From Data Privacy to Location Privacy: Models and Algorithms

Ling Liu
Distributed Data Intensive Systems Lab
School of Computer Science, Georgia Institute of Technology
lingliu@cc.gatech.edu

ABSTRACT

This tutorial presents the definition, the models and the techniques of location privacy from the data privacy perspective. By reviewing and revising the state of art research in data privacy area, the presenter describes the essential concepts, the alternative models, and the suite of techniques for providing location privacy in mobile and ubiquitous data management systems. The tutorial consists of two main components. First, we will introduce location privacy threats and give an overview of the state of art research in data privacy and analyze the applicability of the existing data privacy techniques to location privacy problems. Second, we will present the various location privacy models and techniques effective in either the privacy policy based framework or the location anonymization based framework. The discussion will address a number of important issues in both data privacy and location privacy research, including the location utility and location privacy trade-offs, the need for a careful combination of policy-based location privacy mechanisms and location anonymization based privacy schemes, as well as the set of safeguards for secure transmission, use and storage of location information, reducing the risks of unauthorized disclosure of location information.

The tutorial is designed to be self-contained, and gives the essential background for anyone interested in learning about the concept and models of location privacy, and the principles and techniques for design and development of a secure and customizable architecture for privacy-preserving mobile data management in mobile and pervasive information systems. This tutorial is accessible to data management administrators, mobile location based service developers, and graduate students and researchers who are interested in data management in mobile information syhhhstems, pervasive computing, and data privacy.

1. DESCRIPTION

Location Privacy is a particular type of data privacy. It is defined as the ability to prevent other unauthorized parties from learning ones' current or past location. In Location Based Services (LBSs), there are conceivably two types of location privacy: *personal subscriber level privacy* and *corporate enterprise-level privacy*. Personal subscriber-level privacy must supply rights and options to individuals to control when, why, and how their location is used by an application. With personal subscriber-level privacy, each individual has liberties to ``opt in" and ``opt out" of services that take advantage of their mobile location. Corporate enterprise-level privacy is fundamentally different in that corporate IT managers typically control when, why, and how mobile location capabilities provide application benefits to the organization as a whole. Within the enterprise, personal subscriber-level privacy is sometimes irrelevant because location is a critical requirement for staff to function productively while on the road. Asset tracking and workforce management are examples of location-enabled enterprise applications. However, companies need enterprise-level privacy to preserve corporate secrets and maintain competitive edge.

Location Privacy Threats refer to the risks that an adversary can obtain unauthorized access to raw location data, derived or computed location information by locating a transmitting device, hijacking the location transmission channel, and identifying the subject (person) using the device. For example, location information can be used to spam users with unwanted advertisements or to learn about users' medical conditions, alternative lifestyles or unpopular political views. Inferences can be drawn from visits to clinics, doctors' offices, entertainment districts, or political events. In extreme cases, public location information can lead to physical harm, for example, in stalking or domestic abuse scenarios [5,6].

Several approaches have been proposed for protecting location privacy of a user. Most of them try to prevent disclosure of unnecessary information by techniques that explicitly or implicitly control what information is given to whom and when. These techniques can be classified into three categories: (1) Location protection through user-defined or system-supplied privacy policies [3,5]; (2) Location protection through anonymous usage of information, such as location cloaking, by reducing temporal and spatial resolutions of location information [1,2,3,6,7]; and (3) Location protection through pseudonymity of user identities, which uses an internal pseudonym rather than the user's actual identity [5]. Such pseudonyms should be different for different services and frequently changing to prevent applications tracking them. More importantly, such pseudonyms should be generated in

VLDB '07, September 23-28, 2007, Vienna, Austria.

such a manner that makes the linking between the old and the new pseudonym very hard [5].

Location Privacy and Location Service Quality
On one hand, the quality of the LBS depends on the accuracy of the location of mobile users, and on the other hand, the more accurate the location information is disclosed, the higher risk of location privacy is being invaded. There is an inherent tradeoff between the utility of a LBS that users wish to receive and the location privacy they can afford to risk. An important question is how much privacy protection is necessary. Perfect privacy is clearly impossible as long as communication takes place. Moreover, different users may have varying privacy needs in different contexts. Furthermore, location privacy is context sensitive. Different users may require different levels of privacy at different times. A user's willingness to share location data may depend on a range of factors, including different contextual information about the user. Therefore, it is important to develop customizable/personalized privacy protection mechanisms that can help users finding a comfortable balance between the extreme of fully disclosed and completely withheld location data. This includes (i) the qualitative and quantitative analysis of the inherent tradeoff between the quality of service provided by the LBS and the desired location privacy of the user, (ii) how to determine and model the fuzziness of the location information sent by a mobile user to the LBS in order to reach such a tradeoff, and (iii) what types of user-defined privacy rules need to be combined with a personalized anonymization model to allow users to tailor the system-level privacy protection strategies to meet their personal privacy preferences.

Location Anonymization is a system capability to obfuscate the location information such that a state of a subject is not identifiable within the anonymity set. Depending on the definition of the state and the definition of the anonymity set, the goal and the process of location obfuscation may differ for different users and in different contexts. We argue that there is a need for more privacy requirement measures than merely location *k*-anonymity. In this tutorial we will discuss location *l*-diversity and location *m*-invariant and how they compliment the location *k*-anonymity in supporting location privacy of varying degree [2,3].

According to [5], anonymity can be seen as ``a state of being not identifiable within a set of subjects, the anonymity set''. The concept of *k*-anonymity is originally introduced in the context of relational data privacy research [9]. In the context of LBSs and mobile users, location *k*-anonymity refers to *k*-anonymous usage of location information. A subject is considered *k*-anonymous with respect to location information if and only if the location information sent from a mobile user to a LBS is indistinguishable from the location information of at least *k-1* other subjects (e.g. different mobile nodes) [6]. A larger *k* indicates more uncertainty in linking a location to a particular user. Location *k*-anonymity typically refers to k different users (subjects) forming the anonymity set. However, when all *k* users reside in the exact same location (such as the same clinic office or church), the location *k*-anonymity alone fails to prevent the location of a subject being not identifiable. Thus, we argue the importance of incorporating location l-diversity to allow the mobile users to provide the second dimension of location anonymization in terms of a set of distinct locations (such as postal addresses, and identifiable symbolic addresses). By providing both location *k*-anonymity and location *l*-diversity, mobile users can use the value of *k* and the value of *l* in her location privacy policy as the parameters to control her desired level of privacy in terms of the set of subjects (the anonymity set of users) and the set of locations (the anonymity set of locations). One can further introduce other location privacy measures such as m-invariant to control the number of alternative routes a mobile user would like to maintain anonymous in situations where all *k* users are traveling along the same route segments passing through the same set of identifiable locations.

Location perturbation is known to be an effective technique for implementing location *k*-anonymity. The technical challenge is how to adequately control the location cloaking process in terms of location *k*-anonymity, location *l*-diversity, and location *m*-invariant efficiently to meet both location privacy and location service quality requirements. Many research results, such as distance preserving transformation techniques [4] and variants of k-anonymity [10] in data privacy area can be extended to protecting location privacy.

ACKNOWLEDGMENTS: This research is partially supported by grants from NSF CyberTrust, HIS, IBM faculty award.

2. REFERENCES

[1] Bugra Gedik and Ling Liu. ``Protecting Location Privacy: A Personalized Anonymization Model. IEEE ICDCS 2005.

[2] Bugra Gedik and Ling Liu. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms. IEEE Transactions on Mobile Computing, 2007 (to appear).

[3] Bhuvan Bamba and Ling Liu. PrivacyGrid: Supporting Anonymous Location Queries in Mobile Environments. Technical Report, May 2007. Georgia Institute of Technology.

[4] Keke Chen and Ling Liu. A Random Rotation Perturbation Approach to Privacy Preserving Data Classification, ICDM'05.

[5] A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. IEEE Pervasive Computing, 2003

[6] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. MobiSys 2002.

[7] Mohamed Mokbel, Chi-Yin Chow, Walid Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. VLDB 2006.

[8] Gabrief Ghinita, Panos Kalnis, and Spiros Skiadopoulos. Prive: Anonymous Location-based Queries in Distributed Mobile Systems. WWW 2007.

[9] P. Samarati and L. Sweeney. Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement through Generalization and Suppression, SRI-CSL-98-04.

[10] A. Machanavajjhala, J. Gehrke, D. Kifer. *l -Diversity*: Privacy Beyond k -Anonymity. ICDE06.