Techniques for countering privacy threats of Location-based services

Lauri Suomalainen

Seminar
UNIVERSITY OF HELSINKI
Department of Computer Science

Helsinki, March 3, 2016

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Computer Science

Tekijä — Författare — Author

Lauri Suomalainen

Työn nimi — Arbetets titel — Title

Techniques for countering privacy threats of Location-based services

Oppiaine — Läroämne — Subject

Computer Science

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Seminar	March 3, 2016	10

Tiivistelmä — Referat — Abstract

Abstract

Avainsanat — Nyckelord — Keywords

Geo-location, Privacy, Social Networks

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

List of Figures

Abstract

Most of the social networking services today exploit location-based data in one way or another. Services like Facebook allow users to GeoTag their current location and tag themselves and their friends. In exchange the service can use the data to offer recommendations, news and so on. While the users of the services are often aware that they are providing personal locational data, its pervasiveness and accuracy may come as a surprise and have serious repercussions when it comes to users' real-life privacy. A malicious actor such as a burglar could exploit the data a user provides to for example find out their street address and times when they are not home. This seminar paper reviews and evaluates several techniques used to preserve and protect users' privacy when using Location-based services. The goal for the techniques is to counter several threats Location-based Services face in such manner that users can still use said services without compromising their security.

1 Introduction

Location-Based Services (LBS) are applications that operate on geographical or other location based data. Original LBSs can be traced back to mid-20th century, but what is understood by LBS today is tied to the mid-21th century emergence of G3 networks and hand-held devices with GPS capability [1]. Many contemporary LSBs take a form as a part of a social networking applications such as Facebook, Foursquare and Instagram. There are several ways LBSs work. They can make use of users' mobile device's GPS and provide services based on users' whereabouts when requested or continually monitor users' location [9]. Some services allow users to *check-in* to pre-determined venues or points of interest (POI) whereas other allow more exact location with the GPS data. This practice is commonly known as *GeoTagging*.

LBSs do not come without security risks. By publishing their location on the internet exposes users to threats in the real world. A malicious agent, for example a burglar, could monitor one user's LBS usage and e.g. determine when the user is not at their residence. Using that information the attacker could orchestrate a break-in without the risk of being caught red-handed [2]. While many applications require user to explicitly publish

their locational data, users also often do it unknowingly. For example many contemporary smart phones as well as digital cameras automatically embed metadata to photographs and it can contain privacy sensitive information such as coordinates and the time the photograph was taken [2]. In some cases the user themselves does not expose their location data explicitly, but with certain LBSs supporting *User Tagging*, a user's acquaintance can expose the user without them having a say for it. Thus a user can be associated with places, people and other information they would rather keep private.

The obvious privacy issues have been researched a lot and many techniques have been presented which address one or more possible privacy threats LBSs face [9]. These range from offering users options to rule how, where and when can their locational and identity data used to more technical solutions such as query enlargements and encryption-based techniques. Each of them have have their uses and shortcomings: They can only address a certain set of privacy threats and may come with pre-assumptions and computational costs.

In this paper we first review threats LBSs face in section 2. Then in section 3 we take a look at the techniques designed to address different kinds of privacy threats in LSBs. Section 4 compares and evaluates the techniques in respect of their intended use as well as other techniques. Section 5 will be the concluding part summarising the state of privacy in Location-Based Services and also shortly addressing the future trends of the field.

2 Threats in Location-based Services

A privacy threats in LBS can be defined as events during which an adversary can gain information about the user which they consider sensitive[9]. Threats fall into two major categories: Release of sensitive location information and re-identification through location information. In the first case, the identity of the user is known but the location information associated with that identity is considered sensitive and somehow made available for the adversary. In the latter case the user would like to have their identity kept secret but the adversary can exploit their location data to narrow down their possible identities in the set of other associated identities thus reducing their degree of anonymity. [9] list four different privacy threat vectors and [4] add one more. The vectors are location privacy, absence privacy, co-location privacy,

identity privacy and correlation privacy.

2.1 Location Privacy

Location privacy is the most straightforward of attack vectors. Whenever a user publishes their location they associate their identity with that location and possibly reveal information they consider sensitive. Even though a user can generalise their location it is not guaranteed to be safe. For example, a user would GeoTag themselves in a LBS whilst visiting a bar with a friend, but instead of using the exact location they would use the more general location like the part of town as the tag. Now if the aforementioned friend would happen to meet another friend at the bar too and GeoTag them both but now using the exact location of the bar, an adversary with the access of both GeoTags could easily deduct first user's exact location as well as the fact that they are accompanied by the third friend, both being pieces of information the first user did not explicitly wish to expose [9].

2.2 Absence Privacy

Absence Privacy can be regarded as the other side of the location privacy. As the user publishes information about where they are at the given moment, they simultaneously expose explicitly where they are not. This threat factor also has a temporal dimension. For example, an adversary planning a break-in to a user's home can use the user's location information to estimate how long they will approximately be absent in order to avoid being caught while on the crime scene.

2.3 Co-location Privacy

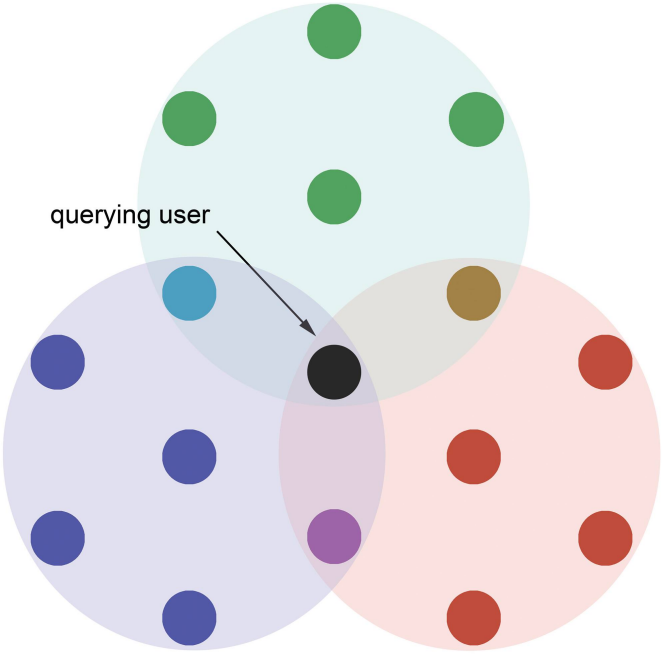
Co-location Privacy refers to a privacy threat in which the adversary monitors multiple users' location information in order to associate other users with each other and different locations. For example, if user A GeoTags themselves and their friend user B to a location and simultaneously user C tags themselves and their friend user D at the same place, an adversary with the access to the information of only A and C can possibly violate the privacy of both B and D as being associated with someone on the other party at the given place can be considered sensitive by users.

2.4 Identity Privacy

Identity Privacy refers to a scenario in which a user employs an obfuscated identity such as a pseudonym to protect their real identity but their identity can be revealed by associating them with a certain location. For example, the user is a 30-year-old male who is using a location-based dating service at his workplace. In the service he uses a pseudonym because he considers the fact that he is using the service to be embarrassing and thus sensitive. Now if he happened to be the only male of 30 years at his workplace and his supervisor would happen to check out the dating service and only search for users in their immediate vicinity, the supervisor would find out that the 30-year-old user is using the service during work hours effectively violating the user's privacy. In a scenario like this the pseudonym does naught to protect the users real identity.

2.5 Correlation Privacy

Correlation Privacy is sort of a special case when it comes to the nature of the privacy threat vectors. Correlation attack is possible when using query enlargement techniques such as k-anonymity [8]. For instance, consider a user using a LBS which takes advantage of the user's mobile device's satellite positioning capabilities and requires user to send their location information to the service periodically. To protect user's identity, the LBS employs k-anonymity by sending user's data among with many other users' in the vicinity so that the user cannot be singled out of the data set. However, the anonymisation is done case-by-case. As a result of this, if the user would send the data consecutively and an adversary would be able to access these queries, the user would be included in every data set and could be easily identified with a simple interjection of the data sets. Figure 1 illustrates, how multiple queries may possible identify the user present in all data sets.



Credit: *Kiia Pitkänen, used with permission*

Figure 1: A user present in all data sets can be easily singled out with a simple interjection.

3 Techniques for countering the privacy threats

As there are possible privacy threats in LBS so is there techniques designed to address and mitigate their severity. [5] classify different threat countering techniques into four classes.

- **Query enlargement techniques** take the user's location and generalize it thus increasing the user's degree of anonymity.
- **Dummy-based techniques** generate fake locations and send them along with the user's actual location effectively hiding their whereabouts among the dummies.
- **Progressive retrieval techniques** retrieve candidate query results iteratively from the server so that the user's exact location will not be revealed.

- **Transformation-based techniques** use cryptography to hide users' data, but provides users deciphering keys to decrypt the data they need.

3.1 Query Enlargement techniques

Query enlargement techniques rely on the notion, that when sending location data to the server, if included is other identities from the given location, the user sending the data can not be easily identified from among other identities. The idea was first formulated by Latanya Sweeney[8] and in her paper it is applied to traditional databases. We assume that the database is made of rows of entries which consist of columns of attributes. The purpose of k-anonymity is to have a database with at least k occurrences of a certain attribute. When someone queries the database with these attributes they will get at least k entries which correspond to the given query, but the query can not be used to find a specific row. In this manner the data of a single user is safe but the dataset as a whole can still be used meaningfully for statistical study. To achieve a database which satisfies k-anonymity, one must first identify the so-called *Quasi-identifiers* which are usually non-unique attributes in the data set but can be used together with other quasi-identifiers to form unique identifiers. To make sure that the Quasi-identifiers cannot be used to identify single entries, they either have to be *suppressed*, made less distinguishable by replacing the value or part of it with a wild card notation, or *generalised* so that multiple values fall to a certain broader category. Figure 2 illustrates a small fictional anonymised database of medical records. The k in the database is 2 which means that there are at least two of each entries with the same quasi-identifier.

Race	Birth	Gender	ZIP	Problem
Black	1965	M	0214*	Short breath
Black	1965	M	0214*	Chest pain
Black	1965	F	0213*	Hypertension
Black	1965	F	0213*	Hypertension
Black	1964	F	0213*	Obesity
Black	1964	F	0213*	Chest pain
White	1964	M	0213*	Chest pain
White	1964	M	0213*	Obesity
White	1964	M	0213*	Short breath
White	1967	M	0213*	Chest pain
White	1967	M	0213*	Chest pain

Figure 2: Example of k-anonymity, where k=2 and the quasi-identifier consists of {Race, Birth, Gender, ZIP}. Original figure: [8]

In LBS context k-anonymity for the users' identities is achieved by generalising the location in the user's query into the so-called *Cloaking Range*. This is called *Spatial Cloaking*. When user sends data to the server, along is sent data from other users so that the exact identity and location of the user cannot be distinguished among others [3]. To anonymise the location, an approach called *Temporal Cloaking* can be used. In this approach the messages sent to the server are delayed until k mobile users have visited the location and thus the temporal dimension of the data can not be used to identify the user as the users in the data set are not necessarily present at the time of the query.

3.2 Dummy-based techniques

Dummy-based techniques can be seen as the other side of the query expansion techniques. The general idea is to send the location data in a set made out of faked dummy locations, instead of multiple actual user identities and location data [6]. The server then returns results for all of the locations, real and fakes as well, of which the user extracts the data associated with their real location. If this technique would be used on the case-by-case basis, the real location would be easily found as the fake location would seem to move erratically and irregularly whereas the real location would abide to real world constraints. To counter this Kido et al. provide two different ways to generate dummy locations [6]. The first one called *Moving in a Neighbourhood* (MN)

keeps the previous generation iteration in memory and uses it to generate the next iteration so that the fake locations are moved in relation to their own previous position decreasing the possibility for an adversary to find out the real location by observing the movement patterns. The other algorithm called *Moving in a Limited Neighbourhood* (MLN) requires that the user has access to other users' positional data. It works similarly to MN but if a new dummy would be created to an area which is highly congested by real users, it is generated again somewhere else instead. This is necessary because if the number of persons change in a certain area (moving etc.) there is a real possibility for the dummy location to stand out when compared to the movement of real locations [6].

The caveat of these proposed generation algorithms is, that they do not take the physical area in to account. This is important, as the size of the area correlates with the difficulty for the adversary to locate the user. Increasing the number of generated dummies does not explicitly guarantee the increase in the privacy region's size. [7] propose an enhanced version of the MN and LMN algorithms which distribute the dummies pseudo-randomly to a circular or a grid-based structure so that the size of the privacy region can be controlled. Grid-based distribution can also be used to reduce data transfer costs as the grid configuration can be sent in lieu of the raw data of each real and dummy location.

3.3 Progressive Retrieval Techniques

Progressive Retrieval Techniques are based on the idea, that the user retrieves results incrementally from the server until certain criteria, such as the sought after value is guaranteed to be in the data set or the user simply terminating the query and possibly settling for an approximation or a 'good enough' result. In this manner the user can get a query results without exposing their real location.

A prominent example of a progressive retrieval application is SpaceTwist [10]. SpaceTwist is used to find k nearest neighbours for the location q. First the user generates a fake location *q'* which is located somewhere in *dist(q, q')* radius. The nearest neighbours are sought incrementally in a growing circle starting from the fake location. The algorithm maintains two numeric variables γ and τ. $\gamma$is used to

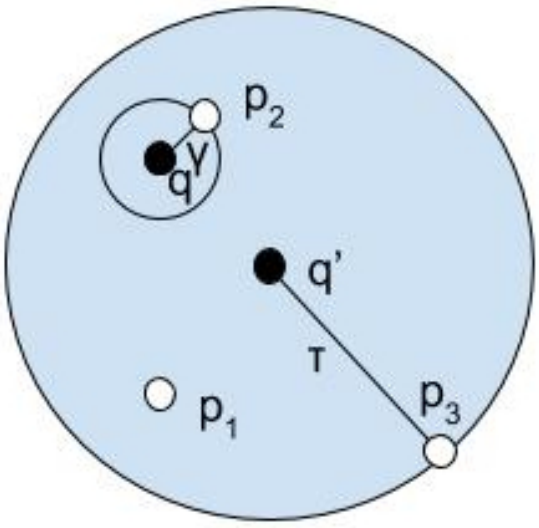


Figure 3: SpaceTwist search results after three iterations

4 Evaluation of the techniques

Sources

[1] Bellavista, P., Kupper, A., and Helal, S.: *Location-based services: Back to the future.* Pervasive Computing, IEEE, 7(2):85–89, April 2008, ISSN 1536-1268.

[2] Friedland, Gerald and Sommer, Robin: *Cybercasing the joint: On the privacy implications of geo-tagging.* In *Proceedings of the 5th USENIX Conference on Hot Topics in Security*, HotSec'10, pages 1–8, Berkeley, CA, USA, 2010. USENIX Association. `http://dl.acm.org/citation.cfm?id=1924931.1924933`.

[3] Gedik, Buğra and Liu, Ling: *Protecting location privacy with personalized k-anonymity: Architecture and algorithms.* IEEE Transactions on Mobile Computing, 7(1):1–18, January 2008, ISSN 1536-1233. `http://dx.doi.org/10.1109/TMC.2007.1062`.

[4] Ghinita, Gabriel, Kalnis, Panos, Khoshgozaran, Ali, Shahabi, Cyrus, and Tan, Kian Lee: *Private queries in location based services: Anonymizers are not necessary.* In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 121–132, New York, NY, USA, 2008. ACM, ISBN 978-1-60558-102-6. http://doi.acm.org/10.1145/1376616.1376631.

[5] Jensen, Christian S., Lu, Hua, and Yiu, Man Lung: *Privacy in Location-Based Applications: Research Issues and Emerging Trends*, chapter Location Privacy Techniques in Client-Server Architectures, pages 31–58. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, ISBN 978-3-642-03511-1. http://dx.doi.org/10.1007/978-3-642-03511-1_2.

[6] Kido, H., Yanagisawa, Y., and Satoh, T.: *An anonymous communication technique using dummies for location-based services.* In *Pervasive Services, 2005. ICPS '05. Proceedings. International Conference on*, pages 88–97, July 2005.

[7] Lu, Hua, Jensen, Christian S., and Yiu, Man Lung: *Pad: Privacy-area aware, dummy-based location privacy in mobile services.* In *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, MobiDE '08, pages 16–23, New York, NY, USA, 2008. ACM, ISBN 978-1-60558-221-4. http://doi.acm.org.libproxy.helsinki.fi/10.1145/1626536.1626540.

[8] Sweeney, Latanya: *K-anonymity: A model for protecting privacy.* Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 10(5):557–570, October 2002, ISSN 0218-4885. http://dx.doi.org/10.1142/S0218488502001648.

[9] Vicente, C.R., Freni, D., Bettini, C., and Jensen, Christian S.: *Location-related privacy in geo-social networks.* Internet Computing, IEEE, 15(3):20–27, May 2011, ISSN 1089-7801.

[10] Yiu, M. L., Jensen, C. S., Huang, X., and Lu, H.: *Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services.* In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 366–375, April 2008.