International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 24 Issue 11 – AUGUST 2017.

MOPE: DATA MINING ON LOCATION BASED SERVICES FOR SECURE DATA QUERY PROCESSING

N. Deshai[#1], P. Penchala Swamy[*2], Dr. G.P. Saradhi Varma[*3]

[#]*Assistant Professor, M.Tech, Department of Information Technology, S.R.K.R Engineering College, Bhimavaram, India*

[*2]*PG Scholar (M.Tech), Department of Information Technology, S.R.K.R Engineering College, Bhimavaram, India*

[*3]*Department of Information Technology, S.R.K.R Engineering College, Bhimavaram, India*

***Abstract*— Nowadays, the information of the current locations such as restaurants, college, home etc can be retrieved, through mobile devices having geo-positioning capabilities. Users are interested in querying their physical locations. The query parameter is sent by the user to know the result of the nearest position, i.e., nearest-neighbors(NNs) [1]. Locations of the user's and their interests may vary. Handling of sensitive information is a very tough task. The storage of the information is also a big issue. Hence the data owner does not make data accessible to all customers. It's allowed only for the paying customers. User sends their current location points and want to know about nearest POI's in NN, but data owner does not have that much storage capacity so we are using cloud service. Cloud provides powerful storage at low cost but is not fully trusted. So we are processing NN queries in an un-trusted outsourced environment, whereas at an equivalent time protective, POI and querying user's location positions. These techniques are based on Mutable Order Preserving Encoding (MOPE) [2]. It is a secure order-preserving encryption and provides increase in performance optimizations process and decrease the computational cost.**

***Index Terms*— Location privacy, Spatial databases, Database outsourcing, Mutable Order Preserving Encoding (MOPE), Point-of-interests (POI's).**

I. INTRODUCTION

In recent years, database outsourcing has gained tremendous popularity. In order to reduce operation and maintenance costs, an existing content distribution network environment may be used for database outsourcing; Database outsourcing involves three types of entities: data owners, service providers, and users. A data owner outsources the database functionality to one or more third parties which are called service providers which have the computational power to support various query processing. Users direct their queries to the service providers.

Database outsourcing has several advantages:

1) As the data owners store their data on the service providers, they do not need to have their own facilities to store and process the data.

2) Using third party service providers is a cheaper way to achieve scalability than fortifying the owner's data center and providing more network bandwidth for every user.

3) The database outsourcing model removes the single point of failure in the owner's data center, hence reducing the databases susceptibility to denial of service attacks and improving service availability.

4) The user can get the query results by a service provider which is close in terms of network latency without need to contact the data owners directly [3].

However, Database outsourcing poses several security challenges because we cannot completely trust the third party service providers which can be corrupted by adversaries. The first challenge is privacy. For instance, in an application to find nearby friends, the server stores the locations of the friends. If the location database is outsourced but not properly protected, unauthorized users may get access on data which causes privacy bleaches to the data owners. In addition, not only the data is stored in the server but also the query is issued to the service provider which is a sensitive information that should be protected. since the service provider can know the location of the user.

The second challenge is authentication. In outsourced databases, the data owners delegate their database functionality like range query, kNN, proximity, top-k, SUM, etc to the service providers. If the service providers are compromised, they could return tampered results to the user. Authenticated query processing techniques guarantee authenticity and completeness of query results in outsourced systems. Authenticity ensures that all the results returned to users originate from the data owners and no spurious results are introduced. Completeness guarantees that all the results which satisfy the query are present in the result set. On the other hand, authentication can be used for location based access control. Location based access control is to give an access to an important information when a user is in a restricted area. In order to determine whether the user is in the restricted area, we can make the user to receive partial keys

from several Location- based service (LBS) devices only when the user is in the area. Then, when the keys are authenticated, the user is given an access to the information.

The third challenge is recovery. Several protocols acknowledge the above authentication issue and provide authentication in the presence of malicious service providers. All these protocols deal with stealthy attacks where the malicious service providers try to modify the result without being detected. Such techniques can verify whether the result is correct or not, and in case they detect that the result has been tampered, they raise an alarm.

Hence, they cannot identify and remove the malicious service providers, leaving the network vulnerable to denial-of-service attacks. So, when the results are not correct, we need to detect the malicious service providers and give the correct results in the next round by excluding them.

In order to achieve location privacy, we propose a novel approach to secure kNN query processing. Our methods support efficient and precise evaluation of conditions based on the ciphertexts of data and queries. Our solution relies on mutable order preserving encoding (MOPE) [2], a transformation that supports comparison between pairs of data items. MOPE has been proposed for the evaluation of numerical comparison. We adapt MOPE to support a broad range of condition evaluations such as polygon enclosure. We propose a secure kNN query processing method. Our solution has a reduced computational overhead and does not incur false positives.

II. LITERATURE SURVEY

Private Information Retrieval is the technique where mobile devices with geo positioning capabilities (e.g., GPS), help to support Location based Services (LBS). For privacy, the user location should not be disclosed. Existing solutions utilize a trusty anonymizer between the users and the LBS. This approach has many drawbacks: (i) All users should trust the third party anonymizer, which may be a single purpose of attack. (ii) An oversized variety of cooperating, trustworthy users is required. (iii) Privacy is bonded just for one photo of user locations; users aren't protected against correlation attacks. A unique framework is used to support personal location dependent queries that work on personal data retrieval. The framework doesn't need a trusty third party, since privacy is achieved via cryptanalytic techniques. Compared to existing work, the method achieves stronger privacy in user's location. It's the first to guarantee privacy against correlation attacks.

Order-preserving symmetric encryption (OPE) is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plaintexts. Here only a single copy is required for efficient encryption and decryption. They allow efficient range queries on encrypted data, that is a remote un-trusted database server is able to index the (sensitive) data it receives, in encrypted form, in a data structure that permits efficient range queries (asking the server to return ciphertexts in the database whose decryptions fall within a given range, say [a, b]). OPE not only allows efficient range queries, but allows indexing and query processing to be done exactly and as efficiently as for unencrypted data, since a query just consists of the encryptions of a and b and the server can locate the desired ciphertexts in logarithmic-time via standard tree-based data structures. OPE has also been suggested for use in network aggregation on encrypted data in sensor networks[6] and as a tool for applying signal processing techniques to multimedia content protection. Yet a cryptographic study of OPE in the provable-security tradition never appeared.

III. EXISTING SYSTEM

In the existing system it preserves both data privacy of the owner and query privacy of the client for processing the queries. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. Most of the existing studies including data outsourcing, address the data privacy and query privacy separately.

IV. PROPOSED SYSTEM

To overcome the problem of existing system we propose the techniques that allow processing of NN queries in an untrusted outsourced environment, while at the same time protecting both the POI and querying users positions. Our techniques rely on mutable order preserving encoding (MOPE), which guarantees indistinguishability under ordered chosen-plaintext attack(IND-OCPA). It also provide performance optimizations to decrease the computational cost inherent to processing on encrypted data, and consider the case of incrementally updating datasets and geometric data structures that enable efficient NN query processing, this system investigate the use of Voronoi diagrams and Delaunay triangulations to solve the problem of secure outsourced kNN queries. Our proposed system emphasize that previous work assumed that the contents of the Voronoi diagrams is available to the cloud provider in plaintext, whereas in our case the processing is performed entirely on cipher texts, which is a far more challenging problem.

Our proposed system consists of following four modules

- Spatial Database Module
- Location Privacy Module
- Database Outsourcing Module
- Voronoi Diagram-Based K Nearest Neighbor (KNN) Module

The spatial database is a database that is optimized to store and query data that represents objects defined in a geometric space. Most spatial databases allow representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverages, linear networks, and TINs. While typical databases are designed to manage various numeric and character types of data.

In Location privacy module, the dataset of points of interest represents an important asset for the data owner, and an important source of revenue. Therefore, the coordinates of the points should not be known to the server. It assume an honest-but-curious cloud service provider. In this model, the

server executes correctly the given protocol for processing kNN queries, but will also try to infer the location of the data points. It is necessary to encrypt all information stored and processed at the server. To allow query evaluation, a special type of encryption that allows processing on ciphertexts is necessary. In this paper, we use the mOPE technique from [2]. mOPE is a provably secure order-preserving encryption method, and our techniques inherit the IND-OCPA security guarantee against the honest-but-curious server provided by mOPE. Furthermore, we assume that there is no collusion between the clients and server, and the clients will not disclose to the server the encryption keys.

In Database Outsourcing Module, The server receives the dataset of points of interest from the data owner in encrypted format, together with some additional encrypted data structures (e.g., Voronoi diagrams, Delaunay triangulations) needed for query processing. The server receives kNN requests from the clients, processes them and returns the results.

V. VORONOI DIAGRAM-BASED K NEAREST NEIGHBOR (KNN) MODULE

A. *Voronoi Diagram*

It focuses on securely finding the 1NN of a query point. Voronoi diagrams[1], are data structures especially designed to support NN queries. An example of Voronoi diagram is shown in the below figure. Denote the Euclidean distance between two points p and q by$d(p,q)$, and let $P=\{p_1,p_2,\ldots,p_n\}$ 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 $dist(q,p)<dist(q,p')$ [1].Answering 1NN query boils down to checking which Voronoi cell contains the query point. In this system model, both the data points and the query must be encrypted. Therefore, we need to check the enclosure of a point within a Voronoi cell securely. Next, we propose such a secure enclosure evaluation scheme.

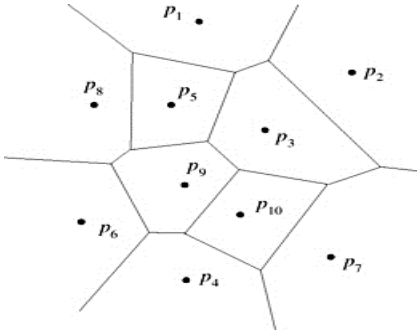


Fig.1. Voronoi diagram

Data Owner sends to Server the encoded Voronoi cell vertices coordinates, MBR boundaries for each cell, encoded right-handside , and encrypted S , for each cell edge. Client sends its encoded query point to the Server. Server performs the filter step, determines for each kept cell the edges that intersect the vertical line passing through the query point and sends the encrypted slope, of the two edges to the client. Client computes the left-handside L,encodes it and sends it to the server. Server finds the Voronoi cell enclosing the query point and returns result to client.

VI. RELATED WORK

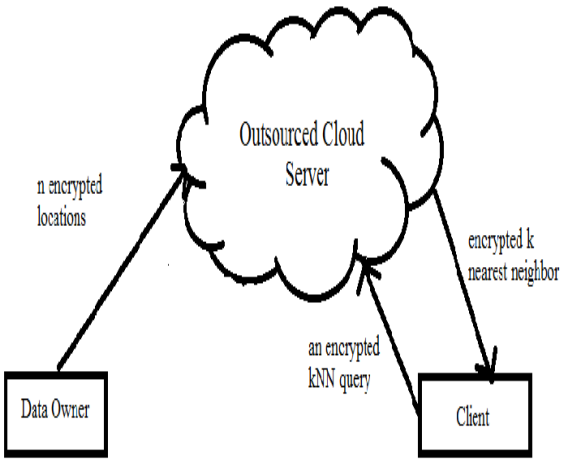
A. *SECURE RANGE QUERY PROCESSING METHOD*

As we are processing kNN queries on encrypted data it requires complex operations, but at the core of these operations sits 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) known to date. The difference between mOPE and previous OPES techniques is that it allows cipher texts to change value over time, hence the mutable attribute. Without mutability secure OPES is not possible [9].

Since our methods use both mOPE and conventional symmetric encryption (AES), to avoid confusion we will further refer to mOPE operations on plaintext/ciphertexts as encoding and decoding, whereas AES operations are denoted as encryption/decryption. The mOPE scheme in a client-server setting works as follows: the client has the secret key of a symmetric cryptographic scheme, e.g., AES, and wants to store the dataset of ciphertexts at the server in increasing order of corresponding plaintexts. The client engages with the server in a protocol that builds a B-tree at the server. The server only sees the AES cipher texts, but is guided by the client in building the tree structure. The algorithm starts with the client storing the first value, which becomes the tree root. Every new value stored at the server is accompanied by an insertion in the B-tree.

The server maintains a mOPE table with the mapping from cipher texts to encodings, Clearly, mOPE is an order preserving encoding, and it can be used to answer securely range queries without need to decrypt cipher texts. In addition, the mOPE tree is a balanced structure. Using a B-tree, it is possible to keep the height of the tree low, and thus all search operations are efficient. In order to ensure the balanced property, when insertions are performed, it may be necessary to change the encoding of certain cipher texts. Note that, the actual cipher text image does not change, only its position in the tree, and thus its encoding, changes. Typically, mutability can be done very efficiently, and the complexity of the operation (i.e., the maximum number of affected values in the tree) is O(logn) where n is the number of stored values.

VII. SYSTEM ARCHITECTURE



VIII. K-NEAREST NEIGHBOUR (KNN)

To support secure kNN queries, where k is fixed for all querying users, we could extend the VD-1NN method from by generating order-k Voronoi diagrams. However, this method, which we call VD-kNN, has several serious drawbacks:

(1) The complexity of generating order-k Voronoi diagrams is either ($k\ nlogn$) or (($n-k$)ogn+$nlog\ n$) , depending on the approach used. This is significantly higher than ($n \cdot logn$) for order-1 Voronoi diagrams.

(2) The number of Voronoi cells in an order-k Voronoi diagram is (($n-k$)),or roughly kn when k<<n. That leads to high data encryption overhead at the data owner, as well as prohibitively high query processing time at the server (a k-fold increase compared to VD-1NN). Motivated by these limitations of VD-kNN, we first introduce a secure distance comparison method (SDCM).

(3) In this paper, we devise Basic kNN (BkNN), a protocol that uses SDCM as building block, and answers kNN queries using repetitive comparisons among pairs of data points. BkNN is just an auxiliary scheme,very expensive in itself, but it represents the starting point for Triangulation kNN (TkNN), presented . TkNN builds on the BkNN concept and returns exact results for k=1. For k>1, it is an approximative method that provides high-precision kNN results with significantly lower costs [7].

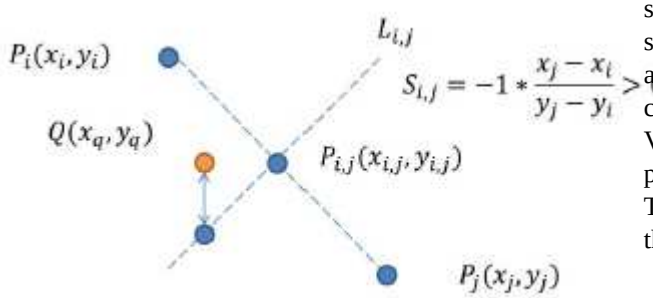


Fig.2. Secure Distance Comparison Method

IX. PRIVACY-PRESERVING QUERY PROCESSING FRAMEWORK

When processing distance-based queries, a multi dimensional index can be treated as traversal on the tree nodes. Very clearly, this may be divided into two alternate processes i.e. node traversal and distance access. The distance access determines the next node to traverse which is depending upon the distances computed from the current node and query point. To safeguard query and data privacy, both procedures must remain secure in the outsourcing model of three parties i.e. when query is being processing not only data owner but the cloud can identify the traversed nodes also or may obtain any information that may point out the query point as the exact distances to the query point. Till time, the client should have no access to the actual node contents during distance access and node traversal [10].

Here, the framework of secure queries processing. Whereas, other part is to protect data privacy, the client has only access to an encrypted version of the index, and must go ahead to process their query together with the cloud, which will decrypt the distances it, computes locally. The distance access is a collective procedure of the client and data cloud, in which not a single party has access to the actual distances [10].

X. ONE NEAREST NEIGHBOR (1NN)

A. Voronoi Diagram-based 1NN (VD-1NN)

In this section, we focus on securely finding the 1NN of a query point. We employ Voronoi diagrams, which are data structures especially designed to support NN queries. An example of Voronoi diagram. Denote the Euclidean distance between two points p and q by (p, q), and let P = {p1 , p2 , ….pn} be a set of n distinct points in the plane. The Voronoi diagram (or tessellation) 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 pi if and only if pi is the 1NN of q, i.e., for any other point pj it holds that (q, pi) <dist(q, pj) [1]. Answering a 1NN query boils down to checking which Voronoi cell contains the query point. In our system model, both the data points and the query must be encrypted. Therefore, we need to check the enclosure of a point within a Voronoi cell securely. Next, we propose such a secure enclosure evaluation scheme based on the secure range query processing method introduced to develop a secure scheme that determines whether a Voronoi cell contains the encrypted query point. Consider the sample Voronoi cell. For simplicity, we consider a triangle, but the protocol we devise works for any convex polygon as a cell. The data owner sends to the server the encrypted vertices of the cell: V1(x1; y1), V2(x2; y2) and V3(x3; y3) [1].

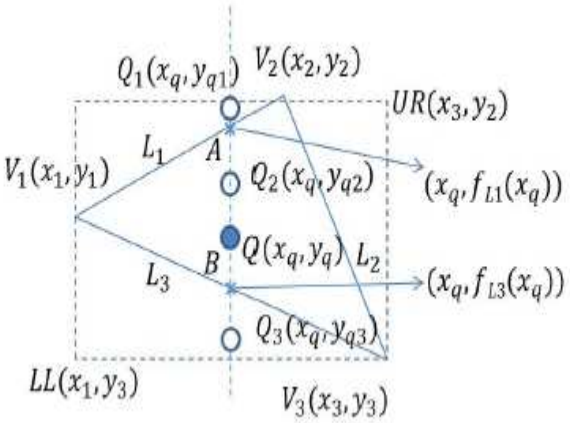


Fig.3. Secure Voronoi Cell Evaluation

B. TRIANGULATION-BASED KNN (TKNN)

Triangulation-based kNN(TkNN) reduces the overhead at the data owner. The Delaunay Triangulation is the dual of the order-1 Voronoi diagram. Any triangulation can be associated with the sorted angle sequence. The increasing sequence of angles (α1, α2… αm) appearing in the triangles of the triangulation. Among all triangulation of a given purpose set, the Delaunay triangulation has the lexicographically largest angle sequence. Since there are only a finite number of triangulations, lexicographically maximum triangulation should be met and this must satisfy the empty circle condition. Hence it is the Delaunay triangulation .We can reduce the data encryption time and query processing time to O (n), and the query encryption time to O (k) [12].

C. PERFORMANCE EVALUATION TKNN & VD-KNN

The main performance metrics used to evaluate the proposed techniques are query response time, encryption time and communication cost. Fig.4. Show the response time which measures the duration from the time the query is issued until the results are received at the client. It provides the computation time at the server and the client, still because the time required for transfer of final and intermediate results between client and server.

Fig.5. show Communication cost (measured in kilobytes) is important given that many wireless providers charge customers in proportion to the amount of data transferred.

Fig.6. shows the data encryption time at the data owner for VD-1NN and TkNN. VD-1NN generates 2*n voronoi points, whereas TkNN has n data points. In addition, the data owner must encrypt the right side, for each edge of every voronoi diagram cell and triangulation object. The total numbers of such edges is 3n for both VD-1NN and TkNN. The overall data encryption overhead of VD-1NN is proportional to 7n. And the TkNN is proportional to 5n. It captures this advantage of approximately 30% that TkNN has over VD-1NN [1].

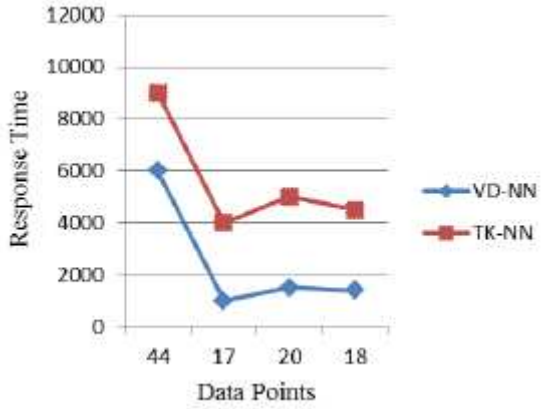


Fig.4. Response Time

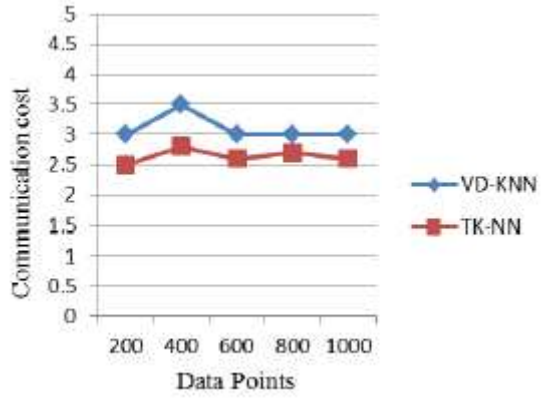


Fig.5. Communication Cost

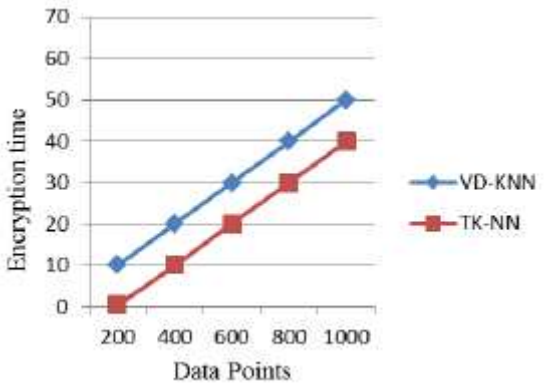


Fig.6. Data Encryption Time

XI. RESULTS AND ANALYSIS

The system is defined in such a way that it works in real time and shall give a better performance and measurable scalability factors. As it works under observation of applying system on various collection of databases which inherits complex, diverse, heterogeneous and generated by autonomous sources from network.

The Data Owner computes the order-1 Voronoi diagram of the dataset, determines the minimum bounding rectangle (MBR) boundaries of each Voronoi cell and encodes using MOPE[6] the cell vertices coordinates, as well as the right side, For each edge of a Voronoi cell. The slopes are encrypted using symmetric encryption (e.g.,AES). Generation time for the Voronoi diagram is (nlogn) using Fortune's

algorithm [7]. The number of Voronoi vertices that require mOPE encoding in a set of data points is at most 2n − 5 [1]. Thus, the time to encode Voronoi points is proportional to 4 since each Voronoi point has a x-coordinate and a y-coordinate. Furthermore, the right side, It must be encoded for each edge. The number of edges in a Voronoi diagram is at most 3 to 6. The total number of mOPE encoding operations is proportional to 7. The slopes are encrypted using AES encryption and do not require mOPE encoding. In total, the Data Owner performs 3*n* AES encryption and 7*n*mOPE encoding operations.

XII. CONCLUSION

In this paper, we implement secure k nearest neighbor query processing: VD-kNN which is based on Voronoi diagrams, and TkNN which relies on Delaunay triangulations. They both use mutable order preserving encoding (mOPE) as building block. VD-kNN provides exact results, but its performance overhead may be high. TkNN only offers approximate NN results, but with better performance. In addition, the accuracy of TkNN is very close to that of the exact method [1].

REFERENCES

[1] Der-Tsai Lee, On k-Nearest Neighbor Voronoi Diagrams in the Plane, IEEE Transactions on Computers, 1982.
[2] Raluca Ada Popa et.al., An Ideal-Security Protocol for Order-Preserving Encoding, IEEE SP, 2013.
[3] Weiwei Cheng. "Authenticating Multi-dimensional Query Results in Data Publishing", Lecture Notes in Computer Science, 2006.
[4] A. Boldyreva et.al., Order Preserving Symmetric Encryption, EuroCrypt, 2009.
[5] A. Boldyreva et.al., Order Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions, Crypto, 2011.
[6] D. Westhoff, J. Girao, and M. Acharya. Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation. IEEE Transactions on Mobile Computing, 5(10):1417–1431, 2006.
[7] Mark de Berg et.al., Computational Geometry, Springer.
[8] Choi, Sunoh, Gabriel Ghinita, Hyo-Sang Lim, and Elisa Bertino. "Secure kNN Query Processing in Untrusted Cloud Environments", IEEE Transactions on Knowledge and Data Engineering, 2014.
[9] research.ijcaonline.org
[10] www.comp.hkbu.edu.hk
[11] Mark de Berg et.al., Computational Geometry, Springer, 3rd Edition.
[12] Haibo Hu, Jianliang Xu, Chushi Ren, and Byron Choi, “Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism,”ICDE’11.
[13] Kalnis P., Ghinita G., Mouratidis K., and Papadias D., "Preserving location-based identity inference in anonymous spatial queries.,TKDE'07 28.
[14] R. Agrawal, J. Kiernan. R. Srikant, and Y. Xu, "Order preserving encryption for numeric data., SIGMOD'04.