USER PRIVACY PROTECTION FROM TRAJECTORY PERSPECTIVE IN LOCATION-BASED APPLICATIONS

Trong Nhan Phan, Tran Khanh Dang
Faculty of Computer Science & Engineering
HCMUT
Ho Chi Minh City, Vietnam
nhanpt@cse.hcmut.edu.vn, khanh@cse.hcmut.edu.vn

Josef Küng
FAW Institut
Johannes Kepler University Linz, Austria
josef.kueng@faw.jku.at

Abstract

Under the help of information and communication technologies, our lives are becoming better than ever in such a way we have been spending. One of popular and worthy supports comes from mobile devices. Integrated with modern positioning-aware technologies like Assisted Global Positioning System (A-GPS), mobile devices nowadays can locate users’ position with acceptable accuracy. However, privacy issues have been emerging as a real challenge to industry as well as researchers world-wide. We cannot deny the advantages of LBS, but one’s privacy can be violated while he enjoys such services. Dealing with this properly aims at protecting users’ personal information while the quality of services (QoS) needs to be assured. This essential balance makes it possible to users’ acceptance and reinforces interests to mobile market. Being a part of that progress, this paper gives an overall perspective about not only trajectory preservation but also its concerns in location-based applications. Moreover, its domains and relevant privacy-preserving techniques are also discussed as a key to the door of future research directions.

Keywords

Information and Communication, Location-based Services (LBS), Trajectory, Privacy-Preserving.

1. Introduction

The more modern our society is, the more advanced equipments are developed to support human life. State-of-the-art technologies today make them possible to gradually improve what they are lack of or limited. Besides, the ability of how to detect user location is also integrated into such mobile devices, which opens and paves the way for the development of what is so-called location-based services (LBS for short). However, there exists one important issue known as user privacy causing the most influence during LBS value chain in (Kupper, 2005) as well as making end-users feel afraid of using these services. That LBS supply customers with the demands of utilities is not
good enough to persuade them to freely use. The reason is that users are bound to sacrifice their position in order to get added-values in return. When enjoying these services, one is uncomfortable about whether his position and other personal information are revealed or misused. For instance, knowing the position or trajectory of a user gives an attacker a chance to re-identify the person to whom the position or trajectory belongs. Then, his personal information will be disclosed one after another. One is unhappy if his personal information he tries to keep by himself is revealed to others. Furthermore, malicious attackers can take advantage of his information to go against him. These matters really make sense in LBS. That is why there is a crucial need to keep user sensitive information safe or inviolate.

How vitally important user privacy is has spread a powerful wave of doing more research in this field of study, for its progress contributes to the final success of LBS. The studies have been documented and reported on privacy techniques suitable for both research and business. Among objects needing to be protected, trajectory is the vulnerable one in terms of spatiotemporal relationship. Protecting only positions at specific time just leads to snapshot-based and space-based protection. In fact, users are interested in services involved in both space and time. Therefore, how to identify threats and find out corresponding privacy defense based on trajectory become more essential than ever. To the best of our knowledge, we classify these into three application domains: (1) Application level; (2) Publication level; and (3) Database level. In addition, privacy objects are considered overall, and challenges when dealing with trajectory problems are also studied. From what we have had up to now, the idea of a new access method towards database level in an effort of unifying moving object data and privacy preservation will be mentioned afterwards.

The rest of the paper is organized as follows. Section 2 presents related work in which trajectory scope is classified into three main levels: application, publication, and database level. Next, trajectory problems in terms of privacy are discussed in section 3. Finally, we introduce our work and clarify our future direction in section 4 before making our conclusion in section 5.

2. Related Work

**Application level:** It makes itself possible to be a fertility land in that privacy techniques are integrated to protect users’ sensitive information. They follow mostly three kinds of architectures including centralized party, non-cooperation, and peer-to-peer (Mokbel, 2007). The authors in (Truong et al., 2010) propose a memorizing algorithm, using trusted middleware that organizes space in an adaptive grid where it cloaks the user’s location information in an anonymization area. In (Hao et al., 2007), the authors propose a dummies-based technique. The basic idea of this approach is to generate trajectory dummies in order for adversaries not to identify the true one and link it to the corresponding user. Mix zone, a concept proposed by the authors in (Beresford & Stajano, 2003), raises an idea of hiding users’ activities in a spatial region where users do not register their applications and let their position be revealed. A similar method known as dynamic mix zone is mentioned by the authors in (Yi et al., 2008). The zone is created dynamically whenever moving objects are close enough. The authors in (Dan et al., 2009) have presented the way called transformation. It aims at making adversaries confused by transforming data from the current space into another space. User requests, after sent to LBS service providers, will be processed in the transformed space, and corresponding results are returned to the users at present. Besides, some obfuscation techniques employ k-anonymity concept to do their jobs. The authors in (Toby et al., 2008) employ user historical positions instead of current neighbor positions. The basic idea is that we can cloak a node’s position based on its nearby footprints left by other people.
Publication level: Hardly can we deny the importance of data mining nowadays. Some researchers have also studied privacy-preserving methods in data mining. These techniques can be applied to processes before mining, during mining, or after mining so that they can achieve their goals. For example, reaching to k-anonymity and some privacy annotations like l-diversity, m-invariance, $\alpha$-diversity, and t-closeness makes trajectory data be generalized or suppressed in order to get extended k-awareness, which can reduce the probability of attacks and is shown in (Pierangela & Latanya, 1998; Győző et al., 2008; Fung et al., 2010; Rinku et al., 2010). The authors in (Mehmet et al., 2008) also exploit k-anonymity to make close trajectories be in a group and then to find a representative trajectory through a process called generalization. Another one is to make attackers so confused because of path perturbation as illustrated in (Ghinita, 2009). That means whenever two users meet, or their distance is less than a pre-defined threshold, their paths in such areas will be forced to cross with each other. This measure drives adversaries to their desperation when they try to monitor or track the trajectory of a user.

Database level: There is little closer work (e.g., towards database level) considering privacy as a significant factor in LBS. The authors in (Divanis & Verykios, 2008) present a privacy-aware trajectory tracking query engine that controls what should be shown to corresponding parties. Having the same approach, the authors in (Chow et al., 2009) propose a privacy-aware query processor embedded inside a location-based database server. It deals with how to process snapshot as well as continuous queries with cloaked regions instead of exact positions from the component known as location anonymizer. Another recent approach comes from the authors in (Vijayalakshmi et al., 2008), who propose a unified index called SPFF-tree. By having authorizations embedded into the tree, it is capable of supporting privileges based on locating and tracking. Actually when striking upon an idea of index structures towards user privacy, the authors in (Dang & To, 2010; To et al., 2011; Dang et al., 2011) have presented a database-centric approach and become pioneers in generating such indices. They have considered temporal and geographical factors other than spatial dimension. One is OST-tree, contributing to make spatiotemporal data obfuscated, and the other is $B^{ob}$-tree, recognizing geographical features inside cloaked regions. The indices open a new access method to spatial-temporal and geographic-aware obfuscation.

3. Trajectory-based Privacy Preservation

3.1. Trajectory Problem

Among location-based services, some refer to trajectory privacy. Here, we have online and offline processing applications. The former points out the direct interaction between a user and the service he is using. If a user is put in a positive place (i.e., he sends or updates his consecutive positions to use his service), we have applications based on continuous queries. Otherwise, we have tracking systems in that they take their initiative in monitoring target locations. The latter is concerned with data mining or data post processing where some data sets are retrieved and processed to publish necessary information expressing their semantics. Some examples for each kind of them are listed below but are not limited:

- Consecutive queries such as “continuously let me know where the nearest gas station is”.
- Tracking systems involved in fleet, asset, or workforce management.
- Publication including traffic planning, trend detection, or behavior discovery.

The population of such services becomes popular to users nowadays. In a general view, each of them leaves trails of user movements called paths or trajectories. According to them, some attackers
can obtain personal information they desire to know or discover items of interests where trajectories cross over or pass by. In particular, each has its own specific characteristics, in terms of privacy, which should be taken into account as well.

### 3.2. Objectives

When a privacy technique is proposed, it is directed toward what it is used for. So objectives need to be considered with other aspects that are significant in requirements in LBS. Here, we take into account criteria as followings: (1) Preventing the combination of a trajectory and an identity with or without the support of background or external knowledge; (2) Preserving the utility of data; (3) Protecting user privacy against untrustworthy parties (LBS server for example); (4) Protecting user privacy when acquiring services; (5) Protecting sensitive location samples (e.g., political buildings, private places, items of interest, etc.); (6) Providing high quality of service; (7) Being applicable to the real-life scenarios as much as possible. They can be found themselves in opposition to each other (e.g., the utility of data versus user privacy, the QoS versus user privacy). However, there should be a trade-off between what is offered and which level of privacy is required.

### 3.3. Approaches

![Figure 1: Trajectory-based privacy preserving approaches](image)

During a LBS supply chain, privacy techniques in general and trajectory-based privacy techniques in particular can be applied to many approaches. In this paper, we classify them into three main levels whose further details are discussed in related work:

- **Application level**: privacy mechanism is implemented at the application level. That means privacy techniques are put outside the storage database.

- **Database level**: privacy mechanism is embedded into the database. Every privacy technique is done before a user request and its relevant data are sent to service providers.

- **Publication level**: privacy mechanism is taken into account at either mining applications or the database. Additionally, where it is applied relies on phases of data mining techniques such as before, during, or after mining.

### 3.4. Important Factors

Through a circle of life for LBS, a request from a service can be affected directly or indirectly by many factors in (Nayot & Indrakshi, 2009). Depending on what we want to meet our needs, we
should be aware of risks from them. Factors listed below more or less reflect their influences on privacy preservation in general and trajectory preservation in particular: (1) **Identity**: identifies the requester who is requesting the location information or services; (2) **Object**: indicates the targets of the service including public or private objects (3) **Query**: specifies the purpose of a user request; (4) **Time**: denotes the time of a request; (5) **Period**: defines a time interval \([T_S, T_E]\); (6) **Location**: shows the position of the requested object or the requester’s; (7) **Velocity**: presents the speed of a moving object; (8) **Acceleration**: cares about the rate of change of velocity over time; (9) **Orientation**: introduces the direction of a moving object, and the angle of the moving is a derived attribute; (10) **Density**: involves the quantitative of objects in a specific area; (11) **Object profile**: gives more extra information of an object like its interests; (12) **Background knowledge**: supplies external information obtained for inference attacks.

### 3.5. Privacy Objects

#### 3.5.1. Identity Privacy

Identity refers to attributes explicitly identifying a user like user identification or social security number (SSN). Thus, **identity privacy** is to protect users’ identities directly from disclosure to attackers. The main goal is to protect users’ identities that could be directly or indirectly inferred from other information including background or external knowledge.

#### 3.5.2. Query Privacy

**Query privacy** is related to the disclosure of sensitive information in the query itself and its association to a user. In other words, a query should be protected from being linked to a specific user. Knowing the content of a query, for example, can indicate user preferences which, for some reason, had better to be kept in private.

#### 3.5.3. Position/Location Privacy

Users’ locations belong to sensitive contents that have to be kept away from malicious attackers. They include those which from the past, current, or even predicted positions. How to protect sensitive locations from being linked to a specific user defines **position/location privacy**. Besides, private or sensitive places are also prevented from adversaries. Many of them such as home, office, hospital, political buildings, items of interest, etc. could be feasible background information leading to re-identifying individuals.

#### 3.5.4. Path/Trajectory Privacy

Relying on location privacy, **path/trajectory privacy** is defined as the ability to prevent other unauthorized parties from learning users’ motions in a close relationship between positions and time. Moreover, sensitive routes can be formed from paths through private or sensitive places. Getting them is usually useful for adversaries to meet their needs, so they have to be preserved, too.

#### 3.5.5. Quasi-Identifier Privacy

Quasi-identifier is a set of attributes that can potentially lead to identity breaches. An adversary exploits them, together with some external information, to re-identify a user. For example, the attribute values consisting of birth date, zip code, and the gender can uniquely determine an individual. So **quasi-identifier privacy** recommends that the values from quasi-identifiers should be kept in private but still assured the utility of the data set in which they are.
3.5.6. Privacy of Sensitive Attributes

In some location-based applications, especially in trajectory data publication, all individual attributes of a user (e.g., salary, disease, disability status, marriage status), not joining quasi-identifier, should be maintained in secret. These should be guaranteed due to the fact that everyone has the right to keep his own personal information, which is shown in (Westin, 1967).

3.6. Challenges

Preserving data privacy is always one of the most emerging issues in database systems, especially in modern information systems. For those who want to achieve their goals have to overcome these challenges: (1) How to make privacy be in close co-operation with both security and policy lays a firm principle on the prosperous development of LBS; (2) There are lots of criteria to classify LBS applications while each also has its own privacy problems and protection mechanisms; (3) Bear in mind that protecting identity including removing the true one or simply using a pseudonym is not sufficient; (4) Private places or items of interest should be also considered; (5) The quality of services should be guaranteed, along with user policy, trade-offs, personalization, efficient query processing, and system performance. Especially due to mobility and some measurement error, the query answer has to be assured; (6) It is a good way to care about both spatial and temporal dimensions and location privacy and query privacy when we deal with user privacy to minimize attacks from adversaries; (7) The more efficiently the variety of queries are supported, the more widely the location-based applications are used; (8) Privacy techniques should be context-aware in order to gain more effectively; (9) Each privacy technique may be attached to some assumptions or constraints. That means it also deals with some attack models. The closer it is to the real world, the most applicable it is to LBS, or at least these assumptions are reasonable; (10) Dealing with exceptions should not be ignored (GPS signal loss because of mobile devices’ power limitation, device failure, or breakdowns of all kinds for example); (11) It is necessary to have a measure (i.e., privacy metric) for user privacy of all kinds as a criterion to judge whether proposed solutions are good enough or not.

Depending on what kinds of trajectory-based applications, we should pay attention to either of issues listed below or their combination: (1) Trajectory preservation: assures that we can preserve the utility of data as much similarity as possible like what they are in the real world and then helps us answer trajectory-based queries efficiently; (2) The outbreak of data: indicates that when time goes by, trajectory data grow quickly. There should be a way to control or deal with this phenomenon; (3) Privacy preservation: especially needed when mining, publishing, or learning some useful information hidden in trajectory data. User privacy should be kept personally and not be violated; (4) Movement constraints: care about what kind of trajectories there are, for one can use lots of different transportations for his moves. If cared properly, they contribute to the quality of services; (5) Data stream management: deals with continuous location updates or queries while given accuracy remains unchanged. Especially in real-time applications, there are numerous data collected from positioning-aware mobile devices or sensors. Therefore, it requires essentially a flexible mechanism to collect and process consecutive data efficiently in an effort of increasing QoS and its performance.

A lot of challenges have been recognized during user privacy analysis. We can possibly take them off the table, but not all of them are resolved in order to have a good privacy method or solution. Among such challenges, some are mutual, but others can be incompatible. One solution is supposed to be fit once it’s suited to its application and objectives.
4. Our Future Work

Although one’s location privacy is assured when location privacy techniques are applied, it may still be exposed once his trajectory is identified. As a consequence, his sensitive data will be leaked. Motivated by this, our research is focused on trajectory-based privacy-preserving problems whereas there is little work in this field.

To cope with these concerns, we can put our investigation in application, publication, or database level, which depends much more on location-based applications, strategies, and what we are going to protect. However, database-centric approach seems to be the most potential above all others. Our approach is how to find out an efficient privacy-preserving access method for moving objects. In other words, an index structure integrated with user privacy will be studied at the database level while previous related work just only cares about how to have an efficient index for moving objects. This approach is chosen in comparison with earlier introduced approaches because of its advantages such as flexibility, efficiency, and performance. Furthermore, it easily spreads to other levels once there is a need (e.g., indexing for data mining or application management).

Alternatively, our goal aims at indexing for the predictability of trajectories supporting user privacy. Such an interesting and significant research direction becomes much more exciting for the next generation of mobile services, for it promotes location-based applications pointing at management, knowledge-based decision making, service optimization, forecast, and intelligence as well. These kinds of LBS inspire more research and industry, promise a strong growth, widen application domains, strengthen users’ expectation, and benefit from what makes them possible.

5. Conclusion

In this paper, a general view and related components about trajectory-based privacy-preserving protection are given. There are lots of criteria, characteristics, and challenges needing to be satisfied and explored deeply in an attempt on maintaining the balance between user privacy and QoS. At the same time, this work also introduces lots of relevant research directions leading to open problems that should be investigated more in the future.

In addition, we address our approach towards database level in dealing with trajectory privacy and its related issues, which can give a potential support in resolving trajectory-based privacy-preserving location-based applications.
6. References


