Location based Queries for Content protecting and Privacy preserving

Ms. Sonali B. Gosavi[1], Dr.Shyamrao.V.Gumaste[2]

[1]M.E. Second Year Student, Department of Computer Engineering,
SPCOE, Otur, Pune, India
[2]Professor, Department of Computer Engineering,
SPCOE, Otur, Pune, India

Abstract: The location server desires to have major control over its data, since data is its big asset. This paper propose a major enhancement by introducing a two stage approach, where the first step is based on Oblivious Transfer stage and the second step is based on Private Information Retrieval stage, to achieve a secure solution for both parties. This report presents a solution to one of the location-based query problems and problem is defined as: (i) a user wants to query a database of location data, known as Points Of Interest (POIs), and does not want to reveal his/her location to the location server due to privacy concerns; (ii) the owner of the location data, that is, the location server, does not want to just distribute its data to all users due to the content preserving concern since data is its asset. The solution present in this report is efficient and practical in many scenarios. Proposed solution can be implemented on a desktop pcs, laptops and mobile phones to assess the efficiency of implemented protocol. Proposed method also introduces a security model and analyzes the security in the context of implemented protocol. Finally this solution highlights a security weakness of previous systems work and presents a novel solution to overcome the disadvantages of previous work.

Keywords: Location based query, Points of Interest, private query, private information retrieval, oblivious transfer.

1. Introduction

The GPS location searching service provided by mobile network on mobile devices such as, mobile phones, GPS devices, pocket PCs are used to find information, entertainment and utility service as per the peoples demand. LBS can offer many services to the users based on the geographical position of their mobile device. The services provided by LBS are typically based on a point of interest database. By retrieving the Points of Interest (POI's) from the database server, the user can get answers to various location based queries, which include but are not limited to discovering the nearest ATM machine, gas station, hospital, or police station. In recent years there has been a dramatic increase in the number of mobile devices querying location servers for information about POI's. Among many challenging barriers to the wide deployment of such application, privacy assurance is a major issue. For instance, users may feel reluctant to disclose their locations to the LBS, because it may be possible for a location server to learn who is making a certain query by linking these locations with a residential phone book database, since users are likely to perform many queries from home.

The Location Server (LS), which offers some LBS, spends its resources to compile information about various interesting POI's. Hence, it is expected that the LS would not disclose any information without fees. Therefore the LBS have to ensure that LS's data is not accessed by any unauthorized user. During the process of transmission the users should not be allowed to discover any information for which they have not paid. It is thus crucial that solutions be devised that address the privacy of the users issuing queries, but also prevent users from accessing content to which they do not have authorization.

2. Related work

Solution proposed by Beresford [2] is having some problems, in which the privacy of the user is maintained by constantly changing the user's name or pseudonym within some mix-zone. It can be shown that, due to the nature of the data being exchanged between the user and the server, the frequent changing of the user's name provides little protection for the user's privacy. A more recent investigation of the mix-zone approach has been applied to road networks [11]. They investigated the required number of users to satisfy the unlinkability property when there are repeated queries over an interval. This requires careful control of how many users are contained within the mix-zone, which is difficult to achieve in practice.

The concept of k-anonymity was introduced as a method for preserving privacy when releasing sensitive records [6]. This is achieved by generalization and suppression algorithms to ensure that a record could not be distinguished from (k − 1) other records. The solutions for LBS use a trusted anonymiser to provide anonymity for the location data, such that the location data of a user cannot be distinguished from (k − 1) other users.

The anonymiser approach allows the users to set their level of privacy based on the value of k [6], [10]. This means that, given the overhead of the anonymiser, a small value of k could be used to increase the efficiency. Conversely, a large value of k could be chosen to improve the privacy, if the users felt that their position data could be used maliciously. Choosing a value for k, however, seems unnatural. There have been efforts to make the process less artificial by adding the concept of feeling-based privacy [13], [14]. Instead of specifying a k, they propose that the user specifies a cloaking region that they feel will protect their privacy, and the system sets the number of cells for the region based on the popularity of the area. The popularity is computed by using historical footprint database that the server collected.

A different technique proposes to distort and confuse the location based data, which include path and position confusion. Path confusion was presented by Hoh and Gruteser [7]. The basic idea is to add uncertainty to the location data of the users at the points the paths of the users cross, making it hard to trace users based on raw location data that was k-anonymised. Position confusion has also been proposed as an approach to provide privacy [4], [5]. The idea is for the trusted anonymiser to group the users according to a cloaking region (CR), thus making it harder for the LS to identify an individual. A common problem with general CR techniques is that there may exist some semantic information about the geography of a location that gives away the user's location. For example, it would not make sense for a user to be on the water without some kind of boat. Also, different people may find certain places sensitive. Damiani et al. have presented a framework that consists of a obfuscation engine that takes a users profile, which contains places that the user deems sensitive, and outputs obfuscated locations based on aggregating algorithms [7].

Other approach for avoiding the use of a trusted anonymiser is to use 'dummy' locations [2], [3]. The basic idea is to confuse the location of the user by sending many random other locations to the server, such that the server cannot distinguish the actual location from the fake locations. This incurs both processing and communication overhead for the user device. The user has to randomly choose a set of fake locations as well as transmitting them over a network, wasting bandwidth. One refer the interested reader to Krumm , for a more detailed survey in this area.

Most of the previously discussed issues are solved with the introduction of a private information retrieval (PIR) location scheme [13]. The basic idea is to employ PIR to enable the user to query the location database without compromising the privacy of the query. Generally speaking, PIR schemes allow a user to retrieve data (bit or block) from a database, without disclosing the index of the data to be retrieved to the database server [6]. Ghinita et al. used a variant of PIR which is based on the quadratic residuosity problem [14]. Basically the quadratic residuosity problem states that is computationally hard to determine whether a number is a quadratic residue of some composite modulus n (x2 = q mid n), where the factorization of n is unknown.

This approach was elaborated to provide database protection [6], [8]. This protocol consists of two stages. In the first stage, the user and server use homomorphic encryption to allow the user to privately determine whether his/her location is contained within a cell, without disclosing his/her coordinates to the server. In the second stage, PIR is used to retrieve the data contained within the appropriate cell. The homomorphic encryption scheme used to privately compare two integers is the Paillier encryption scheme [11].The Paillier encryption scheme is known to be additively homomorphic and multiplicatively-by-a-constant homomorphic. This means that one can add or scale numbers even when all numbers are encrypted. Both features are used to determine the sign (most significant bit) of (a − b), and hence the user is able to determine the cell in which he/she is located, without disclosing his/her location.

3. Motivation

This paper proposes a superior protocol for location based queries that have major performance improvements with respect to the approach by Ghinita at el. [5 and [6]. Like such protocol, the protocol is organized according to two stages. In the first stage, the user privately determines his/her location within a public grid, using oblivious transfer. This data contains both the ID and associated symmetric key for the block of data in the private grid. In the second stage, the user executes a communicational efficient PIR [11], to retrieve the appropriate block in the private grid. This block is decrypted using the symmetric key obtained in the previous stage. This protocol thus provides protection for both the user and the server. The user is protected because the server is unable to determine his/her location. Similarly, the server's data is protected since a malicious user can only decrypt the block of data obtained by PIR with the encryption key acquired in the previous stage. In other words, users cannot gain any more data than what they have paid for. One remark that this report is an enhancement of a previous work [6]. In particular, the following contributions are made. 1. Redesigned the key structure

2. Added a formal security model

3. Implemented the solution on both a mobile device and desktop machine.

4. System Design Overview

A. System Model:

The system model consists of three types of entities Fig. 1: the set of users who wish to access location data U, a mobile service provider SP, and a location server LS.

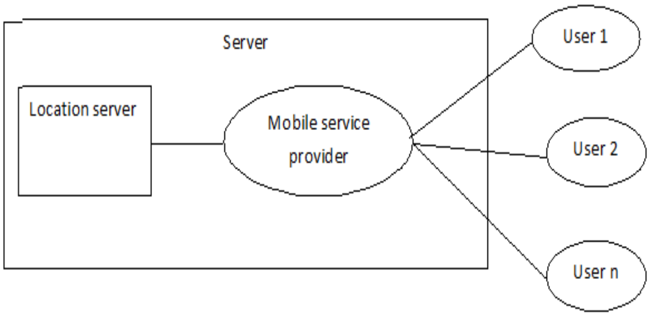

Figure 1: System Model

From the point of view of a user, the SP and LS will compose a server, which will serve both functions. The user does not need to be concerned with the specifics of the communication.

The users in model use some location-based service provided by the location server LS. For example, what is the nearest ATM or restaurant? The purpose of the mobile service provider SP is to establish and maintain the communication between the location server and the user. The location server LS owns a set of POI records ri for 1 ≤ ri ≤ ρ. Each record describes a POI, giving GPS coordinates to its location (xgps, ygps), and a description or name about what is at the location.

One reasonably assume that the mobile service provider SP is a passive entity and is not allowed to collude with the LS. One makes this assumption because the SP can determine the whereabouts of a mobile device, which, if allowed to collude with the LS, completely subverts any method for privacy. There is simply no technological method for preventing this attack. As a consequence of this assumption, the user is able to either use GPS (Global Positioning System) or the mobile service provider to acquire his/her coordinates.

Since authors are assuming that the mobile service provider SP is trusted to maintain the connection, one considers only two possible adversaries. One for each communication direction. One considers the case in which the user is the adversary and tries to obtain more than he/she is allowed. Next one considers the case in which the location servers LS is the adversary, and tries to uniquely associate a user with a grid coordinate.

B. Protocol Description:

Here describe working of protocol. First giving a protocol summary to contextualize the proposed solution and then describe the solution's protocol in more detail. The ultimate goal of proposed protocol is to obtain a set (block) of POI records from the LS, which are close to the user's position, without compromising the privacy of the user or the data stored at the server. One achieve this by applying a two stage Approach shown in Fig. 2. The first stage is based on a two-dimensional oblivious transfer [6] and the second stage is based on a communicationally efficient PIR [11]. The oblivious transfer based protocol is used by the user to obtain the cell ID, where the user is located, and the corresponding symmetric key. The knowledge of the cell ID and the symmetric key is then used in the PIR based protocol to obtain and decrypt the Location data.

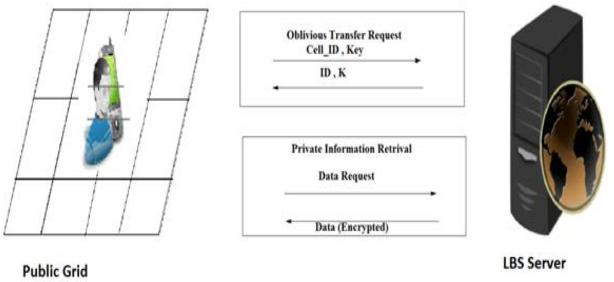

Figure 2: Privately determine user for LBS communication

The user determines his/her location within a publicly generated grid P by using his/her GPS coordinates and forms an oblivious transfer query. The minimum dimensions of the public grid are defined by the server and are made available to all users of the system. This public grid superimposes over the privately partitioned grid generated by the location servers as POI records, such that for each cell Q*i,j* in the servers as partition there is at least one P*i,j* cell from the public grid Since PIR does not require that a user is constrained to obtain only one bit/block, the location server needs to implement some protection for its records. This is achieved by encrypting each record in the POI database with a key using a symmetric key algorithm, where the key for encryption is the same key used for decryption. This key is augmented with the cell info data retrieved by the oblivious transfer query. Hence, even if the user uses PIR to obtain more than one record, the data will be meaningless resulting in improved security for the server's database.

Conclusion

In this paper authors have presented a location based query solution that employs two protocols that enables a user to privately determine and acquire location data. The first step is for a user to privately determine his/her location using oblivious transfer on a public grid. The second step involves a private information retrieval interaction that retrieves the record with high communication efficiency. Authors analyzed the performance of protocol and found it to be both computationally and communicationally more efficient than the solution by Ghinita et al., which is the most recent solution. Authors implemented a software prototype using a desktop machine and a mobile device. The software prototype demonstrates that protocol is within practical limits.

Future work will involve testing the protocol on many different mobile devices. The mobile result that authors provide may be different than other mobile devices and software environments. Also there is need to reduce the overhead of the primality test used in the private information retrieval based protocol.

Acknowledgement

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We also thank the college authority for providing required infrastructure and support.

References

[1] M. Bellare and S. Micali, "Non-interactive oblivious transfer and applications", In Proc. CRYPTO, 1990, pp. 547ˆa557.

[2] A. Beresford and F. Stajano , "Location privacy in pervasive computing", IEEE Pervasive Comput., vol. 2, no. 1, pp. 46ˆa55, Jan.ˆaMar.2003.

[3] C. Bettini, X. Wang, and S. Jajodia, "Protecting privacy against location-based personal identification", in Proc. 2nd VDLB Int. Conf. SDM,W. Jonker and M. Petkovic, Eds., Trondheim, Norway, 2005, pp. 185ˆa199, LNCS 3674.

[4] X. Chen and J. Pang, "Measuring query privacy in location-based services", in Proc. 2nd ACM CODASPY, San Antonio, TX, USA, 2012, pp. 49-60.

[5] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private queries in location based services: Anonymizers are not necessary", in Proc. ACM SIGMOD, Vancouver, BC, Canada, 2008, pp. 121-132.

[6] M. Naor and B. Pinkas,"Oblivious transfer with adaptive queries", in Proc. CRYPTO, vol. 1666, Santa Barbara, CA, USA, 1999, pp. 791-791.

[7] B. Hoh and M. Gruteser,"Protecting location privacy through path confusion", in Proc. 1st Int. Conf. SecureComm, 2005, pp. 194-205.

[8] G. Ghinita, C. R. Vicente, N. Shang, and E. Bertino, "Privacy preserving matching of spatial datasets with

protection against background knowledge", in Proc. 18th SIGSPATIAL Int. Conf. GIS, 2010, pp. 3-12.

[9] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries", IEEE Trans. Knowl. Data Eng., vol. 19, no. 12, pp. 1719ˆa1733,Dec. 2007.

[10] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval", Proc. Scientific and Statistical Database Management(SSDBM), 2007.

[11] G.R. Hjaltason and H. Samet, "Distance Browsing in Spatial Databases ", J. ACM, vol. 45, no. 6, pp. 965ˆa981, 1998.

[12] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms", IEEE Trans. Inform. Theory, vol. 31, no. 4, pp. 469ˆa472, Jul. 1985.

[13] R. Paulet, M. Golam Kaosar, X. Yi, and E. Bertino, "Privacy preserving and content-protecting location based queries", in Proc. ICDE, Washington, DC, USA, 2012, pp. 44-53.

[14] V. Shoup, (2011, Jul. 7). Number theory library [Online], Available: http://www.shoup.net/ntl/.