



## RASP: CONFIDENTIAL AND EFFICIENT HOSTING QUERY SERVICES IN PUBLIC CLOUD

<sup>#1</sup>R.PHANIKUMAR, M.Tech Student,

<sup>#2</sup>RAVI MATHEY, Professor & HOD,

Dept of CSE,

VIDYA JYOTHI INSTITUTE OF TECHNOLOGY, HYDERABAD, TELANGANA, INDIA.

**Abstract:** Cloud computing infrastructures are popularly used by peoples now a day. By using cloud users can save their cost for query services. But some of the data owners are hesitate to put their data's in cloud because, sometimes the data may be hack when they use in cloud unless the confidentiality of data and secure query processing will be provided by the cloud provider. In cloud if the user can get secured query service then the efficiency of query processing will be increased and the workload of the query processing will also be saved. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. The 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 RASP data perturbation method combines order preserving encryption, dimensionality expansion, random noise injection, and random projection, to provide strong resilience to attacks on the perturbed data and queries. It also preserves multidimensional ranges, which allows existing indexing techniques to be applied to speedup range query processing. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the kNN queries.

**Keywords :** *query services in the cloud, privacy, range query, kNN query.*

### I. INTRODUCTION

With the wide deployment of public cloud computing infrastructures, using clouds to host data query services has become an appealing solution for the advantages on scalability and cost-saving. 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. While new approaches are needed to preserve data confidentiality and query privacy, the efficiency of query services and the benefits of using the clouds should also be preserved. It will not be meaningful to provide slow query services as a result of security and privacy assurance. It is also not practical for the data owner to use a significant amount of in-house resources, because the purpose of using cloud resources is to reduce the need of maintaining scalable in-house infrastructures. Therefore, there is an intricate relationship among the data confidentiality, query privacy, the quality of service, and the economics of using the cloud.[1]

Here we summarize these requirements for constructing a practical query service in the cloud as the CPEL criteria: data confidentiality, query privacy, efficient query processing, and low in-house processing cost. Satisfying these requirements will dramatically increase the complexity of

constructing query services in the cloud. Some related approaches have been developed to address some aspects of the problem. However, they do not satisfactorily address all of these aspects. For example, the cryptindex and order preserving encryption (OPE) are vulnerable to the attacks. The enhanced cryptindex approach puts heavy burden on the in-house infrastructure to improve the security and privacy. The New Casper approach uses cloaking boxes to protect data objects and queries, which affects the efficiency of query processing and the inhouse workload. We propose the random space perturbation (RASP) approach to constructing practical range query and k-nearest- neighbor (kNN) query services in the cloud. The proposed approach will address all the four aspects of the CPEL criteria and aim to achieve a good balance on them. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries.[1]

The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee. We have carefully evaluated our approach with synthetic and real data sets. The results show its unique advantages on all aspects of the CPEL criteria.

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, with



indexing and efficient query processing. The range query is used in database for retrieving the stored data's. It will retrieve the records from the database where it can denotes 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. The RASP perturbation embeds the multidimensional data into a secret higher dimensional space, enhanced with random noise addition to protect the confidentiality of data.[2]

Hosting data-intensive query services in the cloud is increasingly popular because of the unique advantages in scalability and cost-saving by means of the cloud infrastructures the service owners can expediently scale up or down the service and only pay for the hours of using the servers. This is a smart feature because the workloads of query services are highly dynamic and it will be expensive and inefficient to serve such dynamic workloads with in-house infrastructures. However because the service providers lose the control over the data in the cloud data confidentiality and query privacy have become the major concerns.

Adversary such as curious service providers can possibly make a copy of the database or eaves drop users' queries which will be difficult to detect and prevent in the cloud infrastructures as new approach are needed to protect data confidentiality and query privacy the efficiency of query services and the benefits of using the clouds should also be preserved. It will not be significant. Let T be a table and  $X_i$ ,  $X_j$  and  $X_k$  be the real valued attributes in T and a and b be a number of constants. Take the including query for example. A typical range query looks like *select count(\*) from T where  $X_i \in [a_i, b_i]$  and  $X_j \in [a_j, b_j]$  and  $X_k = a$* , which calculates the number of records in the range defined by conditions on  $X_i$ ,  $X_j$ , and  $X_k$ . We summarize these needs for constructing a sensible query service in the cloud as the CPEL criteria: data confidentiality query Privacy Efficient query processing and Low in house processing cost. Agreeable these needs will radically increase the complexity of constructing query services in the cloud. Some connected approaches have been developed to address some aspects of the problem.

However they do not satisfactorily address all of these aspects. Future model for crypto index and Order Preserving Encryption (OPE) are vulnerable to the attack. The improved crypto- index approach puts serious load on the in-house infrastructure to improve the safety and privacy. The New Casper move towards uses cloaking boxes to protect data objects and queries which affects the competence of query processing and the in-house workload. We have concise the

weakness of the existing approaches. We propose the Random Space Perturbation (RASP) approach to constructing practical range query and k-nearest-neighbor (kNN) query services.

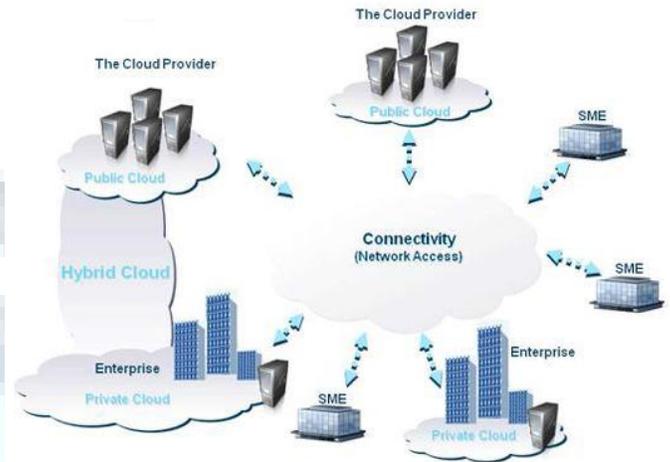


Fig.1. Comparison of Cloud Architecture

## II. RELATED WORK

We review the some most related methods like OPE, crypto-index, DRE, and PIR.

**Order Preserving Encryption:** The order preserving encryption (OPE) preserves the dimensional value order after encryption. Thus, it can be used in most database operations, such as indexing and range query. OPE represents Order Preserving Encryption is used for data that allows any comparison. And that comparison will be applied for the encrypted data; this will be done without decryption. It allows database indexes to be built over an encryption table. The drawback of this process is the encryption key is too large and implementation makes the time and space overhead.

**Cryptoindex:** Cryptoindex is also based on column-wise bucketization. It assigns a random ID to each bucket; the values in the bucket are replaced with the bucket ID to generate the auxiliary data for indexing. To utilize the index for query processing, a normal range query condition has to be transformed to a set-based query on the bucket IDs. Crypto index method is vulnerable to attacks but the working system of the crypto index has many difficult processes to provide the secured encryption and security and also the New Casper approach is used to protect data and query but the efficiency of the query process will be affect. For example,  $X_i < a_i$  might be replaced with . If the attacker manages to know the mapping between the input original query and the output bucket-based query, the range that a bucket ID represents could be



estimated. The width of the bucket determines how precise the estimation could be done. A bucket-diffusion scheme was proposed to address this problem, which, however, has to sacrifice the precision of query results. Another drawback of this method is that the client, not the server, has to filter out the query result. Low precision results raise large burden on the network and the client system. Furthermore, due to the randomized bucket IDs, the index built on bucket IDs is not so efficient for processing range queries as the index on OPE encrypted data is.

**Distance-recoverable encryption:** DRE is the most intuitive method for preserving the nearest neighbor relationship. Because of the exactly preserved distances, many attacks can be applied. Here, dot products are used instead of distances to find kNN, which is more resilient to distance targeted attacks. One drawback is the search algorithm is limited to linear scan and no indexing method can be applied.

**Private information retrieval(PIR) :** PIR tries to fully preserve the privacy of access pattern, while the data may not be encrypted. PIR schemes are normally very costly. This privacy preserving multi keyword search is based on the plain text search. In this the searching process will done by ranking process. The drawback of this concept is because of ranking process in house processing time will be maximized. The research on privacy preserving data mining has multiplicative perturbation methods, which are similar to the RASP encryption, but with more emphasis on preserving the utility for data mining.

### III. METHODOLOGY

#### A. System Architecture

We assume that a cloud computing infrastructure, such as Amazon EC2, is used to host the query services and large data sets. The purpose of this architecture is to extend the proprietary database servers to the public cloud, or use a hybrid private-public cloud to achieve scalability and reduce costs while maintaining confidentiality.

Each record  $x$  in the outsourced database contains two parts: the RASP-processed attributes  $D'=F(D,K)$  and the encrypted original records,  $Z = E(D,K')$ , where  $K$  and  $K'$  are keys for perturbation and encryption, respectively. The RASP-perturbed data  $D'$  are for indexing and query processing. Fig. 2 shows the system architecture for both RASP-based range query service and kNN service.

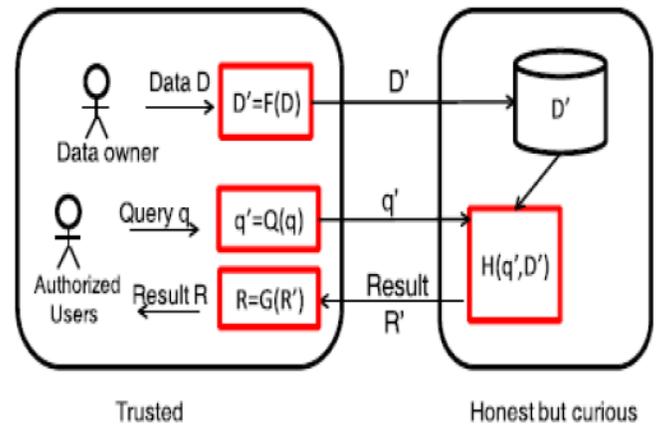


Fig. 2. The system architecture for RASP-based query services.

There are two clearly separated groups: the trusted parties and the untrusted parties. The trusted parties include the data/service owner, the in-house proxy server, and the authorized users who can only submit queries. The data owner exports the perturbed data to the cloud. Meanwhile, the authorized users can submit range queries or kNN queries to learn statistics or find some records. The untrusted parties include the curious cloud provider who hosts the query services and the protected database. The RASP-perturbed data will be used to build indices to support query processing.

There are a number of basic procedures in this framework: 1)  $F(D)$  is the RASP perturbation that transforms the original data  $D$  to the perturbed data  $D'$ ; 2)  $Q(q)$  transforms the original query  $q$  to the protected form  $q'$  that can be processed on the perturbed data; and 3)  $H(q',D')$  is the query processing algorithm that returns the result  $R'$ . When the statistics such as SUM or AVG of a specific dimension are needed, RASP can work with partial homomorphic encryption such as Paillier encryption [24] to compute these statistics on the encrypted data, which are then recovered with the procedure  $G'(R')$ .

#### B. Threat Model

The cloud server is considered as “honest-but-curious” in our model, which is consistent with related works on cloud security. Specifically, the cloud server acts in an “honest” fashion and correctly follows the designated protocol specification. However, it is “curious” to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information. Assumptions: Our security analysis is built on the important features of the architecture. Under this setting, we believe the following assumptions are appropriate:

- Only the authorized users can query the proprietary database. Authorized users are not malicious and will



not intentionally breach the confidentiality. We consider insider attacks are orthogonal to our research; thus, we can exclude the situation that the authorized users collude with the untrusted cloud providers to leak additional information.

- The client-side system and the communication channels are properly secured and no protected data records and queries can be leaked.
- Adversaries can see the perturbed database, the transformed queries, the whole query processing procedure, the access patterns, and understand the same query returns the same set of results, but nothing else.
- Adversaries can possibly have the global information of the database, such as the applications of the database, the attribute domains, and possibly the attribute distributions, via other published sources (e.g., the distribution of sales, or patient diseases, in public reports).

**Protected assets:** Data confidentiality and query privacy should be protected in the RASP approach. While the integrity of query services is also an important issue, it is orthogonal to our study. Existing integrity checking and preventing techniques [33], [29], [18] can be integrated into our framework. Thus, the integrity problem will be excluded from the paper, and we can assume the curious cloud provider is interested in the data and queries, but it will honestly follow the protocol to provide the infrastructure service. Attacker modeling. The goal of attack is to recover (or estimate) the original data from the perturbed data, or identify the exact queries (i.e., location queries) to breach users' privacy. According to the level of prior knowledge the attacker may have, we categorize the attacks into two categories:

- Level 1: The attacker knows only the perturbed data and transformed queries, without any other prior knowledge. This corresponds to the ciphertext-only attack in the cryptographic setting.
- Level 2: The attacker also knows the original data distributions, including individual attribute distributions and the joint distribution (e.g., the covariance matrix) between attributes. In practice, for some applications, whose statistics are interesting to the public domain, the dimensional distributions might have been published via other sources.

## IV. PROPOSED METHOD

The proposed approach has a number of unique contributions. It includes the RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee; the RASP approach preserves the topology of multidimensional range in secure transformation, which allows indexing and efficiently query processing; the proposed service constructions are able to minimize the in-house processing workload because of the low perturbation cost and high precision query results; this is an important feature enabling practical cloud-based solutions.

The project modules are stated below:

- A. User authentication module.
- B. RASP: RANdom Space Perturbation
- C. RASP Range-Query Processing
- D. kNN Query Processing with RASP

The project modules are explained below:

### A. User Authentication Module

In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

### B. RASP: Random Space Perturbation

The data owners have to register their self. The data owner exports the perturbed data to the cloud. RASP is one type of multiplicative perturbation, with a novel combination of OPE, dimension expansion, random noise injection, and random projection. Let's consider the multidimensional data are numeric and in multidimensional vector space. The database has  $k$  searchable dimensions and  $n$  records, which makes a  $d \times n$  matrix  $X$ . The *searchable* dimensions can be used in queries and thus should be indexed. Let  $x$  represent a  $d$ -dimensional record,  $x \in R^d$ . Note that in the  $d$ -dimensional vector space  $R^d$ , the range query conditions are represented as half-space functions and a range query is translated to finding the point set in corresponding polyhedron area described by the half spaces.

### C. RASP Range-Query Processing

Range query in the original space can be transformed to a polyhedron query in the perturbed space, and then develop a secure way to do the query transformation. The users have to register their self and obtain keys from the data owners in order to access data. The authorized users who can only submit queries on RASP-perturbed data. The authorized users



can submit range queries or kNN queries to learn statistics or find some records.

#### D. kNN Query Processing with RASP

kNN query cannot be directly processed with the RASP perturbed data. For that have to design a kNN query processing algorithm based on range queries (the kNN-R algorithm). As a result, the use of index in range query processing also enables fast processing of kNN queries.

The kNN-R algorithm consists of two rounds of interactions between the client and the server.

- i. The client will send the initial upper-bound range, which contains more than k points, and the initial lower-bound range, which contains less than k points, to the server. The server finds the inner range and returns to the client.
- ii. 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.
- iii. The client decrypts the records and finds the top k candidates as the final result.

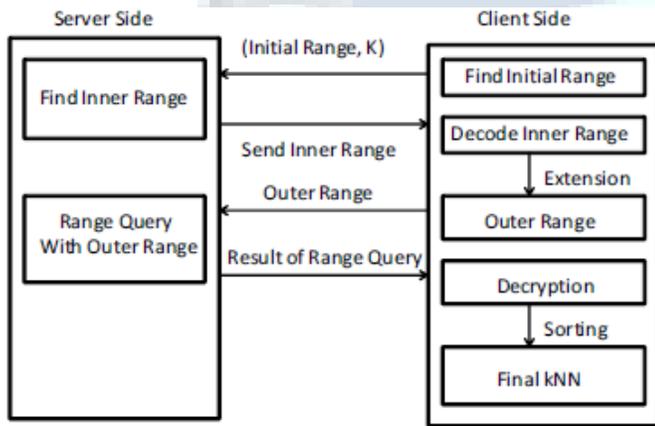


Figure. Procedure for KNN-R algorithm

### V.CONCLUSION

We propose to study an outsourced service based on the CPEL criteria: data Confidentiality, query Privacy, Efficient query processing, and Low inhouse workload. With the CPEL criteria in mind, we develop the kNN-R approach for secure outsourced kNN query service. The kNN-R approach takes advantage of fast and secure RASP range query processing to implement kNN query processing. It can find high precision kNN results and also minimize the interactions between the cloud server and the inhouse client. High precision kNN results and minimized interactions result in low inhouse workload. We have conducted a thorough

security analysis on data confidentiality and query privacy. Compared to the related approaches, the kNN-R approach achieves a better balance over the CPEL criteria.

RASP method with range query and kNN query. 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] 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 Berkeley, 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.
- [13] Atul Phad, Swapnil Patil, Sujeet Purane, Vineet Patil,—Cloud Based SQL Query Processor, in *International Journal Of Engineering And Science*, Issn: 2278-4721, Vol. 2, Issue 4 (February 2013), Pp 01-04.
- [14] 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.
- [15] M. L. Liu, G. Ghinita, C. S. Jensen and P. Kalnis, —Enabling search services on outsourced private spatial data, *The International Journal of on Very Large Data Base*, vol. 19, no. 3, 2010.
- [16] Jing Zhao, Xiangmei Hu and Xiaofeng Meng, —ESQP: An Efficient SQL Query Processing for Cloud Data Management, in *Proceedings of the second international workshop on Cloud data management*. New York, NY, USA: ACM 978-1-4503-0380-4/10/10, Pages 1-8.
- [17] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. and Andy 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 Berkeley, 2009.
- [18] K. Chen, L. Liu, and G. Sun, —Towards attack-resilient geometric data perturbation, in *SIAM Data Mining Conference*, 2007.
- [19] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, —Order preserving encryption for numeric data, in *Proceedings of ACM SIGMOD Conference*, 2004.