteristics:

- Side information: In en-DLS, side information refers to the terrain information in the city. The dummies are not selected from the rivers or mountains in the city, but carefully selected based on the historical query probability in the locations. The problem of reducing in protection caused by side information is solved.
- Cloaking area: To overcome the disadvantage of k-anonymity, in en-DLS, the coverage area of the dummies is selected as large as possible.
- Implementation issues: In en-DLS, accessing to the historical queries is fully considered. Access Points (AP) based method is proposed. The communication overhead is relational.

Although en-DLS solved the problem of privacy leakage in a LBS query, it has vulnerabilities. In this paper, we introduce two attacks to en-DLS, namely long-term statistical attack (LSA) and regional statistical attack (RSA). The attacker can obtain user's privacy contents using historical statistics. For an attacker, after compromising the LBS server, he can obtain a large number of historical data. We introduce an attack named LSA to obtain user's real identity and location using these historical data. We study based on LSA and put forward two methods to preserve privacy named multiple user name (MNAME) and same user name (SNAME). Besides LSA, the attacker can obtain the historical LBS applications from a particular region. Furthermore, the attacker can obtain a lot of information about the user from the region through statistics. For this problem, we propose a method to divide the regions in the map into different privacy levels (PLs). Then, we delete some dummy locations in en-DLS and select some locations from high PL regions to protect the privacy of the regions. We take entxu2007preventingabbas2013collusionropy as the metric to analyze the ability of the proposed methods. The performance analysis and simulation results show that the proposed methods can effectively preserve user's privacy against LSA and RSA. The main contribution of this paper includes the following aspects:

- According to the activities of most users, we introduce LSA. For the attack, we give out two methods to preserve user's privacy.
- We divide the regions in the map into different PLs according to the privacy requirement. We give out an algorithm to make the regions of high PL to be dummies at a high rate and the regions of low PL at a low rate. The problem that the attacker can violate the privacy of a particular region by analyzing the historical data is solved.
- We analyze the ability of preserving user's privacy by entropy. The relation among the frequency of user's LBS query, the division of regions in the map, and the length of the interval of historical information collected is discussed.

The rest of this paper is organized as follows. Section 2 gives out some preliminaries and motivation of this paper. In this section, we give out LSA and RSA. In Section 3, we propose methods to resist LSA and RSA. In Section 4, we discuss the security and performance of the proposed methods. Section 5 presents the simulations. In Section 6, we review the related work. Conclusion and future work are in Section 7.

2. Preliminaries

In this section, we first introduce the privacy metrics and attack model. Then, we give out the motivation of our solution.

2.1. Metrics for privacy

To measure the ability of preserving privacy, we need some metrics. There are five kinds of metrics currently [7]. They are uncertainty-based metric, ''clustering error''-based metric, traceability-based metric, k-anonymity metric, and distortion-based metric. In this paper, we use uncertainty-based metric to measure privacy in communication system. In [8], the author put forward to measure the ability of an attacker by differentiating the real locations from the anonymous set. The author pointed out *k-anonymity* is really achieved if the attacker cannot distinguish the real location from the $k-1$ locations in the same transmission. In [5], the author proposed that the direct method to measure the privacy preserving ability in *k-anonymity* is to use the k. The larger k denotes the higher ability to preserve privacy. However, there are some disadvantages in this measurement. For example, the $k-1$ dummies may be selected in the rivers, lakes, mountains, or in the impossible positions to reach in the path for the limited of speed. The attacker can easily distinguish them as unlikely LBS query positions from the real location. Therefore, simply using k as the metric cannot express the ability of privacy protection accurately. Besides k, entropy is widely used to measure the ability [5,6,8–10]. Entropy is first used to measure privacy in [11]. As we all know, entropy is often used to measure the uncertainty of a system. In privacy protection, entropy can be used to measure the degree of uncertainty of a location belonging to a user. In *k-anonymity*, from the point of view of the attacker, in the anonymous set consists of the real location and $k-1$ dummies, the probability of a location to be the real one is p_i. In the anonymous set, the sum of all probabilities p_i is one. Thus, the entropy H of identifying a real location in the candidate set is

$$H = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i. \qquad (1)$$

When all the k locations in the set have the same probability, the maximum entropy is achieved, where the probability p_i is $1/k$ for all the locations and the maximum of H is $\log_2 k$.