Confidential and Efficient Query Services in the Cloud Using RASP

Shital V Patil
Department of Information Technology,
Sinhgad College of Engineering, Pune
University of Pune, India.
patil.sheetal.01@gmail.com

Prof R S Sonar
Department of Information Technology,
Sinhgad College of Engineering, Pune
University of Pune, India.
rssonar.scoe@sinhgad.edu

***Abstract* —Now a day's cloud computing infrastructures are popularly used by peoples. Because of that cloud user can save their time as well as cost by using query services in cloud environment. In cloud, data owner upload their data in cloud and cloud user access that data. But sometime owner of data does not move to cloud or does not put their data in cloud because they not sure that this data is secure. Because some malicious users are present in cloud and they hack the data. So until cloud provider does not give the guaranteed of confidentiality and privacy of query, data owner does not move to cloud. So to increase the efficiency of query processing and save workload of processing of query, it is required to user to provide secure query services. The main purpose of using Random Space Perturbation approach is to provide confidentiality and efficient range query. RASP is combination of OPE, dimensionality expansion and random projection. To process range query to kNN query here kNN-R algorithm is used. It also used to increase the working process of query.**

Keywords - Query services in cloud, kNN query, rang query, data purturbation.

I. INTRODUCTION

Today's popularity of web-based applications is increased. Also it supports the widely available cloud infrastructures, so service based computing has become a most important computing paradigm. Basically cost of cloud infrastructures is very low, so service providers take this advantage at the same time as service users enjoy suitable services without worrying about the cost of maintaining hardware and software. In cloud, data is collected, stored in large manner. Those data is analyzed in business intelligence and scientific computing for a number of years. Cloud services are like online file storage, social networking sites, webmail, and online business applications. One of the attractive features of cloud infrastructures is that services owners can easily scale up as well as scale down the service and it simply pay for the hours of use of those servers. But sometimes service provider loses the control on the data in cloud, so data confidentiality and query privacy are not preserved. So it becomes the major concerns. For this reasons data owner does not ready to move on cloud until data confidentiality and query privacy are guaranteed because some data might be sensitive. Sometimes, data owners may not be alert of information leakage, which can occur in all kinds of possibilities, if the cloud provider does not want to report the leakage.

Some new approaches are required to maintain data confidentiality and privacy of query, because efficient query service is also important and the benefits of using the clouds must be preserved. A simple method is export dataset on cloud in encrypted format. But searchable encryption is very challenging, and it also shows limited achievement in some specific areas such as document search. The queries in database like range query and kNN query demand for fast processing time and it also support indexing structure.

For providing protection to data and maintain privacy of query, new approaches are needed in cloud. For that purpose here examine CPEL criteria for submit a query in cloud. These CPEL criteria stand for 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 RASP approach is introduced, which stand for Random space Perturbation method, which is used to construct the query. It constructs range query as well as kNN query.

The RASP method is used to provide confidentiality of data and it also used to protect the multidimensional range of queries in secure manner, with indexing and efficient query processing. For retrieving the stored data here range query is used. It retrieves values in between upper and lower bound. The kNN stand for k-Nearest Neighbor query. The combination of RASP and kNN query services is kNN-R which is used to RASP range query services to process the kNN queries.

II. RELATED WORK

To protect the data in cloud, some related approaches are used. But these approaches are not satisfactorily fulfilling all of these features. On the other hand, many approaches developed in the database community focus on performance, only providing very weak security.

Protected Assets. Information confidentiality and query privacy must be protected in the RASP approach. At the same time as the integrity of query services is also an essential issue, it is orthogonal to our study. Existing integrity examination and preventing techniques can be integrated into this framework [2]. So here assume the curious cloud provider is interested in the data and queries, but it will honestly follow the protocol to give the infrastructure service.

A. *Introduction of Preserving Query Privacy*

By using Private information retrieval (PIR) technique privacy of access pattern is preserve without encrypting the data [8]. But this method is very costly. This technique is used to improve the efficiency of query processing by using hash index. It implements efficient privacy preserving data-block operations. But problem of that, it require high throughput range query processing. It also goes thought the problem of query privacy. To jointly process kNN queries it needs the authoritative users, the owner of data, and the cloud. To improve the location privacy, private information retrieval method is used [10]. But this method not thinks about confidentiality of data.

B. *Existing Methodologies*

1) New Casper Approach

It is one type of privacy aware framework which is used for query processing. It is useful for Location Based Services. By using this method mobile and stationary user can obtain continuously services without providing their own private location. To protect data objects and queries here use new Casper approach, it uses a cloaking boxes. The problem of this approach is it affects the efficiency of query processing plus the in-house workload [7].

2) Crypto Index

Now a day's usage of internet is quickly increased. Due to the advances in software and networking organizations can share data for different reasons. Crypto index is formed over sensitive attributes which are considered important in queries in DAS (Database as a service) model. The purpose of Crypto-Index is to minimize work of client and force the server to perform most task of query processing. This method 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 [11].

3) Order Preserving Encryption

By encrypting sensitive data, the problem of outsourced data security will be solved. For that OPE is used. 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. Cloud computing architecture is used to improve the security, it provide a method in the organize that allow interaction in the server, by using that method decryption of private data is avoided. A client use order preserving encryption (OPE) algorithm [1] to encrypt the information. It is a set of data is securely transformed so that the order is preserved but the distribution and domain are changed. The common application of the method is a browser-based webmail application. In this application client receive email from one otherwise more servers and store it in the received email and it has been associated with OPE data[4]. This data is stored on a separate server. But this server cannot be used to send or else receive email. Advantage of that method is indexing/searching on OPE encrypted data. But the problem is once the original distribution is known, OPE is broken.

Attacker Modeling: The purpose of attack is to recover (or estimate) the original data from the perturbed data, or identify the correct queries (i.e., location queries) to break users' privacy. According to the level of previous knowledge the attacker may have, so here categorize the attacks into two categories.

- **Level 1**: The attacker knows simply the perturbed data and transformed queries, lacking any other prior knowledge. This is simply the cipertext-only attack in the cryptographic setting.
- **Level 2**: The attacker as well knows the original data distributions, including individual attribute distributions and the joint distribution (e.g., the covariance matrix) among attributes. In put into practice, for some applications, whose statistics are interesting to the public domain, the dimensional distributions might have been published through other sources.

Security Definition: Unlike from the traditional encryption schemes, attackers can also be satisfied with superior estimation. So, here investigate two levels of security definitions:

(1) It is computationally difficult for the attacker to recover the accurate original data based on the perturbed data,

(2) The attacker cannot effectively estimate the original information.

III. PROPOSED SYSTEM

RAndom Space Perturbation (RASP) approach is used to constructing practical range query and it also used for to construct k-nearest-neighbor query services in the cloud which is known as kNN query services. This approach addresses four features of the CPEL criteria with aim to accomplish a good balance on them. The RASP perturbation is a technique in which range queries are securely transformed into the RASP-perturbed data space. These range queries can be efficiently processed. That support indexing structures in the perturbed space. To increase the confidentiality and efficiency of query services here use kNN-R technique is used. It is combination of RASP range query services and kNN query.

A. *SYSTEM ARCHITECTURE*

The main purpose of Cloud computing infrastructures is to store large datasets and query services. It contain large amount of datasets. The main function of this architecture is to expand the proprietary database servers to the public

cloud, otherwise use a hybrid private-public cloud to obtain scalability and decrease costs even as maintaining confidentiality.

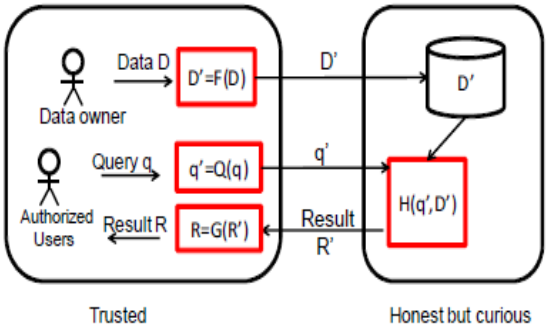

Figure 1. The system architecture for RASP-based query services

Above Figure 1 shows the basic system architecture for RASP-based query services in cloud. This architecture mainly contains two parts: one is trusted parties and another is untrusted parties. Trusted parties contain data owner or service owner and authorized user and untrusted parties contain cloud service provider who are honest but curious. They host the query services and protect database.

Following basic procedure follow this framework:

- F (D) indicates the RASP perturbation. It transforms the original data D into the perturbed data D′
- Q (q) indicate requested query. It first transforms the original query q into the protected form q′ which processed on the perturbed data.
- H (q′,D′) indicate query processing algorithm. This algorithm returns the result R′. When the statistics like SUM or AVG of a particular dimension are necessary, RASP can work through partial homomorphism encryption such as Paillier encryption to estimate these statistics on the encrypted data. After that this can be data recovered by the process G(R′).

B. RASP: RANDOM SPACE PERTURBATION

RASP contains several important features. One of these is RASP does not preserve the order of dimensional values because of the matrix multiplication component, which differentiate itself from order preserving encryption (OPE) schemes. So it does not go through the distribution-based attack. Another is, RASP does not maintain the distances between records, which avoid the perturbed data from distance-based attacks. Last is, the original range queries can be transformed to the RASP perturbed data space. It is the beginning of our query processing approach.

By using RASP perturbation method, two types of queries are constructed: range query and kNN query.

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 id
FROM table name
        WHERE id (
                SELECT top 15*
                FROM India
                WHERE age >60
               );
```

The above example shows the sample query for range query. Here the example query is to retrieve the entries from India it will retrieve the persons who are above 60 years in the top 15 list from the record of India.

C. KNN QUERY

kNN query denotes k-Nearest Neighbor query. It basically 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.

IV. 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.

REFERENCES

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

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

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

[5] K. Chen, R. Kavuluru, and S. Guo, "Rasp: Efficient multidimensional range query on attack-resilient encrypted databases," in *ACM Conference on Data and Application Security and Privacy*, 2011, pp. 249–260.

[6] K. Chen and L. Liu, "Geometric data perturbation for outsourced data mining," *Knowledge and Information Systems*, 2011.

[7] 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.

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

[9] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security*. New York, NY, USA: ACM, 2006, pp. 79–88.

[10] S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbor search with strong location privacy," in *Proceedings of VeryLarge Databases Conference (VLDB)*, 2010.

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