Uncompromised Query Services with Optimal Privacy Preservation Schemes

[1]Iswarya.U [2]Ruhin Kouser.R
[1]Student [2]Assistant Professor,
Kingston Engineering College, Vellore, India

***Abstract* - Query services have experienced very big growth in the past few years. Because of huge usage services need to balanced which brings a need for outsourcing data management to Cloud service providers, which provide query services to clients on behalf of data owners. Data owner needs data confidentiality and query privacy to be guaranteed due to untrusted behavior of cloud service provider. Enhancing data confidentiality must not compromise the query processing performance. It will not be meaningful to provide slow query services as a result of security and privacy assurance. The proposed is random space perturbation (RASP) data perturbation method to provide secure and efficient range query and KNN query services for protected data in the cloud. The scheme enhances data confidentiality without compromising the query processing performance which also increases the user experience.**

***Keywords* - range, KNN, data transformation, privacy, query services**

I. INTRODUCTION

Due to plenty of users query services had shifted to the cloud for their non-interrupted accessibility and to reduce the infrastructure cost. With the cloud infrastructures, the service owners can conveniently scale up or down the service and only pay for the hours of using the servers. However because the service providers lose the control over the data in the cloud, data confidentiality and query privacy have become the major concerns. Adversaries, such as service providers, can possibly make a copy of the database or eavesdrop user's queries, which will be difficult to detect and prevent in the cloud infrastructures.

The existing related approaches such as crypto index are vulnerable to the attacks. The enhanced crypto index approach puts heavy burden on the in-house infrastructure to improve security and privacy. The new Casper approach uses cloaking boxes to protect data objects and queries which affect the efficiency of query processing and then in-house workload. In this project we propose random space perturbation (RASP) approach to constructing practical range query and k-nearest-neighbor (KNN) query services in the cloud.

We design the range query and KNN query in the cloud infrastructure, how it is efficiently processing the query. In structured databases the both of the queries are supporting and also satisfies the CPEL criteria. The good balance of the CPEL must compromise the cloud services user. It satisfies the processing cost and enabled the major attacks to detect in cloud to protect the sensitive data.

The structured construction should implemented and satisfies the privacy schemes. RASP Data Perturbation is basic idea of random noise detection, random injection, dimensionality expansion; these all combination can be satisfied in structured databases.

II. PROBLEM DEFINITION:

The two major problem of these idea that all of not in structured manner, some user wants in unstructured databases. First, In unstructured databases (it is not in in rows and columns) so how can we retrieve the result for respective query in efficient manner.

Second, Range query is not supports the unstructured databases. Range is fast query processing than KNN. Here, KNN gave the result for respective query but it is slow processing because KNN is clustering of data so searching is very complexity.

III. PRELIMINARIES:

3.1ORDER PRESERVING ENCRYPTION

OPE transforms the original form of data to another form of that original data. The transformed data should store in cloud databases for data confidentiality. The OPE schemes have three major technique to applies for transformation of data.

1. Scale, 2. Move, 3. Replace

EX: COMPUTER

Scale is setting some measure to multiply or addition or subtraction that is data owner's wish to complicate the de-transformed that data for hackers in cloud. Move is for scale measure to move one point from another point of the data and then replace the particular whole data.

C is in first place that is 1(C=1,O=2,M=3,P=4,U=5,T=6,E=7,R=8)now c is move to 3rd place and replace it and further alphabets are all transform using the above form of setting methods.

The De-transformed form of the original data of COMPUTER is totally transformed by using the scale, move and replace function to applies in data to store in cloud.

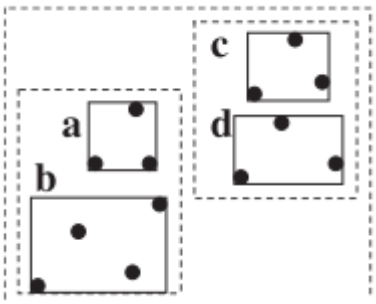
COMPUTER = TRMOPEUC.

Like, the example we have to transformed the whole document of the data should store in cloud, it well secure than other method how means even one letter is hacked in original form that takes

Minimum four annum so whole document means it is not possible.

3.2 RANGE QUERY

A Range query is the database operation that retrieves all the records between min and max bound region.(i.e It gives whole block for respective query, in that block prevalent data and also prevalent data both having) .

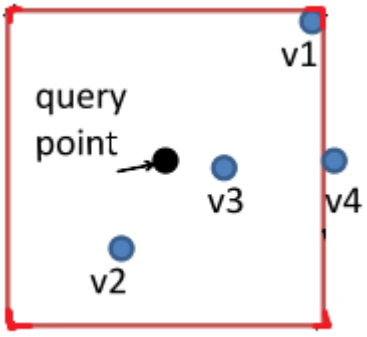


R-Tree Index

The data are store in rectangular blocks that have bounding region to travel around the node that get the respective region for respective given query. It is quick query processing to reduce the working in cloud so cost also low.

3.3 KNN QUERY

KNN (K-nearest neighbor) is the clustering of the data, so searching is very complexity. It gives most relevant result for respective query because of nearest neighbor, many miss rate it gave. User satisfies range query than KNN query. KNN query should support in structured databases and unstructured databases.



KNN structure

The query point which it points in spherical structure to maintain the particular level of the method to follows the data.

KNN (nearest neighbor) to updates the nearest point of respective query points to get the result of the query points.

3.4 MULTIDIMENSIONAL INDEX TREE

Multidimensional Index Tree is most efficiency than binary search tree in time complexity. It is derived from R-Tree algorithm where axis-aligned *minimum bounding region* "bounding boxes". Multidimensional Index Tree is mainly processing two-stage to retrieve the queried result.

In Multidimensional Index Tree compares the MBR and find the result of the queried points with respective to cloud services user. The Transformed queries describes the original data in perturbed space cannot directly processed by multidimensional tree.

IV. CONSTRUCTION:

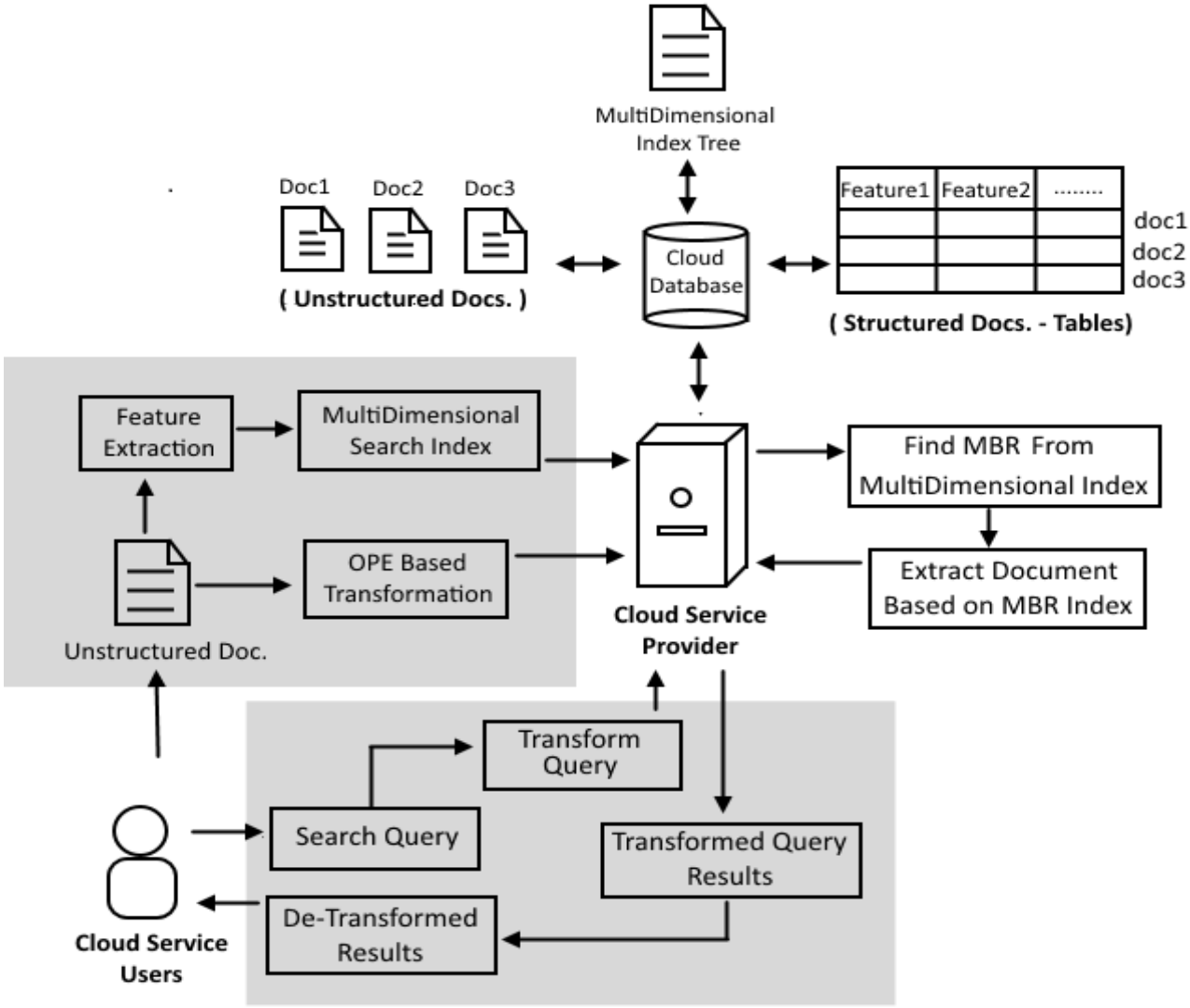
4.1 TRANSFORMING RANGE QUERIES

The Original form of the data can be transformed by using OPE algorithm. Order Preserving Encryption is security of data in the cloud databases. The 2D data can be transformed in 3D data for privacy of data in cloud. Data Confidentiality carried out scale, move and replace technique in string match in lot of word to give the result. The result processing in OPE with Eope in K keys schemes in some amount of data can be set of secured condition is processed. A polyhedron form of data can be transformed in hypercube form of data with respectively maintain the secure level be modeling in the untrusted behavior of third party. The avoidance of these untrusted manner the OPE scheme give the solution of queried answer. The secure set of query condition can also be in transformed to send the cloud and get the result for privacy of query.

4.2 MULTIDIMENSIONAL INDEX TREE

Most of Index tree is derived from R-Tree where algorithm in axis-aligned minimum bounding region MBR "bounding boxes". The Index tree in unstructured database can be processing in maximum stop words to be eliminating in particular

document. MBR is queried region to get the particular amount of the given query the region should analysis the block which matches the queried and indexing by multidimensional index tree algorithm. The MBR compares the queried region and index based on main form of the original data back to users.



System architecture

4.3 TWO-STAGE PROCESSING

4.3.1 FIRST STAGE PROCESSING

In first stage, the queried region will be find by MBR from multidimensional index tree and send to server from proxy of client side. The server then uses set of secured enclosed by MBR and block of the queried answer to be send in these stage.

4.3.2 SECOND STAGE PROCESSING

The whole block of queried answer can be find by MBR from index tree that should be maintaining the efficiency of query processing. The MBR of the polyhedron will possibly enclose the entire data set which is extracted from the first stage and the second stage dataset is reduced.

The return the exact range query result to the proxy server, which significantly reduces the post processing cost that the proxy server needs to take.

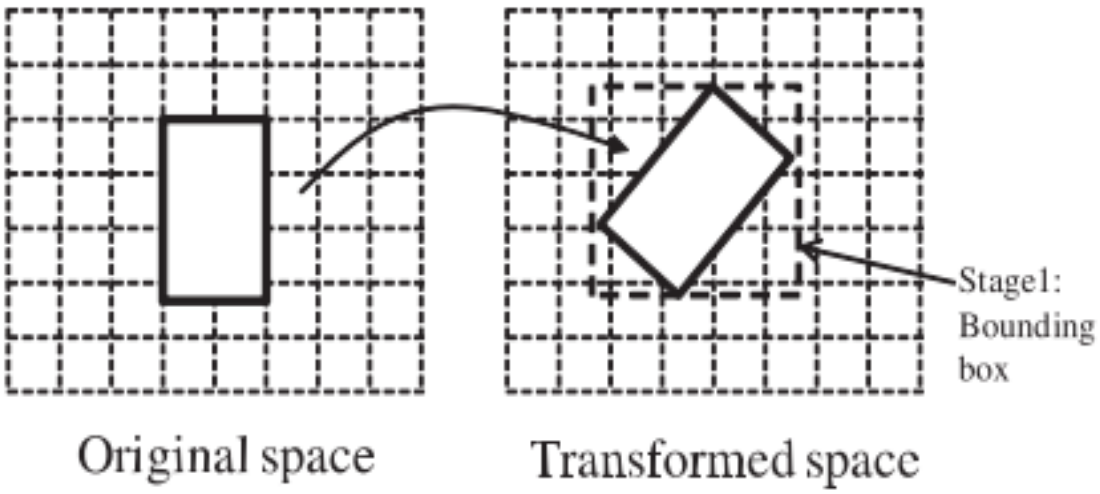


Illustration of Two-Stage Processing

4.4 UNSTRUCTURED DATA INDEXING

The unstructured data cannot be in regular form (there is no rows and columns). It is any other format the document file, text file like we have processing the efficient query processing. From that unstructured document extract the feature of the document. Feature of the document is mainly concentrate on keywords of the file to get efficient processing and construct the tree structure, to eliminate the stop words from the unstructured document to extract the file features. The MBR of given query traveling around the index tree to find the region of queried from multidimensional index tree. The exact document should extract based on MBR index then send back the transformed result to de-transformed result to the cloud services users.

V. IMPLEMENTATION DETAILS:

Step 1: Upload the text file (unstructured document) in cloud service provider.
Step 2: The document should be in transformed form using OPE algorithm.
Step 3: Extract the file feature from the uploaded document.
Step 4: The file feature can be changed into tree structure. (The data will be stored in leaf node)
Step 5: User give the query to retrieve the particular document, send query to cloud service provider that query also in transformed form to reach the CSP (transformed query for privacy of query).
Step 6: Query finds the MBR from multidimensional index tree.
Step 7: The MBR travels all node of the index tree and get the block of the query.
Step 8: In that block extract the exact document based on MBR index for given query.
Step 9: Finally the document is in transformed form, whenever the user gets the de-transformed result from cloud databases.

VI. VI.PERFORMANCE EVALUATION:

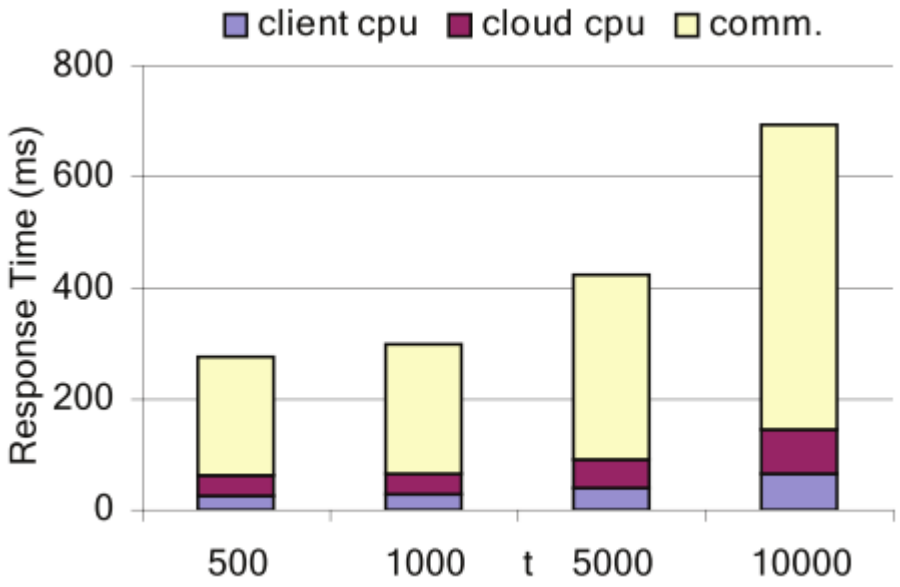
6.1 RANGE QUERY PERFORMANCE

Range queries are plots the client CPU, cloud CPU and the communication cost. The CPU times of both parties increase moderately as τ increases, and even for the largest distance τ = 10,000 (which returns 40 objects on average), both times are below 100 ms.

QUERY RESPONSE TIME	COMMUNICATION COST	CLIENT CPU	CLOUD CPU
0	275	25	100
200	300	50	125
400	410	75	150
600	650	85	175

The query response time, shown in Fig, increases moderately as increases. The same figure also shows the breakdown of the response time, and we find that the CPU times are dominated by the communication time, and their ratios of contributions are

stable regardless of τ. The manually calculate the range queries performance should be implemented in according with CPU cost , cloud cost ,client cost and communication cost all satisfies in optimal solution concerned in the performance evaluation.

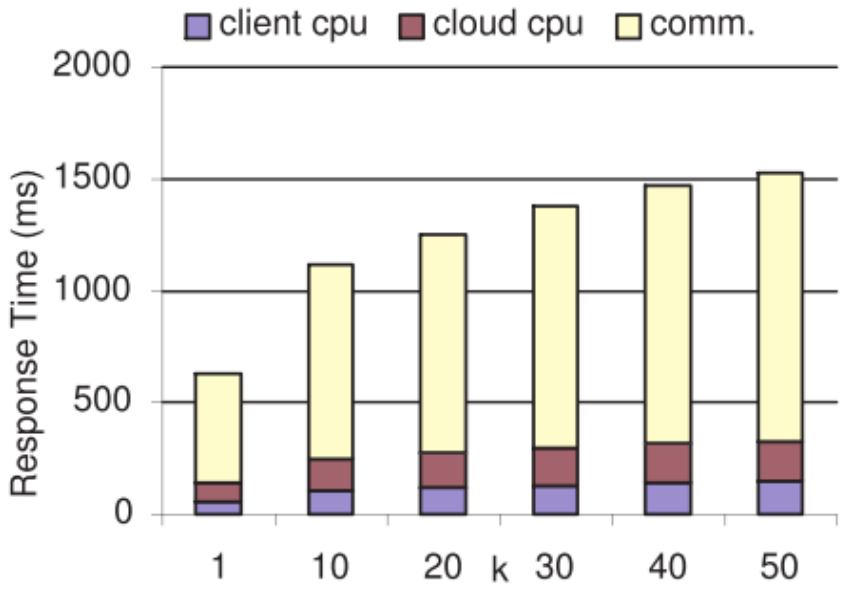


6.2 KNN QUERY PERFORMANCE

For KNN queries, we vary the number of nearest neighbor k from 1 to 50 plots the client CPU, cloud CPU and the communication cost. Similar to distance-range queries, the CPU times of both parties increase sub-linearly ask increases, and they never exceed 200 ms even for the largest k value. The trend in the communication cost coincides with those in CPU times, where the total transmitted data are fewer than 200 kb in all k settings. Although most results are similar to those in the distance-range query, we observe that distance folding gain less and becomes even less as k increases, until eventually when k = 50, its response time is even worse than applying the one-off approach alone. They totally reduce the client CPU time by more than 70%. It is also noteworthy that as τ increases, only the gain of distance folding decreases slowly. As a result, the query response time, increases moderately as k increases. The breakdown of the response time also shows the dominance of the communication time over CPU times, regardless of k. However, ask increases, this heuristic seems less viable — a nonlinear estimation of L should be used.

QUERY RESPONSE TIME	COMMUNICATION COST	CLIENT CPU	CLOUD CPU
0	610	250	75
500	1150	350	125
1000	1300	375	175
1500	1510	450	250

The performance can be compared with the KNN query and the range query. The KNN query will sense the nodes within the region and give the answer for the query whereas the range query is to overlap with the nearby regions, hence it will give the result of the query as much as possible and more queries will be answered in range queries.



The comparison the average wall clock time (milliseconds) per query for the two stages, the RPQ values for stage 1, and the purity of the stage-1 result. The tests are run with the setting of 10K queries, 20K records, 30 percent dimensional query range and 5 dimensions. Since the second stage is done in memory, its cost is much lower than the first-stage cost. Overall, the two stage processing is much faster than linear scan and comparable to the original R*Tree processing.

VII. CONCLUSION

Several sets of experiments to show the efficiency of query processing and the low cost of in-house processing. The two aspects: further improve the performance of query processing for both range queries and KNN queries; and formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality.

FUTURE WORK

The future work is mainly concentrate on storage aspect of server and avoid the duplication of data in perfect manner. The same copy of the data should be stored many times in same region and routing to many node all are control in one main authority also maintained in cloud services provider. The OPE for transformation of data on cloud, that OPE scheme will improve more secure when compare with RSA algorithm because in RSA using encryption takes 4 years to decrypt the single data. So that to improve much more security settings should added in OPE.

REFERENCES

[1]. Agrawal.R, J. Kiernan, R. Srikant (2004)"Order Preserving Encyption For Numeric Data.",Proc. Acm Sigmod Int'l Conf. Management Of Data (Sigmod), Vol.No:2, Issue:1,PP.No:23-32.

[2]. Cheema.M.A, L. Brankovic, X. Lin, W. Zhang (2010) "Multi-Guarded Safe Zone: An Effective Technique to Monitor Moving Circular Range Queries,",Proc. Int'l Conf. Data Eng. (ICDE),Vol.No:2,Issue:3,PP.No:189-200.

[3]. Chen.K And L. Liu (2011)"On Security Of Rasp Data Perturbation For Secure Half Space Queries In The Cloud.", Knowledge And Information Systems, Vol.No:29 Issue:2, PP.No: 657-695.

[4]. Ghinita.G, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L.Tan (2008) "Private queries in location based services: Anonymizers are not necessary". In SIGMOD,Vol.No:1,Issue:3,PP.No:56-64.

[5]. HaiboHu, Jianliang Xu,Chushi Ren, Byron (2011)"Processing Private Queries OverUntrustedDataCloudThroughPrivacyHomomorphism.",IJARCET,Vol.No:2,

[6]. Hore,S. Mehrotra, And G. Tsudik (2004)A Privacy Preserving Index For Range Queries" Proc.Very Large Databases Conf. (Vldb) Vol No:2,Issue:4,PP.No:45-63.

[7]. Huiqi Xu, Shumin Guo, and Keke Chen (2014) "Building Confidential And Efficient Query Servies In The Cloud With Rasp Data Perturbation."IEEE Transcation On Knowledge and Data Engineering",Vol.No:26,Issue:5, PP.No:322-332.

[8]. I-Fang Su,Yu-Chi Chung, PeiChi Liu, Chiang Lee (2013) "Processing Multiple KNearest Neighbour Queries.",International Journal for ebusiness, Vol.No:3 ,Issue:2 PP.No:124-130.

[9]. Jensen.C.S, D.Lin, B. C.Ooi, and R. Zhang (2006)" Effective Density Queries of Continuously Moving Objects", ,ICDE,Vol.No:2,Issue:2,PP.No:45-48.

[10]. Justin Zhan1, LiWu Chang2 and Stan Matwin(2005)"Privacy Preserving K-Nearest Neighbour Classification, "International Journal of Network Security", Vol.No:1, Issue:2, PP.No:46-51.

[11]. Kiruthigapriya Sengoden,Swaraj Paul (2013)"Improving the Efficiency of Ranked Keyword Search over Cloud Data", ,International Journal of Advanced Research in Computer Engineering&Technology,Vol.No:2,Issue:3,pp:881-883.

[12]. Li Liu, Murat Kantarcioglu and Bhavani Thuraisingham (2006) "The Applicability of the Perturbation Model-based Privacy Preserving Data Mining for Real-world Data", 6th IEEE Conference on Data Mining,Vol.No:3,Issue:4,PP.No:34-42.

[13]. Mokbe.M.F, C. Yin Chow, And W.G. Aref (2006) "The New Casper : Query Processing For Location Services Without Compromising Privacy.","Proc. 32nd Int'l Conf. Very Large Databases Conf. (Vldb),Vol.No:2,Issue:3,PP.No:763-774.

[14]. Muralidhar.K., R.Sarathi (2002)"A General additive data perturbation method for data base security", ,"journal of Management Science, Vol No:45,Issue:10, PP.No:1399-1415.

[15]. Nishant Doshi (2012)"A Novel Approach For Cryptography Technique On Perturbed Data For Distributed Environment" IJCIS Vol No:2,Issue:3,PP.No:763-774.

[16]. Nutanong.S, R. Zhang, E. Tanin, and L. Kulik (2008) "A Query-Dependent Approach to Moving Knn Queries" PVLDB Vol No:1,Issue:1,PP.No. 1095-1106. PP.No:58-64.

[17]. RussellPaulet, Md.GolamKaosar, XunYi,andElisa Bertino,Fellow (2013)"Privacy Preserving and Content-Protecting Location Based Queries",IEEE Transcation on Knowledge and Data Engineering,Vol.No:2,Issue-1,PP.No:23-28.

[18]. Sandeep Shivarudranavar,Mohan.k (2014)"RAQP-Range Query Privacy Preserving Scheme For Smart Grid Information System",IJEEDC,Vol.No:2,Issue:5, [19] Saral Elizabeth.E, Ms.K.Padmaveni (20140 "Confidential And Efficient Query Servies In The Cloud.",Vol.No:2,Issue:3,PP.No:456-554.

[19]. Shi.E, J. Bethencourt, T.-H.H. Chan, D. Song, and A. Perrig (2007) "Multi-Dimensional Range Query Over Encrypted Data.",Proc.IEEE Symp. Security and Privacy,Vol.No:1,Issue:3,PP.No:223-322.