A Novel Study for Spatio-Temporal Query Processing using Privacy Preservation

Dileep Kumar, The University of Suwon, Hwaseong-Si Korea

***Abstract*-The prevailing infrastructure of ubiquitous computing paradigm on the one hand making significant development for integrating technology in the daily life but on the other hand raising concerns for privacy and confidentiality. As Location based services (LBS) equip users to query information specific to a location with respect to temporal and spatial factors thus LBS put under extreme criticism when it comes to location privacy and user confidentiality. Here in this paper we are addressing the significance of our pro- posed scheme, a query processing architecture for privacy preservation in LBS, by providing flexible and efficient LBS model to ensure accurate and qualitative result set by employing some indexing scheme at location anonymizer as well as by Identifying possible adversary attacks to breach user privacy in the previous work with respect to location privacy and query privacy. Realizing the need for a unanimous query processing model which can operate in centralize as well as distributed environment, also flexible enough to provide privacy for public queries (snapshot/continuous) as well as private queries (snapshot/continuous) for public and private locations. Finally we will quantify the benefits of our approach using sampled results through experiments that the proposed cloaking algorithm is scalable, efficient and robust to support anonymity irrespective of scale of user queries in real time scenario.**

I. Introduction

Ubiquitous computing is the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively in- visible to the user. LBS play a pivotal role in ubiquitous computing. Due to the proliferation of location based devices, prolific growth has been made to expand the huge base of LBS. For example, you might have observed the abrupt turning up of cabs at the stand since they have been using GPS devices, one of the most recent applications of LBS to track passengers and routes where the mobile user is willing to reveal his or her location. On the other hand, mobile users want to access LBS to locate nearest hospital without revealing his or her location as per privacy concern. Other examples include real time traffic congestion monitoring, detailed directions, integrated search results on dynamic maps, satellite imagery, location-based advertisements and etc. In location-based applications, location-based database server is responsible to process location-based query triggered by user with respect to the revealed spatial information [12]. In fact this authoritative location-based database server is untrusted primarily relying on the revealed user location, thereby raising critical concerns related to privacy and security of the registered users. In order to exploit location-based services, users have to compromise their privacy with such untrusted location-based database servers or simply ensure their privacy by limiting the quality of location-based services. Thus an adversary may access sensitive information related to user by breaching the security of untrusted location-based server or may reveal the user information by examining trends of different publicly available location-based data from such untrusted applications. In order to build up the confidence of users in location based services, there is a pressing need to introduce such privacy preserving models that not only mitigate the privacy and security threats but also provide efficient and scalable computing mechanism. In this regard several cloaking algorithms haven been proposed by research community to preserve user privacy for motivating users towards location based services [9]. In existing approaches [3, 2, 17, 20, 8], user is supposed to indicate his/her privacy requirement in terms of K-anonymity model which is then blurred into a spatial region by cloaking algorithm. Thus the efficiency of cloaking algorithm has direct impact on the characteristics of privacy preservation model with respect to quality, flexibility and privacy.

Following are the shortcomings that motivated us for our contribution for privacy preservation towards location-based services.(i) Existing approaches employ cloaking algorithms on inefficient data structure such as hierarchical data structure [17] which supposed to exhibit poor performance in real time scenario when the scale and speed of mobile users are changed vigorously. (ii) Centralized anonymizer [17] subjects to bottleneck for the overall sys- tem in the presence of poor data structure such anonymizer can be compromised or yield dithered computing performance if the underling data structure is not flexible. (iii) Irrespective of being static or dynamic model several approaches lack efficiency in the absence of particular data indexing model.

Our motivation to conduct this research study is to present a well balanced query processing architecture which is efficient as well as privacy aware while the focus of existing approaches seems to skew towards privacy irrespective of overall efficiency of privacy preservation model. Our contribution made in this paper can be summarized as follows:

- Deploying R-tree based indexing scheme for location anonymizer to make best use of available resources i.e. memory and processing time
- Development of efficient cloaking algorithm to evaluate optimized result set in response to user queries
- Performance analysis of conducted experiments' to demonstrate the effectiveness of the proposed algorithms

The structure of the paper is as follows. Section 2 presents related work in the area of LBS with respect to privacy preservation. Section 3 outlines the architecture of location anonymizer. Cloaking algorithm along with different practical scenarios is discussed in section 4. In section 5, we describe the performance evaluation of our proposed cloaking algorithm in contrast with other existing approaches. Finally, we conclude our work in section 6.

II. RELATED WORK

The K-anonymity model [18] is one of the premier models in the data privacy domain which have influenced research community unanimously. In the most general sense, K-anonymity model addresses the concern of privacy when it comes to release a version of private data by data holder for practical usage with some scientific assurance that privacy of individuals cannot be compromised by matching similar trends of different version of data. L. Sweeney [20] defines K-anonymity as : A relation is said to be Kanonymous provided each record in the relation is indistinguishable from at least K-1 other records with respect to a set of quasi-identifier attributes. In the context of Location Based services, K-anonymity model ensures the privacy of mobile users by making their locations indistinguishable among at least other K-1 users. The synergy of K-anonymity model with spatio-temporal cloaking [9] is one of the open debates among researchers interested in emerging privacy preservation models. Thus K-anonymity is one of the crucial requirements for flexible and efficient location-based query processing model [3] as widely discussed in Clique- Cloak algorithm [2], spatio-temporal cloaking algorithm [9] and peer-to-peer spatial cloaking algorithm [5]. The Clique-Cloak algorithm [2] can only accommodate few users with limited K-anonymity requirements due to computation overhead and also suffer from topographical adversary attacks. The spatio-temporal cloaking [9] technique fails when it comes to scalability since this technique is optimized to track each single movement individually. The peer-to-peer spatial cloaking algorithm [5] finds K-1 NN(Nearest Neighbour) of query source but this approach suffers from ”center-of-ASR” attack. Mokbel et al [6] presents algorithm for privacy preservation in scenarios where user can reveal its location for public queries while can also hide both location and query for private query at private location. The addressed algorithm works on snapshot queries as well as continues queries. The significant part of this work is the identification of centralized location anonymizer for location-based services implying a hierarchical data structure which is supposed to be invulnerable to adversary attacks such as query sampling for snapshot queries and query tracking for continues queries. This scheme is subjected to poor performance in the presence of hierarchical data structure in real time scenario and also proposed cloaking algorithm fails to defend maximum movement boundary attack and public queries versus private queries attack. Prive [8], a distributed system for query anonymization in LBS, comes with HilbASR algorithm. Prive only considers snapshot queries, lack of support for conituous queries and also static version of HilbASR lacks flexibility in terms of data indexing model. Since our approach is quite similar as that of Casper, query processing paradigm discussed in Mokbel et al [16], which relies on third-party middleware to cloak user locations. This framework is suffered from centralize architecture as it can be bottleneck for the overall system or may raise a concern for single point of failure. The third-party middleware can also be compromised or may subject to ban for disclosing sensitive information such as the case of Napster as deploying centralized architecture. Apart from that it requires user to continuously report anonymizer

III. System Architecture

The system architecture has three main components: the mobile user, the location anonymizer and the location based database server. The mobile user interacts with the system by registering privacy profile [7] which specifies the typical privacy requirement of user with respect to the K-anonymity (K) and the minimum area (δ). K-anonymous parameter of privacy profile simply indicates that among how many other users the particular mobile user does not want to be distinguishable, while specifies desired coverage of the cloaked spatial region. Larger values of and results in restrictive privacy thereby degrading quality of service. Different privacy profiles can be set by mobile users to meet the desired privacy level at any time. Next step for mobile users is to report their spatial location and/or spatial query to the location anonymizer. The location anonymizer, after receiving location updates from mobile users, employs spatial cloaking technique to blur the users location and/or queries into cloaked spatial region as per specification of privacy profile. The blurred cloaked spatial region is then sent to the location-based database server which is spatially tuned to deal with the spatial cloaked region instead of exact resultset rather than actual resultset is computed according to cloaked spatial region. The candidate resultset is then returned to the mobile user through location anonymizer which previously computed the blurred spatial region based on the privacy profile, eventually mobile user who initiated the query is responsible to extract actual result from the candidate resultset. The efficiency and accuracy of proposed system directly influenced by data structure scheme and cloaking algorithm Embedded in the core of system i.e. the location anonymizer and the strictness of privacy profile. In the next section, we will discuss how our proposed location anonymizer can really play its part to drive the overall system efficiently and accurately as far as underlying data indexing and cloaking algorithm is concerned. On the other hand, the trictness of privacy profile simply puts operational trade-off between the level of privacy and the QoS (Quality of Service) which is solely manifested by the mobile user.

IV. Location Anonymizer

The location anonymizer is a trusted third party that acts as the middleware between mobile users and back-end location-based database server. The location anonymizer incrementally keeps track the number of users residing in The system and also consistently keep track of continuous movement of mobile users. Therefore, a key question for Developing location anonymizer is: How accurate, scalable and efficient is the cloaking mechanism employed by location anonymizer? In order to address this primary concern for location anonymizer, we propose for the deployment of an indexing scheme to address mobile users efficiently and effectively. The fundamental assumption used for designing efficient location anonymizer is that the spatial datasets are indexed by the structures of R-tree family. R-tree and its variants [11] are considered excellent choices for indexing a range of spatial data such as points, line segments, rectangles, polygons and so forth and have already been deployed commercially (Informix and Oracle).The R-tree [10] is one of the most popular approaches for spatial access methods i.e., for indexing multi-dimensional information such as (x,y) coordinates of geographical data. R-tree is hierarchical in nature where high level node is termed as MBR(minimum bounding rectangle) that supposed to enclose a set of child MBRs at the next level in hierarchy except the lowest level where data objects are stored within the MBRs as depicted. Root node represents the coverage of whole space of system and then the overall system space is hierarchically decomposed. Until now a number of R-tree variants [15] have been developed by the research community [19, 4, 13, 14, 1] as for as optimized performance of this promising indexing scheme for real-world data is concerned. R+ trees [19], a dynamic index for spatial access methods, differ from conventional R-tree by avoiding overlapping of internal nodes by inserting an object into multiple leaves thereby improving performance of point query since all spatial regions are covered by at most one node as well as fewer path traversal. Beckmann et al. [4] proposed the R*-tree which is more efficient in insertion and space utilization than the R-tree based on force-reinserted technique to avoid the overhead of splitting a full node. The Hilbert R-tree [13], which uses a Hilbert space filling curve outperforms the R*-tree by giving a better space localization. SR-tree ”Sphere/Rectangle-tree” [14], an index structure for high-dimensional nearest neighbor queries, differs from other R-tree variants by combining utilization of bounding spheres and bounding rectangles thus showing improved performance on nearest neighbor queries by reducing both the volume and the diameter of regions. Conventional R-tree does not promise good worst-case performance but performing well when it comes to real-world data. The Priority R-Tree [1] is one the efficient R-tree variants and is at the same time worst-case optimal. Irrespective of different R-tree variants, in our proposed model the whole space can be decomposed into several spatial regions. The root of R-tree will represent the whole projected space. Each entry within a leaf node stores two pieces of information related to an element i.e. (i) Bounding box identifier of that element (ii) Number of users in corresponding bounding box. When a mobile user is registered in the system with some unique identifier and privacy profile, a hash table is maintained to store an entry of form (uid,prf,bid) where uid is the user id, prf is user defined privacy profile and bid is the bounding box identifier for bounding box holding that particular user. Any of existing R-tree scheme can be employed to search, insert and delete elements in the data structure. One of the crucial aspect of location-based application is to maintain the mobile user updates as such applications tend to operate in highly dynamic environment thus requires flexible data structure for frequent updates. When a mobile user changes its location, location update is sent to location anonymizer iin the form (uid,x,y) where uid is the useridentifier, x and y are new spatial coordinates after location update. In order to get new bounding box for the updated location of user, location anonymizer simply applies hash function h(x,y). Now depending on the resulting bounding box, If it matches with the previous bounding box then location anonymizer will not do any processing at all. In case if it does not match with the previous bounding box then location anonymizer performs three operations i.e. updating new bounding box, incrementing user in new bounding box and decrementing user in old bounding box. If a new user is registered, a new entry will be marked in hash table and then marked in R-tree by insertion procedure. If a user quits the system then its corresponding entry in hash table is deleted and then removed from R-tree by deletion procedure.

IV. Conclusion

The horizon of LBS is being diminished by the raising concerns for privacy and confidentiality. This paper presents a novel query processing architecture for privacy preservation which can be deployed to motivate mobile users towards LBS without compromising privacy as well as quality through efficient and robust cloaking algorithm. Existing privacy techniques in LBS such as Casper [16] and Prive [8] is suffered by adversary attacks [6], query tracking [6], query sampling [6] and maximum movement boundary attack. Apart from that performance of these existing schemes can further be improved by implying indexing technique such as R-tree. Our proposed scheme can be deployed in centralized as well as distributed environment and is free from all existing adversary attacks by devising robust, efficient and flexible cloaking algorithm for privacy *preservation*.

Acknowledgment

This is research paper is supported by the Research Grant of “The University of Suwon”.

References

[1] L. Arge, M. de Berg, H. J. Haverkort, and K. Yi. The pri- ority r-tree: a practically efficient and worst-case optimal r-tree. In SIGMOD ’04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data, pages 347–358, New York, NY, USA, 2004. ACM.

[2] L. L. B. Gedik. A customizable k-anonymity model for protecting location privacy. *ICDCS*, 2003.

[3] R. Bayardo and R. Agrawal. Data privacy through optimal k- anonymization. In *ICDE*, pages 217–228, Washington, DC, USA, 2005. IEEE Computer Society.

[4] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The r*-tree: An efficient and robust access method for points and rectangles. In *SIGMOD Conference*, pages 322–331, 1990.

[5] C.-Y. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spa- tial cloaking algorithm for anonymous location-based ser- vice. In *GIS '06: Proceedings of the 14th annual ACM inter- national symposium on Advances in geographic information systems*, pages 171–178, New York, NY, USA, 2006. ACM.

[6] C. C.Y and M. Mokbel. Enabling private continuous queries for revealed user locations. In *SSTD*, volume 4605 of *Lec- ture Notes in Computer Science*, pages 258–275. Springer, 2007.

[7] S. Duri, J. Elliott, M. Gruteser, X. Liu, P. Moskowitz, R. Perez, M. Singh, and J.-M. Tang. Data protection and data sharing in telematics. *Mob. Netw. Appl.*, 9(6):693–701, 2004.

[8] P. Ghinita, G. Kalnis, and P. Skiadopoulos. Prive: Anony- mous location-based queries in distributed mobile systems. *WWW*, 2007.

[9] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloak- ing. *MobiSys*, 2003.

[10] A. Guttman. R-trees: a dynamic index structure for spatial searching. In SIGMOD '84: Proceedings of the 1984 ACM SIGMOD international conference on Management of data, pages 47–57, New York, NY, USA, 1984. ACM.

[11] N. M. Hee Kap Ahn and H. M. Wong. A survey on multidimensional access methods. Technical report, 2001.

[12] C. S. Jensen, A. Friis-Christensen, T. B. Pedersen, D. Pfoser, S. Saltenis, and N. Tryfona. Location-based services: A database perspective. In *ScanGIS*, pages 59–68, 2001.

[13] I. Kamel and C. Faloutsos. Hilbert r-tree: An improved r- tree using fractals. In VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases, pages 500–509, San Francisco, CA, USA, 1994. Morgan Kauf- mann Publishers Inc.

[14] N. Katayama and S. Satoh. The sr-tree: an index structure for high-dimensional nearest neighbor queries. *SIGMOD Rec.*, 26(2):369–380, 1997.

[15] Y. Manolopoulos, A. Nanopoulos, A. N. Papadopoulos, and Y. Theodoridis. R-trees have grown everywhere. Technical report, 2003.

[16] M. Mokbel, C. Chow, and W. Aref. The new casper: Query processing for location services without compromising pri- vacy. *VLDB*, pages 763–774, 2006.

[17] M. Mokbel, C. Chow, and W. Aref. The new casper: A privacy-aware location-based databse server (demonstra- tion). *ICDE*, 2007.

[18] P. Samarati. Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engi- neering, 13(6):1010–1027, 2001.

[19] T. K. Sellis, N. Roussopoulos, and C. Faloutsos. The r+-tree: A dynamic index for multi-dimensional objects. In *VLDB* '87: Proceedings of the 13th International Conference on Very Large Data Bases, pages 507–518, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.

[20] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertain. Fuzzi- ness Knowl.-Based Syst*, 10(5):571–588, 2002.