Optimized kNN Query Processing using Clustering in Untrusted Cloud Environment

Lokesh.V
M.Tech. Student
Department of Computer Science and Engineering
Manakula Vinayagar Institute of Technology,
Pondicherry University,
Pondicherry.

Anandajayam. P
Assistant Professor
Department of Computer Science and Engineering
Manakula Vinayagar Institute of Technology
Pondicherry University,
Pondicherry

Shanmugasundaram. S
M.Tech. Student
Department of Computer Science and Engineering
Manakula Vinayagar Institute of Technology,
Pondicherry University,
Pondicherry

ABSTRACT
The query processing optimization is done using an efficient clustering method for the purpose of fast retrieval of the queries. The main desire of a user is to query regarding the point of interest such as nearby restaurants, cafes, etc., The Location Based Service (LBS) enables the user to access their information about their POI (Point Of Interest). Users send their present location as query and they want to get their output as the nearest POI. Due to absence of technical concerns to support query processing on wider scales, they present the data storage and querying to the cloud service provider. The user will present their query to the data owner and, the data owner will present the data storage and querying of the client, to the cloud service provider. The security for this query processing technique is provided by the mutable order preserving encoding (mOPE) and cryptographic security process (AES and DES). Here, they previously used cloaking region and PIR (Private Information Retrieval). Now we propose Voronoi Diagram along with a new clustering method known as Self Updating Clustering Method. By using this clustering method, we can get the nearest neighbor query in a fast and efficient manner.

Keywords
Location based service, point of interest, mutable order preserving encoding, and kNN.

1. INTRODUCTION
The optimized query processing is used in order to provide the result for query asked by the user in an easy and efficient manner. The LBS (Location Based Service) enables the users to access the information about their point of interest. In order to obtain NN Query, the user needs to send their co-ordinates to LBS. For the purpose of optimized query processing, we use two approaches namely cloaking region and Private Information Retrieval (PIR).The CR has a special k-anonymity paradigm and a trusted party is present between the user and LBS and generates a rectangular region that has at least k user locations. This approach is fast but it is in-secure. The second approach used id PIR (Private Information Retrieval).The PIR protocols tends the user to extract an object Xi from a set X = {X1,X2,X3} stored by server, in such a way that the server does not reveal about the value of i.

This approach is sure but it is highly expensive. In order to hide the information of the user from the cloud processor, the work in [4, 5, 6, 7] replaces the location with larger cloaking region. We can still gain confidential information from the cloaking region, so the cryptographic-protection was started in [1] and continued in [2, 3]. The main concept is to lengthen the PIR for binary set to special domains and also permit the LBS to return the NN to the user without revealing information about the user to the cloud processor.

2. RELATED WORK
Mokbel et.al [6], introduce Casper as a technique for mobile users to specify the location based services for a user privacy profile. Location anonymizer which explores the information of location for mobile users into cloaked special areas. Three novel query types that are not supported for location based service.

Gabriel Ghinita et.al [3] process the mobile devices using global positioning capabilities which allows to retrieve Points of Interest (POI). To protect, the user private information, it is important to co-ordinate the untrusted entities that provide location based services. To provide a hybrid, two step approach to private location based queries.

Gruteser et.al [4] provides a middleware architecture algorithm that can be used by a centralized location based service. The adaptive algorithm adjust the resolution of location information. It is used to findthe realistically expected special resolution for different anonymity constraints.

3. PROPOSED TECHNIQUE

Fig. 1. Basic System Model

Client is the one who requests for the query, he send his present location in the form of co-ordinates as a query in an encrypted form. The data owner acts as mediator between the client and cloud environment process. The data owner does not have the technical means in order to support the data storage and the querying from the client. So, he presents both these to the cloud processing environment. The cloud processing, processes the query from the client and does the solving of the client's query and returns it to the data owner, and the data owner returns it to the client. Here the cloud processor does not know about the information about the user since the information are in the encrypted form.

3.1 Voronoi Diagram

We use voronoi diagram in order to yield highly efficient and secure results, when querying about the nearest neighbor than the earlier used cloaking region and private information retrieval (PIR). Here the regions of voronoi diagram is divided by taking the center point between the two axes [8, 10]. The voronoi diagram can be generated using the following equation,

$x_1 < x_q < x_2$ for edge L_1

$x_1 < x_q < x_3$ for edge L_3

Here each region represents each area of location and whole group of region is known grid. The query which falls in the specific part of the grid is known as the region where the query of user lies.

3.2 Self-Updating Clustering in Voronoi Diagram

This clustering method makes the search of user the faster and easy process. This clustering methods separates the range of queries which lie on different zones such educational zone, entertainment zone etc., [9].

4. SYSTEM ARCHITECTURE

The client has a query point Q and wishes to find the point's nearest neighbors. The client sends its encrypted location query to the server and receives k-nearest neighbors as a result. The data owner has a dataset with n-two-dimensional points of interest, but does not have the necessary infra-structure 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. The server receives the dataset of points of interest from the data owner in encrypted format, with some of the extra encrypted data structures (e.g., Voronoi diagrams) needed for query processing. The server receives kNN requests from the clients, solves them and returns the results back to the client. Processing kNN queries on encrypted data requires complex operations, but at the core of these operations is a relatively simple scheme called mutable order-preserving encryption (mOPE). mOPE allows secure evaluation of range queries and is the only provably secure order-preserving encoding system (OPES) until today. The main difference between mOPE and previous OPES techniques is that it allows cipher texts to change value over time, hence the mutable attribute. Then kNN search for query processing takes place and then after the completion of it the voronoi diagram is generated in Fig.1.

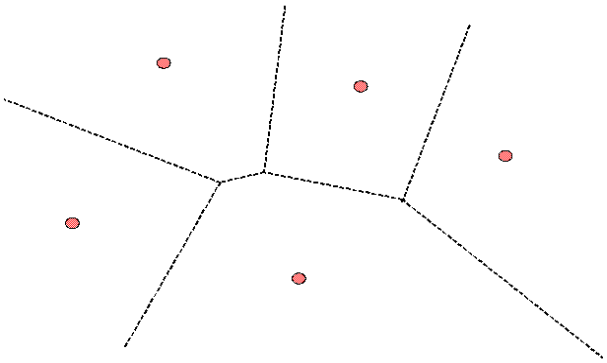

Fig.2. Voronoi Diagram

In order to obtain the nearest neighbor query processing and it also evaluated. Then the "Self Updating Clustering Method" is applied to the voronoi diagram in order to obtain an efficient and fast nearest neighbor query processing. Then finally the result is presented to the client.

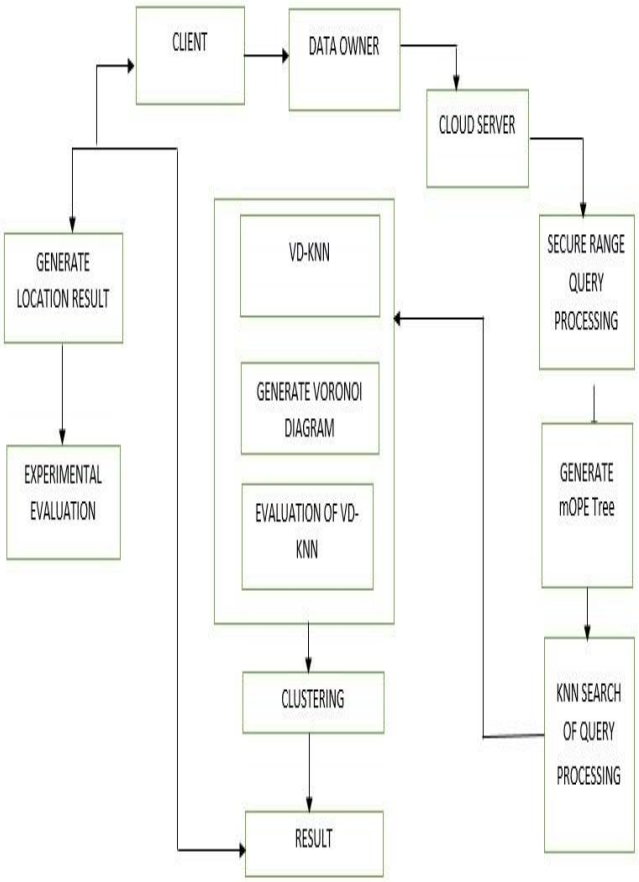

Fig.3 System Architecture

5. CONCLUSION

In this paper, we proposed a VD kNN in order to obtain an optimized nearest neighbor query processing. We include a "Self Updating Clustering Method" along with VD kNN process, in order to obtain fast and efficient nearest neighbor query processing. In future, we will elaborate the usage of VD kNN process with the new clustering technique which can be used to obtain a further efficient and accurate nearest neighbor query processing.

6. REFERENCES

[1] Gabriel Ghinita, PanosKalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan, "Private Queries in Location Based Services: Anonymizers are not Necessary", SIGMOD 2008

[2] Gabriel Ghinita, PanosKalnis, Murat Kantarcioglu, andElisa Bertino, "A Hybrid Technique for Private LocationBasedQueries with Database Protection", SSTD2009

[3] Gabriel Ghinita, PanosKalnis, Murat Kantarcioglu, andElisa Bertino, Approximate and exact hybrid algorithmsfor private nearest-neighbor queries with database protection, Geoinformatica 2011

[4] Gruteser M. and Grunwald D., Anonymous usage oflocation-based services through spatial and temporalcloaking, MOBISYS 2003

[5] Gedik B. and Liu L., Location privacy in mobile systems:a personalized anonymization model, ICDCS 2005

[6] Mokbel M. F., Chow C. Y., and Aref W. G., The newCasper: query processing for location services withoutcompromising privacy, VLDB 2006

[7] Kalnis P., Ghinita G., Mouratidis K., and PapadiasD.,Preserving location-based identity inference in anonymousspatial queries, TKDE 2007

[8] Sunoh Choi, Gabriel Ghinita, Hyo-Sang Lim and Elisa Bertino, Secure kNN Query processing in Untrusted Cloud Environment, IEEE 2014

[9] Wen-Liang Hung, Shou-Jen Chang-Chien and Miin-Shen Yang, Self-updating clustering algorithm for estimating the parameters in mixtures of von Mises distributions, International Journal of Applied Statistics 2012.

[10] Sunoh Choi, Gabriel Ghinita, Hyo-Sang Lim And Elisa Bertino, "Secure Knn Query Processing In UntrustedCloud Environments",IEEE Transactions On Knowledge And Data Engineering Vol: Pp No: 99 Year 2014