Journal of Network and Computer Applications 38 (2014) 125–134

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Balancing trajectory privacy and data utility using a personalized anonymization model

Sheng Gao*, Jianfeng Ma, Cong Sun, Xinghua Li

School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China

ARTICLE INFO

Article history:
Received 11 July 2012
Received in revised form
1 January 2013
Accepted 18 March 2013
Available online 29 March 2013

Keywords:
Trajectory *k*-anonymity set
Trajectory similarity and direction
Trajectory distance
A personalized anonymization model

ABSTRACT

With the widespread use of location-based services (LBS), the number of trajectories gathered by location service providers is dynamically growing. On the one hand, mining and analyzing these spatiotemporal trajectories can help to work out a mobile-related strategic planning; on the other hand, knowledge of each trajectory can be used by adversaries to identify the user's sensitive information and lead to an unpredictable harm. The concept of trajectory *k*-anonymity extends from location *k*-anonymity that has been widely used to address this issue. The main challenge of trajectory *k*-anonymity is the selection of a trajectory *k*-anonymity set. However, existing anonymity methods ignore the trajectory similarity and direction, assuming that it has little impact on privacy. Thus, it cannot provide a preferable trajectory *k*-anonymity set. In this paper, we propose to use trajectory angle to evaluate trajectory similarity and direction, and construct an anonymity region on the basis of trajectory distance. Considering the various preference settings on the proportion of trajectory privacy and data utility in different scenarios, we propose a personalized anonymization model to select the trajectory *k*-anonymity set. Experiment results prove that our method can provide an effective trajectory *k*-anonymity set under various proportions of trajectory privacy and data utility requirements, while the efficiency just reduces a little.

1. Introduction

The explosion of mobile devices equipped with powerful wireless communication ability, together with the rapid progress of mobile positioning techniques such as global positioning systems (GPS), radio frequency identification (RFID) and so forth, have greatly facilitated the prosperity of location-based services. While users sharing the mobile services, a large number of trajectories might be collected by service providers. Mining and analyzing these spatiotemporal trajectories (Ivanov, 2012) can help people to make a mobile-related decision, for instance, merchants can decide the place where to build a restaurant or a supermarket by analyzing trajectories of customers in a certain area (Cao et al., 2010) and tourism company can make a travel recommendation schedule by monitoring trajectories of visitors in a city (Zheng et al., 2009). For such practical applications, the main step is to explore accurate and applicable knowledge, which is out of the scope of this work.

The publication of spatiotemporal trajectories is a double-edged sword. Although mining trajectories can bring many advantages to multiple commercial applications, the disclosure of those spatiotemporal information contained in trajectories may threaten individuals sensitive information, such as home addresses, travel habits, political beliefs, health conditions, personal interests, and so on. To cope with the problem, trajectory *k*-anonymity is presented to anonymize *k* trajectories at least over a time span (Nergiz et al., 2008; Abul et al., 2008; Xu and Cai, 2008; Yarovoy et al., 2009). It is an extension of location *k*-anonymity (Gruteser and Grunwald, 2003), which conceals a user's trajectory with the assistance of the other $k-1$ trajectories at least. Instead of revealing the exact trajectory of a user, an obscure path called *anonymized path* that contains at least *k* trajectories is reported.

To ensure a high quality of anonymization, the main challenge is to determine a trajectory *k*-anonymity set. We observe that the selection of anonymity trajectories affects the trajectory privacy protection level and data utility. Shin et al. (2010a) noticed that existing location *k*-anonymity model regarded the location as sole information when achieving anonymization. However, the disclosure of users' movement directions can cause adversaries to identify a mobile user who submitted a LBS request. They proposed to improve the location *k*-anonymity model by taking a user's direction of movement into account during the anonymous request process. To best of our knowledge, most of trajectory *k*-anonymity methods anonymize the trajectories without taking the similarities and directions among them into consideration. As Shin et al.

* Corresponding author. Tel.: +86 15991720644.
E-mail address: sgao555@gmail.com (S. Gao).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.03.010

(2010a) demonstrated, the trajectory directions affect the trajectory privacy protection level. However, the differences among trajectories may also affect the quality of anonymization. The trajectories can be identified easily with high individual differences. Meanwhile, the data utility of trajectory may reduce with the expansion of anonymity region. In the follow-up method, they also proposed to use optimal trajectory division (Shin et al., 2010b) to strengthen privacy protection and improve the quality of service (QoS). Specifically, through the partition of trajectories with minimum area of anonymity region, the privacy level was increased for the unlinkability over time and the overall quality of service was improved for the smaller anonymous regions.

Motivated by this, in this paper, we take these factors into account. In location privacy protection, Gedik and Liu (2005) proposed a privacy framework on the basis of the requirements of location privacy k and QoS from the perspective of each user and then presented a cloaking algorithm *CliqueCloak* to produce an undirected graph for location privacy protection. However, it only works with a small value of k and fails when the value of k is large. To tackle this defect, Xiao et al. (2007) improved the cloaking algorithm for a robust anonymity while considering both location privacy and QoS. In trajectory privacy protection, trajectory similarity is an important factor for trajectory clustering and anonymization. Pelekis et al. (2007) presented a framework to address the trajectory similarity search problem. The authors transformed this issue into different kinds of similarity queries according to the trajectory characteristics. Moreover, the other works proposed some typical measures for trajectory similarity including Euclidean distance (Abul et al., 2008; Huo et al., 2011; You et al., 2007), edit distance (Chen et al., 2005) and linear spatio-temporal distance (Tiakas et al., 2009). However, in some cases, all of these could not reflect the factor of trajectory similarity and direction very well. To best of our knowledge, in this paper, we first propose to use trajectory angle to evaluate trajectory similarity and direction and construct an anonymity region based on trajectory distance. We construct a personalized anonymization model to balance trajectory privacy and data utility and then translate the selection of a trajectory k-anonymity set into a constrained minimum spanning tree problem. The proportion of trajectory privacy and data utility decided by a user is dependent on the application scenario. Considering that in different application scenarios, the various preference settings on the proportion of trajectory privacy and data utility may affect the selection of trajectory k-anonymity sets, we analyze the actual privacy level and data utility under these different trajectory k-anonymity sets.

In this paper, the main contributions of our work are summarized as follows:

- We propose a personalized anonymization model with taking trajectory privacy and data utility into consideration. In particular, we consider the factors of trajectory similarity and direction for privacy protection and trajectory distance for data utility.
- We transform the optimal k trajectories selection to a constrained minimum spanning tree problem and use Greedy strategy to find an approximate optimal trajectory k-anonymity set in the trajectory graph model we constructed. The weights model the relations between trajectories under various proportions of trajectory privacy and data utility.
- We run a set of evaluations on synthetic dataset. Experiment results prove that our method can provide an effective trajectory k-anonymity set under various proportions of trajectory privacy and data utility requirements, while the efficiency just reduces a little.

The remainder of this paper is organized as follows. Section 2 summarizes related work. Section 3 introduces some basic notions and states the problem of tradeoff between trajectory privacy protection and data utility. In Section 4, we present the personalized trajectory anonymization model and discuss the function of each component in detail. Section 5 discusses the metric in terms of trajectory privacy and data utility. In Section 6, we run and analyze a set of simulations on synthetic dataset to evaluate the selection of a trajectory k-anonymity set under various proportions of requirements on trajectory privacy and data utility, and then compare the effectiveness and efficiency with the previous work. Finally, we conclude this paper and present the future work in Section 7.

2. Related work

Trajectory privacy is a special type of personal privacy, which has been concerned continuously in recent years. According to the time sequence of trajectory, existing trajectory privacy-preserving techniques can be classified into three types (Huo et al., 2011).

2.1. Dummy trajectories

Kido et al. (2005) presented two algorithms to determine the dummy trajectories for trajectory privacy protection. To be specific, the next location of a dummy is selected in a neighborhood of its current location. You et al. (2007) presented two approaches to produce consistent movement patterns in a long term. However, these methods cannot strictly ensure a good similarity between trajectories. Our previous work (Gao et al., 2012) focused on the tradeoff between location and trajectory privacy protection and QoS based on a user's partners' locations and trajectories, and then proposed a method to produce the partners' trajectories that looks like the user's trajectory.

2.2. Suppression technique

Gruteser and Liu (2004) proposed to use suppression technique to protect a user's online trajectory privacy. The sensitivity map divided areas into sensitive and insensitive according to the user's settings. Once the user entered a sensitive area, location updates were suppressed at once. Terrovitis and Mamoulis (2008) studied the privacy-preserving problem in the publication of trajectory databases. They argued that each adversary would possess different portions of users' trajectories and the data publisher was aware of the adversaries' knowledge. They proposed a method that iteratively suppressed some trajectory segments until a probabilistic constraint of disclosing whole trajectories was satisfied. However, if too many trajectory segments are suppressed, it would cause lots of information loss.

2.3. Trajectory k-anonymity technique

Trajectory k-anonymity technique that anonymizes k trajectories together is directly related to our work. As a result of the imprecision of GPS devices, Abul et al. (2008) proposed *Never Walk Alone* (*NWA*) to enforce $(k,\delta)-$*anonymity* model they presented for mobile object databases using trajectory clustering and space translation. Huo et al. (2012) improved the *NWA* by anonymizing the stay points based on *grid-based approach* and *clustering-based approach*. Domingo-Ferrer and Trujillo-Rasua (2012) proposed two heuristic methods to anonymize trajectories. One of them aims at trajectory k-anonymity by microaggregation and the other is to achieve location k-diversity while considering the reachability constraints. A new distance is proposed to improve the *NWA*, which can process those trajectories without time overlap. Nergiz et al. (2008) proposed to enforce k-anonymity by grouping the