Privacy in Location Based Services

Vaibhav Sharma June 7, 2015

Introduction

The increase in the popularity of handheld devices such as Smart Phones, and Tablets with embedded global positioning capability has exponentially increased the applications that use Location Based Services. One of the main reasons for this surge is Geo-social networking application [8] such as Facebook, Yelp, FourSquare, Twitter and many others [7]. Apart from this single-user based application, LBSs are also used to gather traffic information and weather information (The Weather Channel's social weather application). Various crowd-sourcing LBS applications have played an important role in emergencies all over the world (Facebook's Safety Check feature was used recently in Nepal after devastating earthquake [9]).

There are several ways users interact with LBS:

1. Querying about their current position or point of interest
2. Creating location based content such as checking in at some location or using inbuilt location tagging photo applications
3. Companies use stored user location data to generated sanitized anonymous location based data that can be used internally to improve customer service or can be sold to third party applications

LBS applications are extremely useful but they come with their share of potential privacy risks, which depends on the ability of attackers to gain access to geo-location data. By using this data, attackers can infer about the daily routine of users, health status, alternative lifestyles, and political and religious affiliations [1].

This paper will briefly go over the inference based LBS attacks and will primarily cover current solutions such as location generalization using spatial transformation, various cryptographic methods and hybrid (spatial transformation and cryptographic) approaches. We will finally compare all the solutions based on performance and privacy. This paper will explore the conceptual idea of the solutions without delving into complex mathematics involved with each of the solutions.

Inference Attack on Location Based Services

The stripped down version of LBS architecture consist of a device with embedded global positioning capability, a trusted server (anonymizer) and an un-trusted LBS service provider [6]. The information flow from and to the user happens in a following manner using the K-anonymity approach:

1. User sends its current location and query to the anonymizer.
2. Anonymizer process the user request and creates a request bundle, which has the user request and location along with *K-1* requests and location, which could be from other users or plain dummy requests from the Cloaking Region.
3. Anonymizer sends these *K* requests or *K*-anonymizing spatial region (ASR) to the LBS provider.
4. LBS process all the *K* requests and sends the results back the anonymizer.
5. Anonymizer filters the actual query and sends results back to the user.

The K-anonymity makes sure that LBS provider does not have correct user location at any time but it does not stop attacker to infer the position of the user. Inference attack is based on the assumption that the communication between anonymizer and LBS is not secure and an attacker can get hold of the ASR. The next step is to identify the Point of Interests (POIs) for the user so as to characterize them. As per the research done by Colle and Kartridge [4], pair of home and work can almost work as a unique identifier for the individual. Once the POIs have been identified attacker can easily link the user to requests. After this attacker can predict the mobility behavior of the user and can create a profile of the user [2] [3].

Here we have developed a brief idea about the inference-based attacks. In the next sections we will look into various approaches to stop the attacker from identifying user. We will also discuss *K*-anonymity in detail.

Spatial Transformations

The *K*-anonymity approach discussed in the previous section inherently uses spatial transformation. Spatial Transformations can be divided into two categories – Two-tier Spatial Transformation and Three-Tier Spatial Transformation.

1) Two-Tier Spatial Transformation:

The basic two-tier spatial transformation does not involve anonymizer and user performs the work on anonymizer. User generates *K* fake queries and sends those queries to LBS along with the actual query. This approach is highly unsecure as user shares actual location data with the LBS [5]. Following are the two approaches that can be used to resolve this problem:

a) **Hilbert Curve Mapping** involves LBS storing POIs as Hilbert Value. The parameters (curve orientation, scale, etc.) to calculate Hilbert value are kept secret and represent the encryption key. User converts its current location to Hilbert Value and queries LBS for the closes value. Finally, user decrypts output from LBS using the inverse mapping function [5]. This approach is secure but this does not provide guarantee of result accuracy.

b) **The PROBE system** tries to solve both the problems by preventing the attacker to connect user with sensitive location. This is achieved by creating an obfuscated map [5]. The obfuscated map is created by dividing map using some criteria (For e.g. dividing city using blocks) so that no sensitive location should be in one obfuscated region. But practically this is not achievable. Though the PROBE system provides both performance and privacy; it comes with its own set of drawbacks. If a user query from two obfuscated regions and attacker has information about these queries, then attacker can approximate the location of the user by using variables such as velocity and time between the queries. Apart from that, if user is querying from the remote location and attacker knows user is the only person in that region then attacker can link that query to user and breaching user privacy. To solve this problem, system designer has to make sure the obfuscated region should have at least *K*

users, which is assured by Three-tier spatial transformation.

2) Three-Tier Spatial Transformation:

This type basically used the *K*-anonymity where instead of sending the actual location of the user, anonymizer actually sends the information about cloak region CR. To select cloak region, anonymizer uses the quad-tree data structure [5] The system will go specific quadrant represented by quad-tree where there are *K – 1* users along with the actual user and then select parent of that quadrant as cloak region. To construct the proper cloak region, user should continuously share location with the anonymizer. The privacy of the user is highly dependent on the size of *K*. Greater the size of K, more difficult is for attacker to link user with the location. As the value of *K* increases, there is a reduction in performance.

Analysis of Spatial Transformation Approaches:

Two-tier spatial transformation provides speed and accuracy but it is does not preserve the privacy of user. The PROBE system provides some level of anonymity, which is comparable to three-tier spatial transformation, but it has its own set of security issues. Three-tier spatial transformation provide better security but the performance of the whole architecture is of the order is O (*log* N), where N is the number of indexed users [5]. Even though the exact location of the user is not shared but cloaking region gives enough information to attacker to identify approximate location and stage attack. To overcome these drawbacks, Private Information Retrieval (PIR) can be used which is discussed in next section.

Cryptographic Approach

Cryptographic approaches are based on Private Information Retrieval (PIR) protocol, which allows user to retrieve an item from the database without revealing which information is retrieved [5]. Computation of PIR is dependent on Quadratic Residuosity Assumption (QRA), which states that it is computationally hard to find

the quadratic residues in modulo arithmetic of a large composite number $N = q_1 * q_2$ where q_1 and q_2 are large prime numbers [5]. Not going into the mathematics involved in calculating PIR, the underlying idea is if the database is an n-bit binary string of X and the client wants i^{th} bit of X then the client sends the encrypted request (q_i) to the server. The server responds with a value r (X, q_i). The assumption here is attacker cannot identify the value of i and user can easily decrypt the value of X_i [5].

The major challenge in PIR implementation is to convert POIs into indexes. This can be implemented using many know data-structures. One such data structure is *Kd-tree* [10], which divides the space into horizontally and vertically such that every partition has a square root of number of POIs. Each partition becomes the column in the database. What user receives are the boundaries of each partitions and user determines its partition based on its position. User receives all the POIs in the derived partition as per PIR protocol. As opposed to *K-anonymity* approach, PIR approach does not share any location details but instead POIs are retrieved based on the object index and thus provides complete privacy protection [5].

Though PIR provides complete user privacy protection, it comes with a large computation overhead, which is of the order of number of POIs. This makes this approach, though the benefits, less desirable.

Hybrid: Spatial Transformations and Cryptographic

We have previously identified that spatial transformation provide speed but less security and cryptographic approach provides great security but is less efficient. The two approaches are two ends of the line. The most desirable solution would be if we could get the speed of spatial transformation approach and security of cryptographic approach. This is provided by the hybrid approach.

The main problem with the cryptographic approach is that PIR protocol is requested over the whole dataset. The size of the requested dataset can be reduced if we use

K-anonymity approach and send cloaked region as the region that should be portioned. This is achieved by using the following process. User generates an encryption/decryption key pair and sends to location server LS the cloaked region CR, encryption public key and the encrypted user coordinates. LS partition the CR and makes sure there are a fixed set of POIs in each partitions. As opposed of cryptographic approach, user does not receive boundaries of the partition instead there is a cryptographic communication set up between user and LS to determine exact partition where user is located. LS sends back to user the boundary detail of the partition in which user is located. User then requests POIs of that specific partition using PIR protocol [5].

Thus using this hybrid approach, LS will never know the exact location of user but will only know the partition of CR where user is located and user will only know the details of the specific required partition. Also, by using hybrid approach the request PIR protocol is reduced limited partition size and thus reducing the complexity from size of data set to number of data points in specific partition.

Conclusion

In this paper, we have primarily kept ourselves limited to securing users interaction with LBS. But there are other facets of LBS where customer privacy can be compromised. Below are some of the examples:

1. Companies holding user location would want to perform statistical analysis or data mining activities to generate useful information so that they can provide better customer service.
2. Companies can sell sanitized user information to third parties or for research purposes.

In both these cases it is important that no user information is revealed or can be inferred. There are many different existing solutions for each of the problem, which we have not discussed in this paper.

In this paper we discussed about inference attacks in Location Based Services and

how the attacker can stage them. We discussed in detail about the spatial transformation approaches. Both the approaches, two-tier and three-tier spatial transformation, are efficient from performance viewpoint but they have inherent weakness in protecting against inference attacks. To provide better protection from inference attacks, we then discussed cryptographic approach using Private Information Retrieval (PIR) protocol. This approach provides full protection against inference attack but is computationally heavy; thus making it less desirable.

Finally, we discussed the hybrid approach – Spatial Transformation with Cryptographic Approach – which ensured that no information about a user's location is disclosed to location servers; thus protecting users' identity. At the same time by reducing size of PIR protocol request, there is a huge reduction in expensive cryptographic operations and thus improving overall performance.

References

1. Wenyan Zhang; Ximing Cui; Dengfeng Li; Debao Yuan; Mengru Wang, "The location privacy protection research in location-based service," Geoinformatics, 2010 18th International Conference on, vol., no., pp.1, 4, 18-20 June 2010
   URL: http://ieeexplore.ieee.org.ezproxy.depaul.edu/stamp/stamp.jsp?tp=&arnumber=5568118&isnumber=5567473
2. Nunez del Prado Cortez, M.; Frignal, J., "Geo-Location Inference Attacks: From Modeling to Privacy Risk Assessment (Short Paper)," Dependable Computing Conference (EDCC), 2014 Tenth European, vol., no., pp.222, 225, 13-16 May 2014
   URL: http://ieeexplore.ieee.org.ezproxy.depaul.edu/stamp/stamp.jsp?tp=&arnumber=6821108&isnumber=6821069
3. Minami, K., "Preventing denial-of-request inference attacks in location-sharing services," Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on , vol., no., pp.50,55, 6-8 Jan. 2014
   URL: http://ieeexplore.ieee.org.ezproxy.depaul.edu/stamp/stamp.jsp?tp=&arnumber=6799057&isnumber=6799045
4. P. Golle and K. Partridge, "On the anonymity of home/work location pairs," in Proceedings of the International Conference on Pervasive Computing, May 2009, pp. 390–397.
5. Ghinita, Gabriel, "Privacy For Location Based Services" Morgan and Claypool Publishers April 2013, Vol. 4.
6. Hao Zhou; Yingjie Wu; Sisi Zhong; Zhao Luo; Xiaodong Wang, "Preventing Location-based Inference Attack in Location Based Services," Industrial Control and Electronics Engineering (ICICEE), 2012 International Conference on , vol., no., pp.1587,1590, 23-25 Aug. 2012
   URL: http://ieeexplore.ieee.org.ezproxy.depaul.edu/stamp/stamp.jsp?tp=&arnumber=6322708&isnumber=6322296
7. http://techpp.com/2011/01/17/top-10-location-based-service-providers/
8. http://en.wikipedia.org/wiki/Geosocial_networking
9. https://www.facebook.com/about/safetycheck/
10. http://en.wikipedia.org/wiki/K-d_tree