Location Privacy Issues in Pervasive Environments

Dan Lin
Department of Computer Science
Missouri University of Science and Technology
lindan@mst.edu

Abstract

The expanding use of location-based services has profound implications on the privacy of personal information. If no adequate protection is adopted, information about movements of specific individuals could be disclosed to unauthorized subjects or organizations, thus resulting in privacy breaches. In this paper, we provide an overview of a framework for preserving location privacy in pervasive environments. We also discuss potential research topics along this direction.

1 Introduction

With advances in positioning system (e.g. GPS) and wireless communication, tracking of continuously moving objects like vehicles and users with wireless devices, has become feasible in terms of technology and implementation cost. This fosters a new type of services, namely location-based services, which can help travelers locating nearby gas stations and restaurants, provide real-time traffic information, etc. Recently, AT&T has announced that it has deployed assisted GPS technology (A-GPS) within its wireless network to enhance existing and planned location-based services used with A-GPS capable devices. Google map also provides up-to-date traffic condition for more than 30 major U.S. cities.

The very nature of the location-based services is the need to track user movements and locations. It is then easy, based on such information, to discover users' habits and other personal information. However, this brings up the issue of privacy, i.e. exposure and potential misuse of personal information. There is therefore an important concern for location privacy: "how to prevent other parties (including service providers) from learning one's current or past location?" For example, Alice sends a query to her service provider to check whether one of her friends is within one mile of her location so that they can go out for lunch together. Neither she nor her friends would like to disclose their exact locations to the service provider, but they want the correct query results. The problem becomes more challenging when Alice is travelling on the road and sending out continuous queries about nearby gas stations or other locations of interests, which requires anonymizing a trajectory rather than a single location to preserve location privacy.

In this paper, we first review the advantages and disadvantages of existing proposals for location privacy preservation. Then we propose a 3-tier framework and discuss the challenges and potential research topics towards the realization of such framework.

The rest of the paper is organized as follows. Section 2 reviews state-of-art work in the area of location privacy. Section 3 presents an overview of potential research directions. Section 4 concludes the paper.

2 Literature Review

In this section, we review recent work grouped into two categories: (i) k-anonymity-based approaches; (ii) cryptographic-based approaches.

2.1 k-anonymity-based Approaches

Many approaches [1, 2, 3, 5, 6, 7, 11] focus on the development of anonymization techniques to hide users' location information from service providers.

The anonymization is typically carried out by a trusted agent. One simple anonymization is to use pseudo name for users. However, such approach cannot provide sufficient location privacy protection because malicious service providers may utilize background information to map pseudo names to the actual users.

A popular anonymization technique is spatial-temporal cloaking, originally introduced by Gruteser et al. [6]. In their approach, the agent represents a user's location by a region in which at least other $k-1$ users are also present. A service provider receives regions instead of exact locations of users, and hence, it is hard for the service provider to distinguish the users in the same region. However, such approach sacrifices query accuracy. For example, a nearest neighbor of a region may not be the nearest neighbor of a user in the region. Based on the same idea, Gedik et al. [3] proposed a more flexible approach which does not require a global k-parameter but instead allows each user to define its own k-parameter. However, this approach still cannot provide accurate query results. The spatial-cloaking-based model was later enhanced by Mokbel et al.[11]. They improved the query processing at the server side and ensured that a super set of accurate query results is found. This step is then followed by a filtering step at the agent side to remove the false positives.

More recently, Ghinita et al. [5] utilized Hilbert space filling curve to group users into buckets of k. The algorithm works in P2P environments where each user can be the anonymization agent. In this way, the risk of a single agent failure is reduced. However, the approach is not pervasive and requires extensive user involvement.

Existing approaches have one or more of the following shortcomings: First, most of them trust a single anonymization agent and allow the agent to store location information about users, which makes the agent the single target of attacks by malicious parties. Second, the single agent is a performance bottleneck since it needs to process all updates and anonymization. Finally, they only consider privacy of a snapshot of user locations and do not provide any protection on possible attacks on users' trajectory. For example, in case of continuous query (i.e., a user keep sending queries), the user can be easily identified because his/her successive cloaking regions disclose his/her moving trend.

2.2 Cryptographic-based Approaches

Another line of approaches is based on cryptographic techniques. Hore et al. [8] suggested encrypting sensitive data and using a privacy-preserving index for executing range queries over encrypted data. Khoshgozaran et al. [9] proposed a one-way transformation to encode all static and dynamic objects and resolve the k nearest neighbor query blindly in the encoded space. However, these two techniques are unable to generate accurate query results in general. Most recently, Ghinita et al. [4] employed Private Information Retrieval (PIR) technique to prevent service providers from knowing users' location information, which does not require a trusted third-party agent and can return exact results.

Cryptographic-based techniques have two common limitations: First, they support restricted types of queries. For example, [4] only supports queries on static public objects. Second, encryption techniques are usually too costly to be applied in practice.

3 Research Directions

We propose a 3-tier location privacy protection framework [10] which consists of a service provider, multiple agents and users. The main idea is to utilize the agents to carry out appropriate location transformation to distort the true locations, so that the service providers are prevented from knowing the users' real locations and only serve as a computing engine. The specific data flow is as follows. Each time a user needs to update its position, it does not directly contact the server; instead, it randomly selects an agent to which it sends its location information. When querying, a user's query is broadcasted to all agents. The agents are responsible for executing a transformation on the user data or queries, and pass the transformed data to the server. The server processes the transformed data and returns the query results to the agents. After receiving the results from the server, the agents perform a reverse transformation and filtering before returning the results to the user. Such a framework has advantages in terms of both privacy and performance efficiency. First, no single entity (agents or server) is able to track the movement of any user without colluding with other entities in the system. Because each agent only collects a subset of the locations of each user in the system, the level of trust required from each agent does not need to be high. Moreover, the use of multiple agents allows multiple transformations to be applied to the data by the same user. This makes it much harder for the server to keep track of the relative distance among users. Second, by employing multiple agents, data is grouped according to agent IDs at the server side. That means updates and queries will be processed in a relatively small dataset and hence the response time will be shortened.

To realize the framework, the biggest challenge is to design effective location transformation functions, which can preserve relative distance in each subdataset corresponding to each agent. For example, if user A is the nearest neighbor of B in the original space, A should also be closest to B after the location transformation so that the server can compute the correct result of the nearest neighbor query. One possible solution is to employ the combination of three basic types of transformations: scaling, rotation, translation, since each has the relative-distance preserving property. It would be interesting to explore other types of transformation functions which may be more efficient or provide better privacy protection. One thing needs to be considered during the design is whether the continuous movement of the same user can be detected from the transformed data.

The second problem is the query transformation and query result assembling, i.e., how to transform a user query into queries being processed on transformed subdatasets at the server and then collect all the results to form the final answer. Existing query algorithms for moving objects need to be revised.

Finally, it is very important to explore potential threats to a location-based service system as thorough as possible. Based on the understanding of privacy risks, a privacy model and privacy metrics need to be developed for quantifying and evaluating the proposed approaches.

4 Conclusion

In this paper, we present an overview of a privacy protection framework and point out several promising research directions. As more and more people's location information become available and more and more people subscribe to location-based services, the research on location privacy issues will have a tremendous impact on our daily life.

References

[1] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[2] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. Preserving user location privacy in mobile data management infrastructures. In *Proc. Workshop on Privacy Enhancing Technologies*, 2006.

[3] B. Gedik and L. Liu. A customizable k-anonymity model for protecting location privacy. In *Proc. IEEE ICDCS*, pages 620–629, 2005.

[4] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L.Tan. Private queries in location based services: Anonymizers are not necessary. In *Proc. SIGMOD*, 2008.

[5] G. Ghinita, P. Kalnis, and S. Skiadopoulos. Prive: Anonymous location-based queries in distributed mobile systems. In *Proc. World Wide Web*, pages 371–380, 2007.

[6] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. MobiSys*, pages 31–42, 2003.

[7] M. Gruteser and X. Liu. Protecting privacy in continuous location-tracking applications. *IEEE Security and Privacy*, 2(2):28–31, 2004.

[8] B. Hore, S. Mehrotra, and G. Tsudik. A privacy-preserving index for range queries. In *Proc. VLDB*, pages 720–731, 2004.

[9] A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In *Proc. SSTD*, pages 239–257, 2007.

[10] D. Lin, E. Bertino, R. Cheng, and S. Prabhakar. Position transformation: A location privacy protection method for moving objects. In *Proc. ACM SPRINGL*, 2008.

[11] M. F. Mokbel, C. Y. Chow, and W. G. Aref. The new casper: Query processing for location services without compromising privacy. In *Proc. VLDB*, pages 763–774, 2006.