ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

Mining Queries to find the Nearest Neighbors

Alma Mary Margret [1], Kishore Sebastian [2], Liz Maria Mathew [3]
PG Scholar, CSE, SJCET, Palai, India [1,3]
Assistant Professor, CSE, SJCET, Palai, India [2]

Abstract: Mobile devices with Geo positioning capabilities can send location-dependent queries to Location Based Services (*LBS*). To protect privacy of user, the location must not be disclosed to the server. Existing solutions utilize a trusted anonymizer between the users and the LBS. But the users are not protected from correlation attacks. Private Information Retrieval protocol is used to provide privacy to user location. This approach is secure, but it is expensive in terms of computational overhead. This paper specifies the use of Voronoi diagrams to solve the problem of secure outsourced *k*NN queries. Itemset with high profits are mined from the database. The itemset will be the query to the service provider. As the data owners do not have the infrastructure to process the query, so the dataset is outsourced to a cloud service provider. Secure voronoi cell enclosure evaluation is done to find the nearest neighbour.

Keywords: Location Based Service, Point of Interest, Anonymity, Private Information Retrieval, Voronoi Diagrams**.**

I. INTRODUCTION

The explosive growth of location-detection devices (e.g., cellular phones, GPS-like devices, and RFIDs) results in a wide spread of location-based applications. Examples of these applications include location-based store finders, traffic congestion and location-based advertisements. Users who are registered with location-based services continuously send their location information to the location-based database server. Upon requesting to a service, a registered user has to issue a location-based query to server. This query is executed at the server based on the knowledge of the user location [4, 5, 11, 15].

Although location-based applications along with the location-based query processing promise safety and provide privacy and security to their customers. The location-based query processor depends mainly on the implicit assumption that users agree to reveal their private locations. In order to obtain a location-based service, a user has to send her location [20]. If a user wants to keep her location information private, then she has to switch-off her location aware device and unsubscribe from the service temporarily. With untrustworthy servers, such model provides several privacy threats. For example, an employer may check on her employee behaviour by looking the places she has visits and the time of each visit, the medical records of a person can be inferred by knowing which all clinics a person has visits. In many cases, GPS devices have been used in stalking personal locations [1, 3]. The traditional approach of using a fake identity [2] is not applicable to location based applications where a location of a person can directly lead to the true identity. For example, asking about the nearest coffee shop to my home using a fake identity will reveal my true identity, i.e., a resident of the home.

Advances in sensing and tracking technology enable location-based applications but they create some privacy risks. Anonymity provides high degree of privacy. It save users from dealing with service providers privacy policies. It will also reduce the service providers requirements for safeguarding private information [25]. Here we replace the locations with larger cloaking regions to prevent the disclosure of exact user location [7].

But still the LBS can access the information of user from cloaking regions. To overcome this problem cryptographic strength protection was started in [22] and continued in [18,19]. The main idea is to improve existing Private Information Retrieval (PIR) protocols and allow the LBS to return the NN to users without learning any information about users locations. This work brought together encryption and geometric data structures that enable efficient NN query processing. The work in [7] use Voronoi diagrams [6] to solve the problem of secure outsourced kNN queries.

In the existing systems GPS devices are used to find the current location which is an online process. But in our proposed system is an offline system. Here we are not using the GPS device in order to find the current location of a user. Using HTML5 core technology we can find the current location without a GPS device.

II. RELATED WORK

In paper [8] presents middleware architecture and algorithms that can be used by a centralized location based broker service. The algorithms adjust the resolution of location information into spatial or temporal dimensions. It is to meet specified anonymity constraints based on the entities that *may* be using location services within a given area. This paper investigates a complimentary approach that concentrates on the principle of minimal collection. In this approach location-based services collect and use only de-personalized data—that is, *practically anonymous* data [14]. This approach promises benefits for both parties. For the service provider, practically anonymous data causes less overhead. It can be collected, processed, and distributed to third parties without user consent. For data subjects, it removes the need to evaluate potentially complex service provider privacy policies. Practical anonymity requires that the subject cannot be reidentified (with reasonable efforts) from the location data.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

Consider a message to a road map service that comprises a network address, a user ID, and coordinates of the current location. Identifiers like the user ID and the network address are obvious candidates for Reidentification attempts. For anonymous service usage, the user ID can be omitted and the network address problem is addressed by mechanisms such as Crowds [12] or Onion Routing [13]. It provides sender anonymity.

A. Anonymizing Location Information

In our system model, the mobile nodes communicate with external services through a central anonymity server. This anonymity server is part of the trusted computing base. In the initialization phase, the nodes will make an authenticated and encrypted connection with the anonymity server. When the mobile node sends time and position information to an external service, the anonymity server will decrypts the message, removes identifiers such as network addresses. It perturbs the position data according to the cloaking algorithms to reduce the Reidentification risk. The anonymity server can acts as a mixrouter [10]. It randomly reorders messages from several mobile nodes, for preventing an adversary from linking outgoing and ingoing messages at anonymity server. Finally, the anonymity server will forwards the message to the external service.

As per Koehntopp and Pfitzmann, "Anonymity is considered as the state of being not identifiable within a set of subjects, the anonymity set [14]. Inspired by Samarati and Sweeney [9] k-*anonymous is* considered as a subject with respect to location information, if and only if the presented location information is indistinguishable from the location information of at least k - 1 other subject. Location information is represented by a tuple containing three intervals ([x1, x2]; [y1, y2]; [t1, t2]). The intervals [x1, x2] and [y1, y2] describe a two dimensional area where the subject is located. [t1, t2] describes a time period during which the subject was present in the area.

B. Adaptive-Interval Cloaking Algorithms

The degree of anonymity is represented by the parameter k_{min}, is the minimum acceptable anonymity set size. The algorithm takes the current position of the requester, current positions of all other subjects in the area and coordinates of the area covered by the anonymity server as input. The algorithm is inspired by quad tree algorithms [16]. It subdivides the area until the number of subjects in the area falls below the constraint k_{min}. The quadrant, which still meets the constraint, will be returned.

The key idea is to delay the request until k_{min} vehicles have visited the area chosen for the requestor. The cloaking algorithm is modified to take an additional spatial resolution parameter as the input. It then identifies the monitoring area by dividing the space until the specified resolution is reached. The algorithm observes vehicle movements over this area. When k_{min} different vehicles have visited the area, a time interval [t1, t2] is computed as follows: t2 is set to the current time, and t1 is set to the difference between time of request and random cloaking factor. The time and area interval are then returned.

In paper [17], it present a framework for preventing location based identity inference of users who issue *spatial queries* to Location Based Services. The paper proposes transformations based on the *K*-anonymity concept. It is to compute exact answers for range and nearest neighbour search, without providing the information of query source. This method will optimize the process of anonymizing the requests and it will process the transformed spatial queries. Experimental studies suggest that the proposed techniques can be applicable to real-life scenarios with numerous mobile users.

Instead of directly sending the query to the LBS, he uses an *anonymizer*, which is a trusted server. He establishes a secure connection (e.g., SSL) with the anonymizer, which removes the user id from the query and forwards it to the LBS. The answer from the LBS is also routed to Bob through the anonymizer. Specifically, it prevent an attacker from inferring the identity of the query source by providing the *K*-anonymity. A dataset is *K-anonymized*, only if each record is indistinguishable from at least *K*-1 other records with respect to certain identifying attributes. A user sends his location and query to the anonymizer through a secure connection. The anonymizer removes the user id and transforms his location using a technique called *cloaking*. This technique hides the actual location using a *K-anonymizing spatial region* (*K*-ASR or ASR), which is an area that encloses the client that issued the query, as well as at least *K*-1 other users.

III.. LBS WITH PRIVATE INFORMATION RETRIVEL PROTOCOL

This framework does not require a trusted third party, because privacy is achieved using cryptographic techniques. With respect to existing work, this approach provides stronger privacy for user locations. It provides privacy guarantees against correlation attacks. We optimize query execution by employing data mining techniques. Users can ask location-dependent queries, such as "find the nearest coffee shop", and it is answered by Location Based Services (*LBS*) like Map request or Google Maps. However, queries may provide sensitive information about individuals lifestyle, habits, or may result in unsolicited advertisement (i.e., spam).

Recent research on PIR [4] resulted in protocols that allow a client to privately retrieve information from a database. Without these database

Server learns what information the client has requested. Most techniques are expressed in a theoretical setting, where the database is an *n*-bit binary string *X***.** The client wants to find the value of the *ith* bit of *X* (i.e., *Xi*). To preserve privacy, the client sends an encrypted request *q* (*i*) to the server. The server returns a value $r(X, q(i))$, which allows the client to calculate X_i. We focus on *computational* PIR, which employs cryptographic techniques, and relies on the fact that it is computationally intractable for an attacker to find the value of *i*, from the given value of *q* (*i*). Furthermore, the client can easily find the value of X_i on the basis of response from server $r(X, q(i))$.

DOI 10.17148/IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

U is the querying user and the LBS contain four points of interest *p*1, *p*2, *p*3 *and p*4. In an off-line phase, the LBS develop a KD-tree index of the POIs and partition the space into three parts *A*, *B*, *C*. To answer a query, first the server sends the regions *A*, *B*, *C* to *u*. The user finds the region (i.e., *A*) that contains him, and utilizes PIR to request all points within *A*. So the server does not know about the region it has retrieved. The user receives the POIs in *A* is encrypted form and calculates *p*4 as his NN. Implementations based on the Hilbert curve and on an R-tree variant. Upon asking a query, the client first retrieves the granularity of the grid, and calculates the grid cell that contains him (i.e., *C*2). Then, he employs PIR to request the contents of *C*2. He receives *{p*3, *p*4*}* (encrypted) and calculates P3 as the exact NN.

IV. PROPOSED METHOD

Utility mining is an important fact in data mining. Process of finding high utility itemset from database is the process of finding itemset with high profits. Utility of items in a transaction database consist of two aspects: external utility and internal utility. The method of finding the high utility itemsets consists of three steps: 1) Construct a global UP-Tree it scans the database twice. 2) Using UP-Growth or by Up-Growth+ it generate Potential High Utility Itemset (PHUIs) from global UP-Tree and local UP-Tree. 3) From the set of PHUIs it identifies the actual high utility itemsets. The itemset will act as the query to the service provider.

Each node N in UP-Tree consists of N.name, N.count, N.nu, N.parent, N.hlink and a set of child nodes. N.name is the node's item name. N.count is the node's support count. N.nu is the node utility, N.parent records the parent node of N. N.hlink is a node link which points to a node whose item name is the same as N.name. The construction of a global UP-Tree can be performed with two scans of the original database. In the first scan, *Transaction Utility* (TU) of each transaction is computed. At the same time, *Transaction Waited Utility* (TWU) is also accumulated. By *Transaction Weighted downward Closure* (TWDC) property, an item and its supersets are unpromising to be high utility itemsets if its TWU is less than the minimum utility threshold. Such an item is called an unpromising item. During the second scan of database, transactions are inserted into UP-Tree. The unpromising items and their utilities should be removed from the transaction.

A. *LBS With kNN Query Processing*

Users send their current location as the parameters of a query, and wish to receive the nearest POIs. User will send his or her interest and he current location of user is obtained through offline process. But typical data owners do not have the technical infrastructure to process queries on a large scale, so the data owners outsource data storage and querying to a cloud service provider. In paper [7] they provide model with three entities: 1. Data owner, 2. Cloud service provider and 3.Client. The data owner has a dataset with *n* two dimensional points of interest, but does not have the necessary infrastructure to run and maintain a system for processing nearest- neighbor queries from a large number of users. Therefore, the data owner outsources the data storage and querying services to a cloud provider.

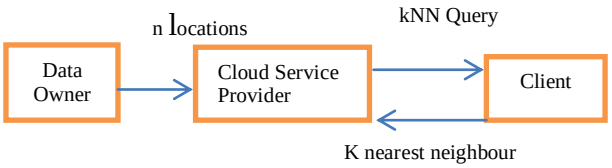

Figure.4.1 System Model

The server receives the dataset of points of interest from the data owner with additional encrypted data structures (e.g., Voronoi Diagram) needed for query processing. The server receives *k*NN requests from the clients, processes the query and returns the results to client. The client has a query point Q and wishes to find the point's nearest neighbors. The client sends its location query to the server, and receives k nearest neighbors as a result

B. *Secure Voronoi Cell Enclosure Evaluation*

We employ Voronoi diagrams [6], which are data structures especially designed to support NN queries. Denote the Euclidean distance between two points p and q by (p, q), and let $P = \{ , p , \dots , p \}$ be a set of n distinct points in the plane. The Voronoi diagram of P is defined as the subdivision of the plane into n convex polygonal regions (called *cells*) such that a point q lies in the cell corresponding to a point p if and only if p is the 1NN of q, i.e., for any other point p it holds that $dist(q, p) < dist(q, p)$ [6]. Answering a 1NN query boils down to checking which Voronoi cell contains the query point. we develop a secure scheme that determines whether a Voronoi cell contains the encrypted query point. The data owner sends to the server the encrypted vertices of the cell: $V(x_1, y_1)$, $V(x_2, y_2)$ and $V(x_3, y_3)$.

C. *VD-1NN protocol*

1. Data Owner sends to Server the encoded Voronoi cell vertices coordinates, MBR boundaries for each cell, encoded right hand side $R_{i,j}$, and encrypted $S_{i,j}$ for each cell edge.
2. Client sends its query point to the Server.
3. Server performs the filter step, for each kept cell it determines the edges that intersect the vertical line passing through the query point and sends the slope $S_{i,j}$ of the two edges to the Client.
4. Client computes the left hand side $L_{i,j}$, encodes it and sends it to the Server.
5. Server finds the Voronoi cell enclosing the query point and returns result to Client.

D. *Performance Analysis*

Runtime of all algorithms increases with increasing Transaction (T) because when T is larger, transactions and databases become longer and larger. Also, runtime of the methods is proportional to the number of candidates. The difference of the performance between the methods appears when T is larger than. The best method is UPT&UPG and the worst one is IHUPT&FPG.

DOI 10.17148/IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

The number of candidates generated by UPT&UPG is the smallest. This shows that UP-Growth can effectively prune more candidates by decreasing overestimated utilities when transactions are longer. In other words, UP-Growth is more efficient on the data sets with longer transactions. Memory usage of all methods increases with decreasing min_util since less min_util makes IHUP-Trees and UP Trees larger. On the other hand, memory usage increases with increasing database size. UP&UPG uses the least memory among the three methods. This is because the strategies effectively decrease the number of PHUIs and local UP-Trees. High utility itemsets are efficiently identified from the set of PHUIs which is much smaller than HTWUIs generated by IHUP. By the reasons mentioned above, the proposed algorithms UP-Growth achieve better performance than IHUP algorithm.

V. CONCLUSION

In this paper we have analysed the technical feasibility of anonymous usage of location-based services. Using new Technologies are able to find the correct location of the user without GPS devices. It describes that location data generate will create new and potentially more privacy risks than network addresses pose in conventional services. Reidentification and the location tracking risk can be reduced through k-anonymous data. It is affected by plaintext attack. To overcome this extension of PIR protocol is performed. But this process is expensive. VD-kNN has introduced to support k-nearest neighbour query processing.

ACKNOWLEDGMENT

The authors would like to thank Department of Computer Engineering, SJCET and indebted to our guide **Prof. Kishore Sebastian** for his guidance and sagacity without which this paper would not have been designed. He provided us with valuable advice which helped us to accomplish the design of this paper. We are also thankful to our HOD **Prof. Smitha Jacob** (Department of Computer Engineering) and our Coordinator **Prof. Teena Thomas** for their constant encouragement and moral support. Also we would like to appreciate the support and encouragement of our friends who helped us in correcting our mistakes and proceeding further to produce the paper with the required standards.

REFERENCES

[1] Foxs News. Man Accused of Stalking Ex-Girlfriend With GPS.http://www.foxnews.com/story/0,2933,131487,00.html.Proceedings. Tenth International Conference on, pp. 135-142. IEEE, 2004.
[2] A. Pfitzmann and M. Kohntopp. Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. In Proceedings of the Workshop on Design Issues in Anonymity and Unobservability, 2000.
[3] USAToday. Authorities: GPS system used to stalk woman. http://www.usatoday.com/tech/news/2002-12-30-gpsstalker
[4] C. S. Jensen. Database Aspects of Location-Based Services. In Location-Based Services, pages 115–148. Morgan Kaufmann, 2004.
[5] M. F. Mokbel and W. G. Aref. PLACE: A Scalable Location-aware Database Server for Spatio-temporal Data Streams. IEEE Data Engineering Bulletin, 28(3):3–10, 2005.
[6] Mark de Berg et.al., Computational Geometry, Springer
[7] Choi, Sunoh. "Secure kNN Query Processing in Untrusted Cloud Environments." (2014).
[8] Nabil R. Adam and John C.Worthmann. Security-control methods for statistical databases: a comparative study. ACM Computing Surveys (CSUR), 21(4):515–556, 1989.
[9] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, Computer Science Laboratory, SRI International, 1998.
[10] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 24(2):84–90, 1981.
[11] K. Mouratidis, D. Papadias, and M. Hadjieleftheriou. Conceptual Partitioning: An Efficient Method for Continuous Nearest Neighbor Monitoring. In SIGMOD, 2005.
[12] Michael K. Reiter and Aviel D. Rubin. Crowds: anonymity for Web transactions. ACM Transactions on Information and System Security, 1(1):66–92, 1998.
[13] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. Communicationsof the ACM (USA), 42(2):39–41, 1999.
[14] Andreas Pfitzmann and Marit Koehntopp. Anonymity, unobservability, and pseudonymity — a proposal for terminology. In Hannes Federrath, editor, Designing Privacy Enhancing Technologies — Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability, volume 2009 of LNCS. Springer, 2000.
[15] O. Wolfson, H. Cao, H. Lin, G. Trajcevski, F. Zhang, and N.Rishe. Management of Dynamic Location Information in DOMINO. In EDBT, 2002.
[16] Hanan Samet. The Design and Analysis of Spatial Data Structures. Addison-Wesley, Reading, MA, 1990.
[17] Kalnis P., Ghinita G., Mouratidis K., and Papadias D., Preserving location-based identity inference in anonymous spatial queries, TKDE'2007.
[18] Gabriel Ghinita, Panos Kalnis, Murat antarcioglu, and Elisa Bertino, A Hybrid Technique for Private Location- Based Queries with Database Protection, SSTD'2009
[19] Gabriel Ghinita, Panos Kalnis, Murat antarcioglu, and Elisa Bertino, Approximate and exact hybrid algorithms for private nearest-neighbor queries with database protection, Geoinformatica'2011
[20] Mokbel M. F., Chow C. Y., and Aref W. G., The new Casper: query processing for location services without compromising privacy, VLDB'2006
[21] B. Gedik and L. Liu. A customizable k-anonymity model for protecting location privacy. Technical Report GIT-CERCS-04-15, Georgia Institute of Technology, April 2004.
[22] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan, Private Queries in Location Based Services: Anonymizers are not Necessary, SIG-MOD' 2008
[23] Gedik B. and Liu L., Location privacy in mobile systems: a personalized anonymization model, CDCS '05
[24] Gruteser M. and Grunwald D., Anonymous usage of location-based services through spatial and temporal cloaking, MOBISYS'2003