Location And Query Privacy In K Nearest Neighbour Queries

Kaluvala Swapna & Kauda BhanuPrasad

[1]PG Scholar, Department of CSE, Vaagdevi College of Engineering, Bollikunta Warangal, Telangana, Mail id: kaluvalaswapna@gmail.com.

[2]Assistant Professor, Department of CSE, Vaagdevi College of Engineering, Bollikunta Warangal, Telangana, Mail id: banu.kbp@gmail.com

ABSTRACT

In mobile communication, spatial queries pose a serious threat to user location privacy because the location of a query may reveal sensitive information about the mobile user. In this paper, we study k nearest neighbor (kNN) queries where the mobile user queries the location-based service (LBS) provider about k nearest points of interest (POIs) on the basis of his current location. We propose a solution for the mobile user to preserve his location privacy in kNN queries. The proposed solution is built on the Paillier public-key cryptosystem and can provide both location privacy and data privacy. In particular, our solution allows the mobile user to retrieve one type of POIs, for example, k nearest car parks, without revealing to the LBS provider what type of points is retrieved. For a cloaking region with n × n cells and m types of points, the total communication complexity for the mobile user to retrieve a type of k nearest POIs is $O(n+m)$ while the computation complexities of the mobile user and the LBS provider are $O(n + m)$ and $O(n 2m)$, respectively. Compared with existing solutions for kNN queries with location privacy, our solutions are more efficient. Experiments have shown that our solutions are practical for kNN queries.

I. INTRODUCTION

The integration of positioning capabilities (eg GPS) into mobile devices facilitates the emergence of location-based services (LBS), which is considered the next "deadly application" in the wireless data market. LBS allows customers to query a service provider (such as Google or Bing Maps) ubiquitously to retrieve detailed information about nearby points of interest (POIs) (restaurants, hospitals, etc.). The LBS provider processes spatial queries based on the location of the mobile user. Location information collected from mobile users, knowingly and unknowingly, can reveal much more than the latitude and longitude of a user. Knowing where a mobile user is can mean knowing what he's doing: attending a church service or support meeting, visiting a doctor's office, buying an engagement ring, conducting non-business activities or spending an evening at the corner of the office. bar. He could reveal that he is interviewing for a new job or "out" as a participant in a gun rally or peace protest. It may mean knowing who he / she spends time with and how often. When location data is aggregated, it can reveal its habits and routines - and when it deviates from them. A 2010 survey for Microsoft in the United Kingdom, Germany, Japan, the United States and Canada found that 94% of consumers who used location-based services thought they were useful, but the same survey found that 52% privacy1. In this article, we investigate the k nearest neighbors (kNN) queries where the mobile user queries the LBS provider on k closest POIs. In general, the mobile user must submit his location to the LBS provider who then discovers and returns to the user the nearest k POIs by comparing the distances between the mobile user's location and the nearby points of interest. . This reveals the location of the mobile user to the LBS provider. There have been many techniques that can provide some degree of location intimacy. These techniques mainly include: • Access control to information. LBS requests based on access control, mixing zone, and kanonymity require the service provider or middleware that retains all user locations. They are vulnerable to the misbehavior of the third party. They offer little protection when the service provider / middleware is the property of an unreliable party. Private data has been inadvertently disclosed on the Internet in the past. k-anonymity is initially used to protect the confidentiality of the identity. It is generally inadequate for the protections of the confidentiality of places, where the notion of distance between places is important (unlike the distances between identities). The effect of anonymity-based LBS queries k strongly depends on the distribution and density of mobile users, which, however, are beyond the control of the location confidentiality technique. Factual location-based LBS requests require the mobile user to randomly select a set of false locations, send the false locations to the LBS, and receive false reports from the LBS over the mobile network. This entails computing and communication costs in mobile devices. For efficiency reasons, the mobile user may choose fewer false locations, but the LBS provider may restrict the user to a small subspace of the total domain, resulting in low privacy. LBS queries based on the transformation of geographic data are likely to access model attacks [24] because the same query always returns the same coded results. For example, the LBS can observe the frequencies of the returned ciphertext. Having knowledge of the context of the database, it can match the POI in the most popular raw text with the most frequently returned ciphertext and, thus, unravel the information on the query. PIR-based LBS requests offer strong cryptographic safeguards, but are often expensive in terms of computation and communication. To improve efficiency, reliable hardware was used to perform PIR requests for LBS [17]. This technique is based on a hardware-assisted PIR [23], which assumes that a trusted third party (TTP) initializes the system by setting the secret key and permutation of the database. Like LBS requests based on access control, mixing zone and anonymity k, this technique is vulnerable to bad behavior third parties. It is a challenge to provide practical solutions for kNN queries with location confidentiality based on PIR. In this paper, we build solutions for kNN queries based on PIR with the Paillier public key cryptosystem [15]. We have three main contributions: • LBS requests based on PIR [6], [7], [18], [19] usually require two steps. In the first step, the mobile user retrieves the index of its location from the LBS provider. In the second step, the mobile user retrieves the POIs based on the index of the LBS provider. To simplify the process, we give a solution to kNN queries that only require one that is, the mobile user sends his (encrypted) location to the LBS provider and receives the nearest (encrypted) POIs from the LBS provider. • Current PIR-based LBS requests only allow the mobile user to find the nearest POIs, regardless of POI type. For the first time, we take into account the type of POIs in kNN requests and give the mobile user a solution to find the nearest PIO of the same type without revealing to the LBS provider the type of POI that interests him. , our solution allows the nomadic user to find the car parks closest to the LBS provider. • Current PIR-based LBS requests must all fix a hiding region based on which the LBS provider generates the mobile user's query responses. If the hiding region is large, the LBS queries are inefficient. If the hiding region is small, LBS requests have low privacy.

II. RELATED WORKS

The main current techniques to maintain the confidentiality of the localization for LBS are as follows. • Information access control [12], [28]: User locations are sent to the LBS provider as usual. This technique relies on the LBS provider to restrict access to stored location data through rule-based policies. It supports three types of location-based queries: 1) user location queries (query the location of one or more specific users, identified by their unique identifiers); 2) enumeration queries (querying user lists at specific locations, expressed either in terms of geographic or symbolic attributes); 3) Asynchronous requests (querying "event" information, such as when users enter or leave specific areas). This technique requires the LBS provider to maintain all user locations. He is vulnerable to the misbehavior of the LBS provider. • Mixing zone [2]: Trusted middleware between mobile users and the LBS provider. Before forwarding requests based on the location of users to the LBS, the middleware anonymizes their locations by pseudonyms. The basic idea is that when a user enters a mixing zone, the middleware assigns him a pseudonym, through which the user queries LBS. The communication between the user and the LBS goes through the middleware and the nickname changes each time the user enters the mixing zone. Recently, the mixed zone has been applied to road networks [16]. This technique requires the middleware to anonymize user locations. He is vulnerable to mishandled middleware. • k-anonymity [22]: This technique ensures that a record can not be distinguished from k-1 other records. Instead of sending the exact location of a single user to the LBS, kanonymy-based schemes collect user locations k and send a corresponding (minimal) bounding region to the LBS as a query parameter. The collection of different mobile user locations is performed either by a trusted third party [11], [1] between the users and the LBS, or by a 641 peer-to-peer collaboration [5] among the users. Because kanonymity is achieved, an opponent can only identify the user of a location with a probability not exceeding 1 / k. This technique relies on the third party or a peer user to collect different mobile user locations. It is vulnerable to bad behavior by the third party or the peer user. • "dummy" locations [10], [21]: The basic idea is that when the mobile user queries the LBS, it sends many other random locations with its location to the LBS provider to confuse its location so that the server can not distinguish the actual location from the wrong locations. Different from anonymity-based schemes k, this approach includes falsified or fixed locations, rather than those of other mobile users, as parameters of the requests sent to the LBS provider. False dummy locations are randomly generated, and fixed locations are selected from special locations such as road intersections. In any case, user locations are hidden from the service provider. Although this technique is not based on any third party, the LBS provider can restrict the user in a small subspace of the total domain, resulting in low confidentiality. • Private Information Retrieval (PIR) [14]: This technique allows a user to retrieve a record from a database server without revealing the record they are retrieving. PIR based protocols [6], [7], [18], [19] are proposed for POI queries and consist of two steps. In the first step, the user determines in private the index of his location through the service provider without disclosing his coordinates. In the second step, the user executes a PIR protocol with the service provider to retrieve the POIs corresponding to the index. The difference between Ghinita et al. [6], [7] and Paulet et al. [18], [19] The protocols based on PIR are in the first stage, where Ghinita et al. approach is based on homomorphic encryption [15] whereas the technique of Paulet et al. is based on the unconscious transfer [13]. In addition, trusted hardware was used to perform PIR requests for LBS [17]. Their technique is built on hardware-assisted PIR [23], which relies on a trusted third party (TTP) to define thesecret and the permutation of the database. Like LBS requests based on access control, mixing zone and anonymity k, this technique is vulnerable to third party misbehavior. • Transformation of geographic data [25], [8], [26]: This technique involves three parts: 1) A data owner who has a D database and wishes to outsource D to a server (ex.) That does not can not be entirely reliable. 2) A user who wants to access and query the database D. 3) An honest but potentially curious server in D tuples and / or user requests. A server may be curious either because it is simply curious or because it has been compromised to become curious in the name of a third party without its explicit knowledge. In this setting, the owner of the data is different from the LBS. The owner transforms the database (using a coding method) before sending it to the LBS. To allowThe user who owns the secret transformation keys issues a coded request to the LBS. Both the database and the queries are unreadable by LBS and, as a result, the privacy of the site is protected. The goal is to provide the LBS with coded data search capabilities. Wong et al. [25] provide a secure point transformation that preserves the relative distances of all POIs in the database to any request point. Another solution [26] uses the order-preserving encryption [3], [4], given only the encryption of the location point E (q) and the encryption of the database E (D), the server can returning an encrypted relevant partition E (G) from E (D), so that E (G) is sure to contain the response to the request NN. These techniques allow an approximate NN search directly on the transformed points. They are prone to accessing model attacks [24] because the same query always returns the same coded results

III PRIVATE K NEAREST NEIGHBOR

So far, we have considered approximate consultations, for which there is an inherent compromise between the amount of POI disclosed and the accuracy of the results. The most popular POI increases the probability that one of them is the real NN, or the distance between the user and the approximate NN is close to the distance of the real NN. However, an ideal private query technique should return a single point of interest that is the exact NN. In this case, an optimal result is obtained both from the point of view of the user interrogator, which receives its exact NN, and the perspective of the database as a single data point is made known by the application. In this section, we will present a method that achieves this optimal result. Previous work in [5] introduced a method (described in Section 2) for private private NN queries using Voronoi and PIR tiling. The idea is to use a regular 2D grid and create a network diagram in which each cell of the 2D grid is assigned to all data points whose Voronoi cells intersect the grid cell. At the time of the request, the client makes a PIR request for the cell in the grid surrounding its location. Although the method ensures that the exact NN is part of the result received by the user, the compartmentalization of Voronoi cells in the grid cells implies an inherent loss of precision due to the inability to have a sufficiently fine grid for a single cell of Voronoi is hash in every cell of the network. In fact, the experimental results of [5] with a set of real data show that the average number of hash data points in a cell of the grid (and therefore disclosed in a single application) is 15. This number is much to be optimal a diluent of the 2D grid cell could potentially reduce this number, but the results in an overload of computing and the most important communication since PIR computational complexity is linear with respect to the number of cells in the grid . In addition, if the data are distorted, the use of a finer grid does not necessarily result in a decrease in the disclosed POI. There is another important drawback of the Voronoi cube solution cells: in addition to the total number of cells in the grid, the maximum number of similar measures Voronoi 12A was used for R trees [28]. Although MBRs are used in the evaluation of benefits, the resulting partition is not pruned in the MBR due to the requirement that the index cover the entire data space. Geoinformatic cells cut into a grid cell are also a factor with a linear influence on PIR complexity, this time in terms of computational cost and communication. More precisely, this collision factor dictates the "depth" of the PIR matrix, in other words, the PIR protocol is executed separately for each collision point. Keep in mind that even if there is only one cell with a large number of collisions, this also affects all other cells because the server should not be able to know which cell is being recovered based on the depth of that cell. . all cells must be padded to the depth of the grid cell with most collisions. To aggravate the problem, this maximum depth parameter depends on the data set and an upper limit can not be determined: in the worst case, the depth is linear in relation to the size of the database. All the limitations mentioned above can be followed up to a factor: only in the PIR framework, the client and the server can not afford any form of interactive filtering, preserving the confidentiality of the results without disclosure. Specifically, all the client is able to do is recover a fraction of the database privately, but the recovery is made based on the index of the data, and not directly based on the spatial information. The only time spatial information is taken into account is in the compartmentalization phase, but, as we have indicated, the accuracy of compartmentalization may be low. In addition, the compartmentalization is done independently of the request. As a result, it may not be possible to find a compartmentalization that optimally addresses all requests. We propose a different approach, in which query processing uses the techniques developed in Section 4.1 for the interactive evaluation of arithmetic conditions through homomorphic encryption. In particular, we will show that the user can privately identify the cell index of bucketization

IV Conclusions

This article proposes hybrid techniques for approximate and accurate private NN queries that provide protection to users and the service provider. Our solutions rely on cryptographic protocols for the private evaluation of a point-in-rectangle and polygonal point-to-convex enclosure. Hybrid techniques allow much lower disclosure of POIs than their CR only counterparts and PIR only. In fact, the exact hybrid NN method is optimal with respect to POI disclosure. The proposed techniques are also effective in practice, and outperform pure PIR methods in most cases, with the sole exception of large CR queries for the optimal exact NN solution. In future work, we plan to expand our work to support the private assessment of more advanced spatial conditions and more complex query types, such as kNN queries for k> 1 and skyline queries. Geoinformatics

REFERENCES

1. Gruteser M, Grunwald D (2003) Anonymous use of location-based services through spatial

and temporal camouflage. In: Proc. from USENIX MobiSys

2. Gedik B, Liu L (2005) Location Confidentiality in Mobile Systems: A Customized Privacy Model. In: Proc. of the ICDCS, pp 620-629

3. MF Mokbel, Chow CY, Aref WG (2006) The new Casper: query processing for location services without compromising privacy. In: Proc. from VLDB, pp 763-774

4. Kalnis P, Ghinita G, Mouratidis K, Papadias D (2007) Preservation of location-based identity inference in anonymous spatial queries. IEEE TKDE 19 (12): 1719-1733

5. Ghinita G, Kalnis P, Khoshgozaran, Shahabi C, Tan KL (2008) Private requests in location-based services: anonymizers are not needed. In: SIGMOD, pp 121-132

6. Kido H, Yanagisawa Y, Satoh T (2005) An anonymous communication technique using dummies for location-based services. In: International Conference on Invasive Services (ICPS), pp 88-97

7. Yiu ML, Jensen C, Huang X, Lu H (2008) SpaceTwist: Managing trade-offs between location privacy, query performance, and query accuracy in mobile services. In: International Conference on Data Engineering (CESI), pp. 366-375

8. Cheng R, Zhang Y, E Bertino, Prahbakar S (2006) Preserving the confidentiality of users' location in mobile data management infrastructures. In: Privacy Enhancement Technologies (PET), pp 393-412

9. Chow CY, Mokbel MF (2007) Enabling private continuous queries for revealed user locations. In: SSTD, pp 258-275

10. Gruteser M, Liu X (2004) Protection of confidentiality in continuous localization applications. IEEE Secur Priv 2: 28-34

11. Damiani M, Bertino E, Silvestri C (2008) PROBE: an obfuscation system for the protection of sensitive location information in LBS. Technical Report 2001-145, CERIAS

12. Khoshgozaran A, Shahabi C (2007) Blind evaluation of the requests of the nearest neighbors by using the transformation of the space to preserve the confidentiality of the places. In: SSTD, pp 239-257

13. Chor B, Goldreich O, Kushilevitz E, Sudan M (1995) Retrieval of private information. In: IEEE Symposium on the Foundations of Informatics, pp 41-50

14. Kushilevitz E, Ostrovsky R (1997) Replication is not necessary: single database, computer-based information retrieval. In: FOCS, pp 364-373

15. Flath DE (1998) Introduction to number theory. Wiley, New York

16. Atallah MJ, Du W (2001) Secure multipartite computational geometry. In: WADS '01: Proceedings of the 7th International Workshop on Algorithms and Data Structures, pp 165-179

17. Luo Y, Huang L, Zhong H (2007) Securing the problem of including two-party point circles. J Comput Sci Technol 22 (1): 88-91

18. O Goldreich, Micali S, Wigderson A (1987) How to play any mental game. In: Proceedings of the ACM Colloquium on the Theory of Computing (STOC), pp. 218-229

19. Fischlin M (2001) A profitable multiplication payment comparison method for millionaires. In: CT-RSA 2001: Proceedings of the 2001 conference on the themes of cryptology, pp. 457-472

20. Blake IF, Kolesnikov V (2004) Strong conditional unconscious transfer and computation over intervals. In: Advances in Cryptology-ASIACRYPT 2004, pp 515-529

21. Lin HY, Tzeng WG (2005) An effective solution to the problem of millionaires based on homomorphic encryption. In: Intl. conference on Applied Cryptography and Network Security, pp 456-466

22. Yao AC (1982) Protocols for secure calculations. In: SFCS '82: Proceedings of the 23rd Annual Symposium on the Foundations of the informatique, pp 160-164