Survey on Confidential and Efficient Query Services in the Cloud

Shital V Patil[*] Mrs.R. S. Sonar Mrs.Y. A. Jakhade
IT EnggUniversity of Pune *IT EnggUniversity of Pune* *IT EnggUniversity of Pune*

Abstract— Today's, peoples are popularly used cloud computing infrastructures. So user can save their cost and time by using query services in cloud. But sometimes data owner does not move to cloud, because data may be hack from the malicious users when they use in cloud if not the confidentiality data and also privacy of a query is guaranteed. In cloud, to increase the efficiency of query processing and to save the workload of query processing, it is necessary to provide secure query service to user. The aim of propose this system is by using Random Space Perturbation approach provides confidentiality and efficient range query. RASP is combination of OPE, dimensionality expansion and random projection. KNN-R algorithm is design to process range query to kNN query. This method also provides secure the multidimensional range which is used to increase the working process of query.

Keywords— query services in cloud, kNN query, rang query, RASP, data perturbation

I. Introduction

Hosting data-extensive query services in the cloud is popularly increased because of the single advantages in scalability and cost-saving. Cloud computing is the internet based storage technique. It is mainly used for storing the files and applications in it infrastructures. Peoples uses the cloud because of its smart features like secure service, unlimited of storage, it will satisfy the user experience, low cost and multiple user can access the files and applications. In cloud, the query service process are often used because, the user can save their cost. The owners in the cloud will give the amount only for their using time of server. This is an important feature because, the working time of query service in cloud is very high and it is more expensive [2].

To protect the data and query privacy, new processes are need in the cloud. But if the new approaches for providing security will provide sloe query process is not an advantage. We examine the CPEL criteria for submit a query in cloud. This CPEL criterion denotes Confidentiality of data, query Privacy, Efficient query processing and Low working cost. This method also used to increase the complexity of query service.

In this paper the Random space Perturbation (RASP) method used to construct the query. Here also separate the query as range query and kNN query. The proposed RASP method will use the four concepts of the CPEL criteria and here the multidimensional data can be transformed with the combination of order preserving encryption, random projection and random noise injection.

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 way, with indexing and efficient query processing. It is also used to construct practical range query and kNN query services within the cloud system [1]. The range query is used in database for retrieving the stored data. It will retrieve the records from the database where it can denote some value between upper and lower boundary. The kNN query denotes k-Nearest Neighbor query. K denotes positive integer and this query are used to find the value of nearest neighbor to k.

II. QUERY SERVICES IN CLOUD

Query is mainly used for searching purpose. Queries are constructed by using structured query language. By retrieving the needed information from the database, queries are used. Query services are the method for services that are exposed through an implementation of service provider. Here by using RASP, range query and kNN query in cloud provide secure [7], fast storing and retrieving process of encryption and decryption of a data from database.

2.1 SYSTEM ARCHITECTURE

Cloud computing infrastructures like Amazon EC2, is used to store large datasets and query services. The architecture shows mainly two main parts in it. The data owner exports the perturbed information in the cloud space. The data's can be stored in the cloud via data owners d=n (d, k) here d represents data, n represents common form of data, k represents key value specified by the data owner. This format will be saved in the cloud as encrypted form d=e(d, k), here e stands for encryption. The above Fig.1 shows that clearly two separated parties: Customer and Cloud provider. Customer is the trusted party who store their data in cloud. And another is cloud provider, it store data in encrypted format. The customer party it contain the data/service owner, the internal proxy server, and the authoritative users who can only submit queries. The data owners export 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. Here the owner can store their data in cloud whereas those data will encrypted in cloud and stored in the cloud database and also the data owner will give key value by using that key value only cloud will encrypt the data by using random space perturbation method.

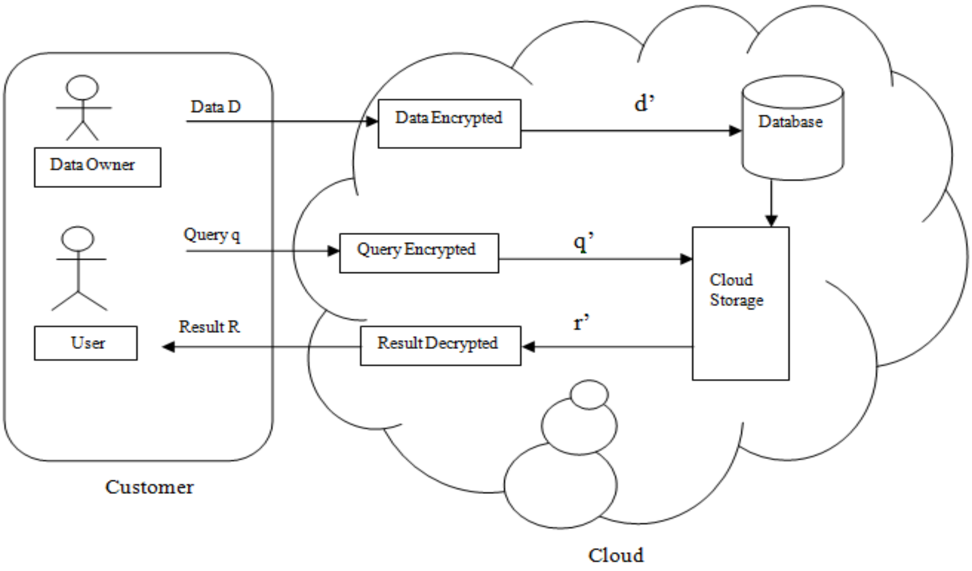

Fig. 1 System Architecture for RASP method

The untrusted parties consist of the curious cloud provider who hosts the query services and the protected database. The RASP-perturbed data will be used to build indices to maintain query processing.

2.2 SECURITY ANALYSIS

The some security analysis in the design shows the following

- Only authorized users have a key value which provided by the owner. So an authoritative user is not a malicious and will not purposely break the confidentiality. So only individual's users can send the queries for retrieving the information.
- The communication procedure among the user, owner and cloud and client system are well secured and no protected data records and queries can be leaked from cloud.
- RASP approach is used to give the protection of the query privacy and confidentiality of the data.

Attacker Goals: The major goal of attacker is to hack the original data from the database or identify the exact queries (For example, location queries) and break the privacy.

According to the level of preceding knowledge the attacker may have, we classify the attacks into two categories [1]

- Level 1: The attacker knows only the perturbed data and transformed queries, without any other previously knowledge. This corresponds to the cipertext-only attack in the cryptographic setting.
- Level 2: The attacker also knows the original information distributions, in addition individual attribute distributions and the joint distribution among attributes. In place into practice for some applications, whose statistics are gorgeous to the indefinite domain, the dimensional distributions may have been published through new sources.

III. MODULES

Following three modules are used. They are

- RASP
- Range query
- kNN query

3.1 RASP

RASP denotes RAndom Space Perturbation. The RASP data perturbation method is combination of OPE, random noise injection, and random projection, to give strong resilience to attack on the perturbed data as well as queries. It also conserves multidimensional ranges, which allow presented indexing techniques to be applied to speed up range query processing. OPE denote Order Preserving Encryption is a method of encrypting data so that it's possible to make efficient inequality comparisons on the encrypted items without decrypting them. Random projections are a powerful way for dimensionality reduction. Random projection is a process of projecting original high-dimensional data onto a lower-dimensional data representation.

Random noise injection is mostly used to adding noise to the input to obtain proper output when we compare it to the estimated power. The RASP method and its combination provide confidentiality of data and this approach is mostly used

to protect the multidimensional range of queries in secure mode and also with indexing and efficient query processing will be done. RASP has some important features. In RASP the use of matrix multiplication does not protect the dimensional values so no need to go through from the distribution based attack.
RASP does not preserve the distances between records, so it prevents the data that are perturbed from distance based attacks [8]. And also it won't protect more difficult structures it may be a matrix and other components. The range queries can be send to the RASP perturbed data and this range query describe open bounds in the multidimensional space.
In Random space perturbation, the perturbation is used to do collapsing this process will take place according to the key value that is specified by the owner. In this module the data owner have to register like owner and have to provide owner name as well as key value. And then the users have register and obtain the key value and data owner name from the owner to do access in the cloud. In this user can submit their query as range query or kNN query and obtain their answer. We examine and show the result with encrypted and also in decrypted format of the data for the query build by the user.

3.2 RANGE QUERY

Range query is the common database operation. It retrieves the data value from the database that values are in between upper bound & lower bound. The range query is not common because user won't know in advance about the result for the query, how many entries will come as result for the query.
For example,

```
        SELECT EMP_ID
        FROM   table name
        WHERE EMP_ID (

                        SELECT top 20*

                        FROM Canada

                        WHERE age>60
                );
```

The above given example shows the sample query for range query. This example retrieves the entries from Canada it will retrieve the Employee who are above 60 years in the top 20 list from the record of Canada.

The range search is mainly used to return the values which are present among the two specified values given in the query. For example database name is EMP_INFO then

```
Go
                SELECT EMP. id
                FROM EMP_INFO.EMP
                WHERE EMP_AGE BETWEEN 40 and 60
```

The above example will show one more example of range query search it will provide the entries of what are employee id that are present in EMP database with age above 40 and within 60. So by using range query user can simply retrieve the data's from records and this query process will be complete in secure manner as well as the speed of the query process will also increased.

3.3 kNN QUERY

kNN query denotes k-Nearest Neighbor query. This query is usually used to retrieve the nearest neighbor values of k. Here k is used to denote positive integer value. kNN algorithm is mainly used for classification and regression. The use of kNN-R algorithm is to process the range query to kNN query. This algorithm consists of two methods. That is used to make interaction between the client and the server. The client will send the query to the server with initial upper bound and lower bound. This upper bound range has to be more than the k points and the lower bound range have to be less than the k points. This process is shown in Fig. 2
The following figure shows the entire process of k-nearest neighbor query.

The above process is used to provide the inner range of the database by the server. With that inner range the client will calculate the outer range and send this outer range to the server. After that the server will search and find the records in the outer range from the database and send it to client and then the client will decrypt the record and get the top k files to provide the final result. This algorithm is used to find the compact inner square range for provide high accuracy and it has two not easy processes in it.

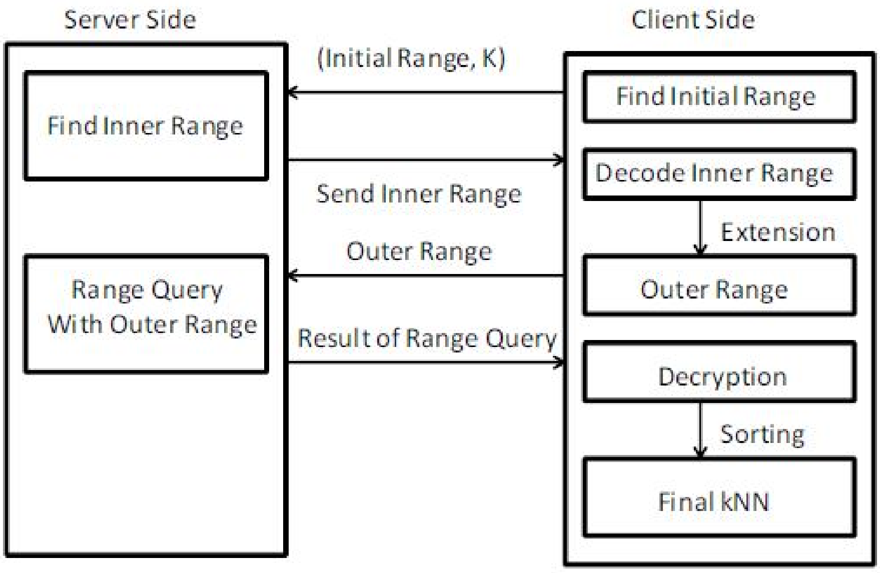

Fig.2 kNN query process

They are to find the number of points that are present in the square range and update of the boundary (i.e.) upper bound and lower bound is difficult because range queries are well secured by using random space perturbation. The security of kNN query and range query is equivalent.

IV. STUDY OF EXISTING PROCESS

In this we summarized about the study of exiting process.

OPE: OPE represents Order Preserving Encryption [1]. It is used for data that allows any comparison. It encrypts data. For that it possible to make efficient difference comparisons on the encrypted items without decrypting them. It allows database indexes to be built over an encryption table. The disadvantage of this process is the encryption key is too large and implementation makes the time and space overhead.

Crypto-Index: This approach is used for providing security and confidentiality of data within cloud. But it is vulnerable to the attack. The enhanced crypto-index approach put bulky load on the in-house infrastructure to develop the security and privacy [12].

Preserving Query privacy: This privacy preserving [5] multi keyword search is based on the simple text search. In this the searching method will done by ranking process. The problem of this concept is because of ranking process in-house processing time will be maximized.

New Casper approach: To protect data objects and queries here use new Casper approach, it uses a cloaking boxes. This approach affects query processing efficiency and the in-house workload [8].

V. CONCLUSION

We surveyed few methods that are used to provide a security to data in the cloud. 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 has contains CPEL criteria in it. By using the range query and kNN query user can retrieve their data's in secured manner and the processing time of the query is minimized.

REFERENCES

[1] Huiqi Xu, Shumin Geo, Keke Chen, "Building confidential and Efficient Query services in the Cloud with RASP Data Parturbation" IEEE Transaction on knowledge and data Engineering vol:26 no:2,2014.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. andAndy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," Technical Report, University of Berkerley, 2009.

[3] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model,"
in Proceedings of ACM SIGMOD Conference, 2002.

[4] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in *Proceedings of ACM SIGMOD Conference*, 2004.

[5] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *ACM Computer Survey*, vol. 45, no. 6,
pp. 965–981, 1998.

[6] B. Hore, S. Mehrotra, and G. Tsudik, "A privacy-preserving index for range queries," in *Proceedings of Very Large Databases Conference (VLDB)*, 2004.

[7] J. Bau and J. C. Mitchell, "Security modeling and analysis," *IEEE Security and Privacy*, vol. 9, no. 3, pp. 18–25, 2011.

[8] K. Chen, L. Liu, and G. Sun, "Towards attack-resilient geometric data perturbation," in *SIAM Data Mining Conference*, 2007.

[9] M. F. Mokbel, C. yin Chow, and W. G. Aref, "The new casper: Query processing for location services without compromising
privacy," in Proceedings of Very Large Databases Conference (VLDB), 2006, pp. 763–774.

[10] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *ACM Computer Survey*, vol. 45, no. 6, pp. 965–981, 1998.

[11] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model," in *Proceedings of ACM SIGMOD Conference*, 2002.

[12] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy preserving multi-keyword ranked search over encrypted cloud data," in *INFOCOMM*, 2011.