Building Confidential and Efficient Query Services in the Cloud Using kNN-R and RASP Data Perturbation

MS. DIPALI S. SHINTRE[1] & DR. S. M. JAGADE[2]

[1]Department of M.E.(C.S.E.), [2]Principal T.P.C.T.'s College Of Engineering, Osmanabad, India.

***Abstract : In this paper, i*n cloud, to expand the performance and efficiency of query processing and to spare the workload of query handling, it is important to give secure query service to end-user. We propose this system is by using RASP approach to gain confidentiality and efficient range query and kNN query services for protected data in the cloud. Random Space Perturbation(RASP) is combination of many approaches such us random projection , dimensionality expansion and order preserving encryption(OPE). KNN-R algorithm is design to process range query to k-Nearest Neighbor(KNN) query and also these approaches are used to increase the working process of query by secure multidimensional range query processing. The kNN-R algorithm is intended to work with the RASP range query algorithm to process the kNN queries. We have thoroughly analyzed the attacks on information, data and queries under an absolutely characterized threat model and practical security assumptions. Broad experiments have been directed to demonstrate the focal points of this approach on security and efficiency of query processing in cloud environment.**

Index Terms : query services in the cloud, kNN query, range query,

I. INTRODUCTION

Cloud computing is the web based storage strategy. It is mostly utilized for storing and retrieving files, records and applications in its bases of datacenter. Many people utilizes the cloud on account of its smart features like unlimited of storage, secure service, great user stratification, low price and any time access, and also multiple user can access data and application at any time. With the developing of data on World Wide Web [1], Search Engines have turned into the main perspective to get to data on the web.

Manuscript received Dec, 2015.

Ms. Dipali S. Shintre, Master of Computer Science & Engineering, T.P.C.T.'s College of Engineering Osmanabad, India.
Dr. S. M. Jagade, Ph.D(E & TC), Principal, T.P.C.T.'s College of Engineering Osmanabad, India.

A typical actuality in Web Search is that a user frequently needs numerous iterations of query refinement to discover the desired results from an internet search engine.

Query services in the cloud are prominently increased due to the single points of interest in scalability and cost-saving. In cloud, the query service process are frequently utilized because, the user can save their expense and time. The owners in the cloud will give the amount just for their utilizing time of server. This is a most important feature that, the working time of query processing in cloud is extremely high and it is more expensive.

To secure the user data and query privacy, new techniques are need in the cloud. At the same time if the new approaches for giving security will give query processing not favorable element. We analyze the CPEL criteria for suggest a query in cloud. This CPEL paradigm indicates Confidentiality of data, efficient query processing, Privacy of query, Low working cost and less time. This technique likewise used to build the complexity of query service. In this paper the Random space Perturbation (RASP) technique used to build the query. Likewise separate the query as range query and kNN query. The proposed RASP technique will utilize the four ideas of the CPEL criteria and here the multidimensional data can be changed with the blend of order preserving encryption random noise injection and random projection.

The RASP technique and its combination gives confidentiality of information and this approach is mostly used to ensure the multidimensional range of queries in secure way, with indexing and query processing. Likewise it is used to develop functional extent query and kNN query services inside the cloud framework [1]. The range query is utilized as a part of database for recovering the stored data. It will recover the records from the database where it can mean some value between upper and lower limit.

II. RELATED WORK

The most relevant work about perturbation techniques includes the random noise addition methods and the condensation-based perturbation technique.

Authorized users with keywords: Their approach requires that the data owner provides the indices and keys for the server, and authorized Users use the data in the server. While in the cloud database scenario, the cloud server takes more responsibilities of indexing and query processing. Secure keyword search on encrypted documents scans each encrypted document in the database and finds the documents containing the keyword, which his more like point search in database. The research on privacy preserving data mining has discussed multiplicative perturbation methods [7], which are similar to the RASP encryption, but with more emphasis on preserving the utility for data mining.

Private information retrieval : *(PIR*) tries to fully preserve the privacy of access pattern, while the data may not be encrypted. PIR schemes are normally very costly. Use a pyramid hash index to implement efficient privacy preserving data block operations based on the idea of Oblivious RAM.

Another line of research facilitates authorized users to access only the portion of data in the authorized range with a public key scheme. The underlying identity based encryption used in these schemes does not produce indexable encrypted data. The untrusted service provider in our setting is responsible for both indexing and query processing. Secure keyword search on encrypted documents scans each encrypted document in the database and finds the documents containing the keyword, which is more like point search in database. The research on privacy preserving data mining has discussed multiplicative perturbation methods which are similar to the RASP encryption, but with more emphasis on preserving the utility for data mining.

Random Noise Addition Approach : The random noise addition approach can be briefly described as follows: A new decision-tree algorithm for the randomization approach is developed in order to build the decision tree from the perturbed data. Randomization approach is also used in privacy-preserving association-rule mining.

While the randomization approach is intuitive, several researchers have recently identified privacy breaches as one of the major problems with the randomization approach. The authors demonstrated that the randomization approach preserves little privacy in many cases.

Furthermore, there has been research addressing other weaknesses associated with the value based randomization approach. For example, most of existing randomization and distribution reconstruction algorithms only concern about preserving the distribution of single columns. There has been surprisingly little attention paid on preserving value distributions over multiple correlated dimensions.

Second, value-based randomization approach needs to develop new distribution-based classification algorithms. In contrast, our random rotation perturbation approach does not require modify existing data classification algorithms when applied to perturbed datasets. The randomization approach is also generalized to improve the balance between the privacy and accuracy.

Condensation-based perturbation approach : The condensation approach aims at preserving the covariance matrix for multiple columns. Different from the randomization approach, it perturbs multiple columns as a whole to generate entire “perturbed dataset”. The condensation approach can be briefly described as follows. The authors demonstrated that the condensation approach can preserve data covariance well, and thus will not significantly sacrifice the accuracy of classifiers if the classifiers are trained with the perturbed data. However, we have observed that the condensation approach is weak in protecting the private data. The *K NN* based data groups result in some serious conflicts between preserving covariance information and preserving privacy.

As the authors claim, the smaller the size of the locality in each group, the better the quality of preserving the covariance with the regenerated *k* records is. We design an algorithm that tries to find the nearest neighbor in the original data for each regenerated record. The result shows that the difference between the regenerated records and the nearest neighbor in original data is very small, and thus, the original data records can be estimated from the perturbed data with high confidence

III. PROBLEM FORMULATION

A. System Architecture

Cloud computing infrastructures used to store huge datasets and question administrations. The system architecture demonstrates two fundamental parts in it. The system data can be stored in the cloud database by data owner represented as d=n, here n represent as normalize form of data,d represent data and k represents key value provided by data owner, this key value used to encrypt original data. Encrypted data in cloud represented as d=e(d,k),here e is encryption key.

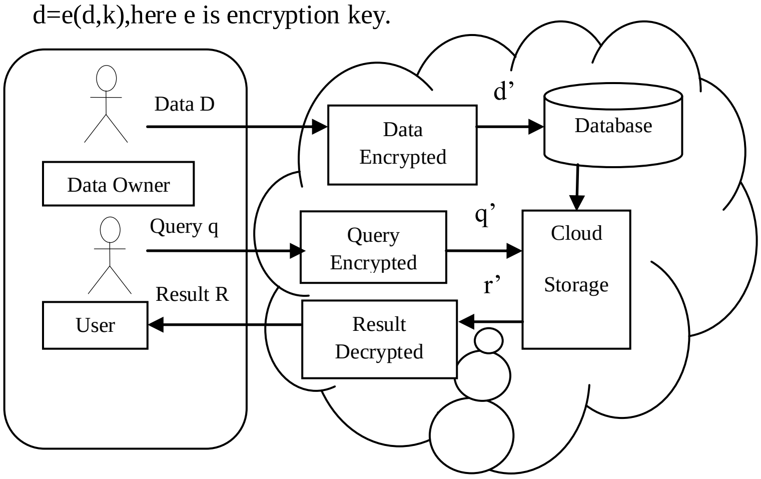

Fig. 1 System Architecture for RASP method.

The system architecture shown in below Fig. 1. The system diagram shows two types of parties or people involved in data access in cloud. Customer and Cloud service provider, here customer represent as end user who store their data in cloud. Cloud service provider has a responsibility to store customer data in secure format .Cloud service provider do encryption and decryption to ensure secure data processing. In customer party side we have data owner, end user, internal proxy server, and the users who can only submit queries. The data owners upload the perturbed data to the cloud. In the period in-between, the authorized users can submit range queries or kNN queries to find some records. The approved customer can submit range queries or kNN queries to discover a some records. Here the data owner can store their information in cloud while those information will encrypted in cloud and stored in the cloud database furthermore the data owner will give encryption key by utilizing this key value just cloud will encode the data by utilizing random space perturbation method.

The untrusted parties comprise of the inquisitive cloud service provider who hosts the query services and the ensured protected database. The RASP-perturbed data will be utilized to fabricate records to keep up query processing.

B. Modules

Five modules are used. They are user interface design, range query processing, kNN query processing, server protecting data, data confidentiality analysis.

User Interface Design :

The User Interface Design plays an important role for the user to move login the Application. This module has created for the security purpose. In this login page we have to enter user name and password, it will check username and password, if valid means directly go to home page, invalid username or password means show the error message and redirect to registration page. So we are preventing from unauthorized user entering into the login page to user page. It will provide a good security for our project.

Range Query Processing :

Based on the RASP perturbation method, we design the services for two types of queries: range query and kNN query. This section will dedicate to range query processing. We will first show that a range query in the original space can be transformed to a polyhedron query in the perturbed space, and then we develop a secure way to do the query transformation.

kNN Query Processing :

The original distance-based kNN query processing finds the nearest k points in the spherical range that is centered at the query point. The basic idea of our algorithm is to use square ranges, instead of spherical ranges, to find the approximate kNN results, so that the RASP range query service can be used.

Server Protecting Data :

Server protect the original data from attackers they provide random space data for user the user will use the data without loss in original data from server. Any attacker to corrupt the data random space data will be loss so the server what happen to delete the corrupt data and insert new clone of the original data from database. Server protect the original data from attackers.

Data Confidentiality Analysis :

As the threat model describes, attackers might be interested in finding the exact original data records or estimating them based on the perturbed data. Once the attacker is revocation that node will be eliminate from server. The server to analysis the user attribute and which user to use application in our server to maintain the user attribute and secure and to relieve the attackers in service.

C. RASP : RAndom Space Perturbation

RASP denotes Random Space Perturbation. RASP is one type of multiplicative perturbation, with a novel combination of OPE, dimension expansion, random noise injection, and random projection.

Random projection is mainly used to process the high dimensional data into low dimensional data representations. It contains features like good scaling potential and good performances. Random noise injection is mainly used to adding noise to the input to get proper output when we compare it to the estimated power. The RASP method and its combination provide confidentiality of data and this approach is mainly used to protect the multidimensional range of queries in secure manner and also with indexing and efficient query processing will be done. RASP has some important features.

Algorithm 1 RASP Data Perturbation

```
RASP Perturb(X,RNG,RIMG,Ko)
Input: X: $k \times n$ data records, RNG: random real value generator
        that draws values from the standard normal distribution,
        RIMG : random invertible matrix generator,
        Kope: key for OPE Eope; Output: the matrix A
$A \leftarrow 0$;
$A3 \leftarrow$ the last column of A;
$v0 \leftarrow 4$;
while A3 contains zero do
    generate A with RIMG;
end while
for each record x in X do
    $v \leftarrow v0 - 1$;
    while $v < v0$ do
        $v \leftarrow$ RNG;
    end while
    $y \leftarrow A((Eope(x,Kope))T , 1, v)T$ ;
    submit y to the server;
end for
return A;
```

In random space perturbation, the word perturbation is used to do collapsing this process will happen according to the key value that is given by the owner. In this module the data owner have to register as owner and have to give owner name and key value. And then the user have register and get the key value and data owner name from the owner to do access in the cloud. Here user can submit their query as range query or kNN query and get their answer. We analyze and show the result with encrypted and also in decrypted format of the data for the query construct by the user.

D. KNN-R: Using Range Queries to Process kNN Queries

we have mentioned, the quality of secure outsourced kNN query service can be summarized as the CPEL criteria: data Confidentiality, query Privacy, Efficient query processing, and Low inhouse workload.

The kNN-R algorithm aims to improve the confidentiality guarantee while preserving the efficiency of query processing. The basic idea is to use the RASP encryption to protect the confidentiality of data, and to use secure range query to protect the privacy of kNN query. The key is to develop an efficient kNN query algorithm based on the RASP encrypted data and queries. The design of kNN-R algorithm keeps the following problems in mind. (1) While the RASP protects the data confidentiality, it does not preserve distances or distance ranks. Therefore, the traditional distance-based kNN search algorithm does not work with the RASP encrypted data. Can we design a kNN search algorithm based on existing RASP range query algorithm? (2)Because of the limited computing capacity of the client side, the new algorithm should minimize the client's responsibility in query processing, which includes pre-processing, post-processing, and in-processing aid. Thus, the second question is how we design the algorithms to minimize the client's costs.

The kNN-R algorithm consists of five steps involving both the client and the server. The client will generate the initial upper bound range (that contains more than *k* points) and the lower bound range (that contains less than *k* points) and send them to the server. The server finds the inner range and returns to the client. The client calculates the outer range based on the inner range and sends it back to the server. The server finds the records in the outer range and sends them to the client. The client decrypts the records and pick the top *k* candidates as the final result.

Algorithm 1 KNN-R algorithm

```
1: The client generates the initial range and sends its secure form to the server;
2: The server works on the secure range queries and finds the inner range covering at least k points;
3: The client decodes the secure inner range from the server and extends it to the outer range, which is sent back to the server;
4: The server returns the points in the outer range
5: The client decrypts the points and extracts the k nearest points;
```

IV. EXPERIMENTAL RESULTS

In this experiment, we study the costs of the components in the RASP perturbation. The major costs can be divided into two parts: the OPE and the rest part of RASP.

RASP does not preserve the order of dimensional values because of the matrix multiplication component, which distinguishes itself from order preserving encryption schemes, and thus does not suffer from the distribution-based attack

Fig. 2 shows the cost distributions for 10K records at different number of dimensions. The dimensionality has slight effects on the cost of RASP perturbation.

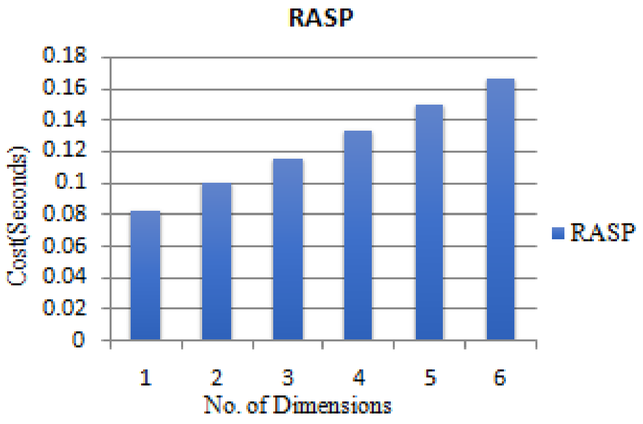

Fig. 2 The cost distribution of the RASP scheme.

An OPE scheme maps a set of single-dimensional values to another, while keeping the value order unchanged. Fig. 3 shows the cost distributions for 10K records at different number of dimensions. The dimensionality has slight effects on the cost of OPE.

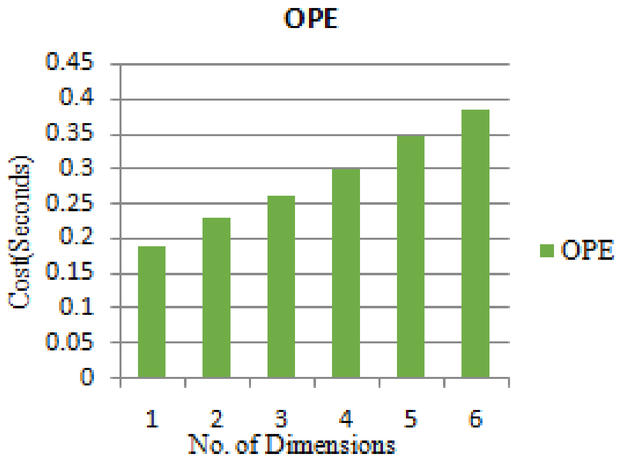

Fig. 3 The cost distribution of the OPE scheme.

V. CONCLUSION

To provide secure and efficient query services in cloud, RASP approach is used. Cloud base RASP data perturbation for building confidentiality and efficiency query services provide secure and efficient query services in cloud environment. To fulfill the requirement on low in house workload, cloud computing provide quality query services which is more efficient and very secure. This method mainly used to perturb the data given by the owner and saved in cloud storage. It also combines random injection, order preserving encryption and random noise projection and also it contains CPEL criteria in it. By using the range query and kNN query user can retrieve their data in secured manner and the processing time of the query is minimized.

VI. ACKNOWLEDGMENT

I would like to express my deep gratitude to my research, Dr. S. M. Jagade has a research guide, for their patient guidance, enthusiastic encouragement and useful critiques of this research work. I would also like to extend my thanks to the technicians of the laboratory of the Computer department for their help in offering me the resources in running the program.

Finally, I wish to thank my friends and parents for their support and encouragement throughout my project tenure.

REFERENCES

[1] Xu, H., Guo, S., and Chen, K. “Building confidential and efficient query services in the cloud with RASP data perturbation”, IEEE Transactions on Knowledge and Data Engineering 26, 2 (2014).
[2] K. Chen, R. Kavuluru, and S. Guo, “RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases,” Proc. ACM Conf. Data and Application Security and Privacy, pp. 249-260, 2011.
[3] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order Preserving Encryption for Numeric Data,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2004.
[4] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.K. Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing,” technical report, Univ. of Berkerley, 2009.
[5] J. Bau and J.C. Mitchell, “Security Modeling and Analysis,” IEEE Security and Privacy, vol. 9, no. 3, pp. 18-25, May/June 2011.
[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE INFOCOMM, 2011.
[7] K. Chen and L. Liu, “Geometric Data Perturbation for Outsourced Data Mining,” Knowledge and Information Systems, vol. 29, pp. 657- 695, 2011.
[8] K. Chen, L. Liu, and G. Sun, “Towards Attack-Resilient Geometric Data Perturbation,” Proc. SIAM Int’l Conf. Data Mining, 2007.
[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, “Private Information Retrieval,” ACM Computer Survey, vol. 45, no. 6, pp. 965-981, 1998.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,” Proc. 13th ACM Conf. Computer and Comm. Security, pp. 79-88, 2006.
[11] R. Marimont and M. Shapiro, “Nearest Neighbour Searches and the Curse of Dimensionality,” J. Inst. of Math. and Its Applications, vol. 24, pp. 59-70, 1979.
[12] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, “Executing SQL over Encrypted Data in the Database-Service-Provider Model,” Proc. ACM SIGMOD Int’l Conf. Management of
Data (SIGMOD), 2002.

AUTHOR PROFILE

Ms. Dipali Sambhaji Shintre, is a M.E student of Computer Science & Engineering from T.P.C.T.’s College of Engineering, Osmanabad India. She graduated in Computer Science and Engineering from BAMU University, Aurangabad. Her current research work focuses on ”Building Confidential and Efficient Query Services in the Cloud Using kNN-R and RASP Data Perturbation”.

Dr. S. M. Jagade received Ph.D. in (Electronics & Telecommunication) from SGGS IE & T Research center Nanded, ME (Ec) specialization in Computer Science from SGGS, College of Engineering & Technology, Nanded. completed BE (Electronics and Telecommunication) Degree from Govt. Engineering College, Aurangabad. He is currently working as a Principal in T.P.C.T.’s College of Engineering, Osmanabad India.