Privacy in Location-based Services: A System Architecture Perspective

Chi-Yin Chow Mohamed F. Mokbel
Department of Computer Science and Engineering
University of Minnesota
{cchow, mokbel}@cs.umn.edu

Introduction

Location-based services (LBS, for short) are information and entertainment services that are conveniently accessible by mobile users through GPS-enabled portable devices and mobile networks (e.g., 2G/3G cellular telephone and Wi-Fi networks). Examples of LBS include resource finding (e.g., where is my nearest gas station), route finding (e.g., what is the shortest route from my current location to a shopping mall, social networking (e.g., where are my friends), and location-based gaming (e.g., GPS online game). LBS rely mainly on an implicit assumption that mobile users are willing to reveal their private locations. With untrustworthy LBS providers, the revealed private location information could be abused by adversaries. For example, an adversary may infer a user's medical record by knowing that she regularly visits a specialized clinic. There are already several life scenarios that took place where personal GPS locations are abused, e.g., see [8, 27, 28].

Unfortunately, the traditional approach of *pseudonymity*, i.e., using a fake identity, cannot overcome such a privacy threat in LBS, where personal locations can be used as identities. For example, asking about the nearest Pizza restaurant to a personal house using a fake identity will immediately reveal the customer identity as a resident of the house. Recently, there is huge interest to enable privacy-preserving LBS in which users can entertain high quality location-based services without compromising their privacy. In general, two main issues need to be considered: (a) anonymizing personal locations, and (b) obtaining high quality services on top of the anonymized locations. In this article, we will briefly discuss these two main research issues with respect to five different system architectures for privacy-preserving LBS. Then, we will discuss future research directions.

Client-Server Architecture

This is a centralized architecture where mobile users directly communicate with the LBS provider. Existing work in this architecture can be classified into three main categories. (1) *False dummies* [22]. For every location update, a user sends n different locations to the server where only one of them is true while the rest are *dummies*. Thus, the server cannot know which one of these reported locations is the actual one. The query processor finds an answer set that includes the answer to each location. After the user gets the answer set, she computes the exact answer. (2) *False locations* [17, 31]. The main idea is that users will send false location(s) to the server. This approach can go as simple as just sending the location of a nearby landmark or a significant object

to the user location, in which the database will give the query answer with respect to the chosen landmark [17]. A much better approach, i.e., more accurate, is *Space twist* [31]. where a user sends a nearest-neighbor query along with a false location to a database server, the database server keeps sending the nearest objects to the false location to the user. The user caches the received objects and terminates the request until the answer derived from the cached objects satisfies the user privacy and accuracy requirements. (3) *Space transformation* [10, 21]. This approach converts the original location information of data and queries into another space through a third party. The space transformation maintains the spatial relationship among the data and query, in order to provide approximate query answers [21] or exact query answers [10] obtained through private information retrieval.

Trusted Third Party Architecture

The main idea of this architecture is to employ a trusted third party, termed *location anonymizer*, to be placed between mobile users and the LBS provider. The location anonymizer is responsible for blurring user locations into cloaked areas that satisfy user's personalized privacy requirements [1, 2, 7, 9, 13, 19, 23, 29, 30]. In this case, the user privacy requirements are mostly presented in terms of the K-anonymity model [25, 26], i.e., a cloaked area A contains at least K users making each user indistinguishable among at least K users. Other location anonymization techniques employ this architecture approach for avoiding location tracking for continuous location updates [14, 16] or continuous queries [2, 29, 30].

With the *location anonymizer*, the trusted third party architecture supports three new query types for privacy-preserving LBS [23], namely, *private queries over public data* (e.g., a person (private query) asks about nearest gas station (public data)), *public queries over private data* (e.g., an administrator (public query) asks about the number of mobile users (private data) within a certain area), and *private queries over private data* (e.g., a person (private query) asks about her nearest buddy (private data)). Since the query processor embedded inside the database server does not know the actual location information of the query and/or data, it can return only an answer set that includes the exact answer to the user regardless of the actual user's location within the cloaked area. The existing privacy-aware query processing frameworks can deal with rectangular cloaked areas [6, 18, 23, 24] or circular cloaked areas [19] as the query and/or data location information.

Distributed Architecture

In this model, mobile users communicate with each other through a fixed communication infrastructure, e.g., base stations [11, 12]. The basic idea of the location anonymization techniques in this architecture is that users collaborate with other peers to maintain a distributed data structure where the stored location information is used by the users to blur their location information into K-anonymous cloaked areas. Then, the query processing could be similar to the one used in the trusted third party architecture where the user sends to the server its query along with a cloaked area that includes the user location.

Mobile Peer-to-Peer Architecture

In mobile peer-to-peer networks, there is no fixed communication infrastructure or centralized/distributed servers. Instead, mobile users directly communicate with their peers through multi-hop routing to blur their locations into cloaked areas that satisfy their personalized K-anonymity and/or minimum area privacy requirements [5]. Similar to the *distributed model*, the

proposed peer-to-peer location anonymization technique uses the privacy-preserving query processing framework designed for the trusted third party architecture. After a user finds a cloaked area as her location, she randomly selects a peer within the cloaked area as an agent. The user sends the query along with the cloaked area to the agent, and then the agent communicates with the database server on behalf of the user. When the agent gets an answer set from the database server, the agent forwards the answer set to the user. Finally, the user computes the exact answer from the answer set.

Wireless Sensor Networks

Research in wireless sensor networks include two main directions: (a) Dividing the system space into hierarchical levels based on physical units, e.g., sub-rooms, rooms, and floors [15]. If a unit contains at least K users, the algorithm cloaks the subject count by rounding the value to the nearest multiple of K. Otherwise, the algorithm cloaks the location of the physical unit by selecting a suitable space containing at least K users at a higher level. Then, the query processing will be similar to the one used in the trusted third party architecture. (b) Providing an in-network location anonymization algorithm that is suitable for both indoor or outdoor environments regardless of the system's physical structure [3]. The main idea is to let sensor nodes provide aggregate information about the monitored mobile objects. Then, the database server employs a spatio-temporal histogram that estimates the actual object distribution in the system based on the anonymized location information [3, 4]. The database server uses the estimated object distribution to answer range queries that are used to provide aggregate location monitoring services in wireless sensor networks.

Future Directions

Although many research efforts have been focused on privacy-preserving LBS, there still exist many open research issues and challenges in this area that include:

Users' prospective. Existing privacy-preserving LBS frameworks are designed from the technology's prospective. There is still need to study the location privacy issue from the user's prospective. For example, how can a *casual* user define privacy requirements. Is it possible to define privacy levels as low, medium, and strict, and then users would choose among them. How can a user achieve a trade-off between the privacy requirements and the quality of services. How can the user evaluate the privacy risk she has from using a certain LBS.

Privacy measures and adversary attacks. There is a need to define a formal privacy measure and adversary attacks of anonymized location information in different environment settings, e.g., the Euclidean space, road network, and wireless sensor networks, and for different privacy-aware query types, e.g., static and continuous queries. Such measures and attacks can be used to evaluate the degree of privacy protection of existing and forthcoming location anonymization techniques in terms of the tradeoff between privacy and system performance.

Privacy-aware location-based query types. Existing privacy-preserving LBS frameworks support only private range and nearest-neighbor queries over public or private data. One of the future directions is to extend existing frameworks to support other kinds of location-based queries, e.g., reverse nearest-neighbor queries [20] and aggregate nearest-neighbor queries [32] where the query processor does not know the actual location information about the query and/or data.

Road networks environments. Existing location privacy techniques mainly consider the Euclidean space where users can move freely. In reality, most of the object movement is constrained

by the underlying road network. Applying existing location privacy techniques directly to the road network environment is not practical as adversaries would have more information about the possible user locations, derived from the knowledge of the underlying road network. Thus, it is important to design new specialized location anonymization and privacy-preserving query processing techniques for road network environments.

References

[1] B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting anonymous location queries in mobile environments with privacygrid. In *WWW*, 2008.

[2] C.-Y. Chow and M. F. Mokbel. Enabling private continuous queries for revealed user locations. In *SSTD*, 2007.

[3] C.-Y. Chow, M. F. Mokbel, and T. He. Tinycasper: A privacy-preserving aggregate location monitoring system in wireless sensor networks (Demonstration). In *SIGMOD*, 2008.

[4] C.-Y. Chow, M. F. Mokbel, and T. He. Aggregate location monitoring for wireless sensor networks: A histogram-based approach. In *MDM*, 2009.

[5] C.-Y. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based services. In *ACM GIS*, 2006.

[6] C.-Y. Chow, M. F. Mokbel, J. Nap, and S. Nath. Evaluation of range nearest-neighbor queries with quality guarantee. In *SSTD*, 2009.

[7] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *PERVASIVE*, 2005.

[8] Foxs News. Man accused of stalking ex-girlfriend with GPS, `http://www.foxnews.com/story/0,2933,131487,00.html`. September 4, 2004.

[9] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *TMC*, 7(1):1–18, 2008.

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

[11] G. Ghinita, P. Kalnis, and S. Skiadopoulos. Privé: Anonymous location-based queries in distributed mobile systems. In *WWW*, 2007.

[12] G. Ghinita, P. Kalnis, and S. Skiadopoulos. Mobihide : A mobile peer-to-peer system for anonymous location-based queries. In *SSTD*, 2007.

[13] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *MOBISYS*, 2003.

[14] M. Gruteser and X. Liu. Protecting privacy in continuous location-tracking applications. *IEEE Security and Privacy*, 2(2):28–34, 2004.

[15] M. Gruteser, G. Schelle, A. Jain, R. Han, and D. Grunwald. Privacy-aware location sensor networks. In *HOTOS*, 2003.

[16] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *MOBISYS*, 2008.

[17] J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *MOBISYS*, 2004.

[18] H. Hu and D. L. Lee. Range nearest-neighbor query. *TKDE*, 18(1):78–91, 2006.

[19] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias. Preventing location-based identity inference in anonymous spatial queries. *TKDE*, 19(12):1719–1733, 2007.

[20] J. M. Kang, M. F. Mokbel, S. Shekhar, T. Xia, and D. Zhang. Continuous evaluation of monochromatic and bichromatic reverse nearest neighbors. In *ICDE*, 2007.

[21] A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In *SSTD*, 2007.

[22] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *ICPS*, 2005.

[23] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The new casper: Query procesing for location services without compromising privacy. In *VLDB*, 2006.

[24] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The new casper: A privacy-aware location-based database server (Demonstration). In *ICDE*, 2007.

[25] P. Samarati. Protecting respondents' identities in microdata release. *TKDE*, 13(6), 2001.

[26] L. Sweeney. k-anonymity: A model for protecting privacy. *IJUFKS*, 10(5):557–570, 2002.

[27] USA Today. Authorities: GPS system used to stalk woman, `http://www.usatoday.com/tech/news/2002-12-30-gps-stalker$\_$x.htm`. December 30, 2002.

[28] J. Voelcker. Stalked by satellite: An alarming rise in GPS-enabled harassment. *IEEE Spectrum*, 47(7):15–16, 2006.

[29] T. Xu and Y. Cai. Location anonymity in continuous location-based services. In *ACM GIS*, 2007.

[30] T. Xu and Y. Cai. Exploring historical location data for anonymity preservation in location-based services. In *INFOCOM*, 2008.

[31] M. L. Yiu, C. Jensen, X. Huang, and H. Lu. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *ICDE*, 2008.

[32] M. L. Yiu, N. Mamoulis, and D. Papadias. Aggregate nearest neighbor queries in road networks. *TKDE*, 17(6), 2005.