iJRASET
International Journal For Research in
Applied Science and Engineering Technology

Volume: 3 Issue: IV Month of publication: April 2015

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Confidential and Efficient Query Service with RASP Data Perturbation in Cloud

Megaya Jesupillai.M[1], Appandairaj.A[2]

[1] M.E Student, Ganadipathy Tulsi's Jain Engineering College, Vellore Dist, T.N, India

[2] Assistant Professor, Ganadipathy Tulsi's Jain Engineering College, Vellore Dist, T.N, India

Abstract **-Todays, people groups are widely utilized cloud computing foundations and infrastructure. So the end-user can save their expense and time by utilizing query services as a part of cloud. At the same time infrequently data holder does not move to cloud, because information may be hack from the malignant user when they use in cloud if not the confidentiality data furthermore security of a query should be guaranteed. In 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.**

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

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. 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. The kNN query means k-Nearest Neighbor query. K means positive whole number and this query are utilized to discover the estimation of closest neighbor to k.

II. QUERY SERVICES IN CLOUD

Query is primarily used for searching purpose in web world. Queries are developed by using structured query language. It mainly retrieving the required data from the database according to the user need. Query services are the technique for internet services that are exposed through an implementation of service supplier. RASP technique and Knn Query processing provide secure data by encryption and decryption, fast storing and recovering process in database [7].

III. 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. The system architecture shown in below figure

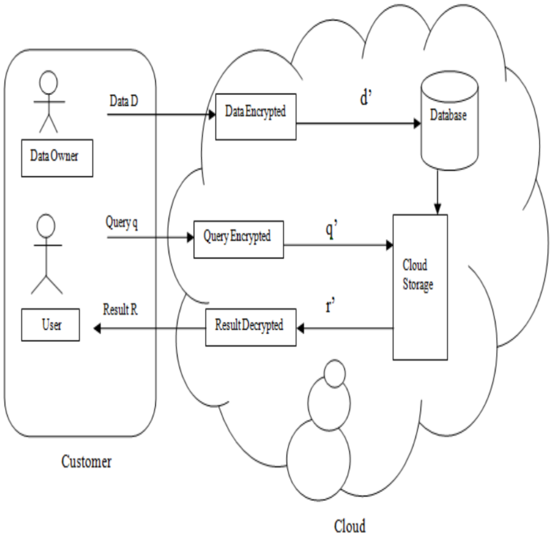

Fig. 1 System Architecture for RASP method

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.

A. Security Analysis

The security analysis of the system design discuss below,

1) Just authorized users have a key which gave by the data owner. So an authoritative user is not a malevolent and won't deliberately break the confidentiality. So only singular's users can send the queries for recovering the data.

2) The communication technique among the end user, data owner and cloud service system are decently secured and ensure no records and queries can be spilled from cloud.

3) RASP methodology is used to provide the security of the query privacy and confidentiality of the owner's data.

B. Attacker Objective

The significant goal of hacker or attacker is to hack the data from the database or identify the queries (for instance, location queries) and break the protection and privacy of the data.

C. Attacker classification

The attacker can be classified into two categories based on their level of preceding knowledge,

A. *Level 1:* The attacker knows just the perturbed data and transformed queries, with no other prior knowledge. This compares to the cipertext-just assault in the cryptographic setting.
B. *Level 2:* The attacker additionally knows the original data distributions, moreover individual trait distributions and the joint distributions among qualities or attributes. Set up into practice for a few applications, whose insights are lovely to the indefinite domain, the dimensional distributions may have been published through new sources.

IV. MODULES

Three modules are used in our system, such RASP, range query and kNN query.

A. RASP

RASP indicates RAndom Space Perturbation. The RASP data perturbation strategy is blend of OPE, random noise injection, and random projection, to give solid versatility to attack on the perturbed data and additionally queries. It additionally monitors multidimensional ranges, which permit introduced indexing techniques to be connected to speed up range query processing. OPE indicate Order Preserving Encryption is a method for encrypting data so that it’s conceivable to make proficient disparity correlations on the encrypted things without decrypting them. Random projections are an effective way for dimensionality reduction. Random projection is a methodology of projecting unique high-dimensional data onto a lower dimensional data representation.

Random noise injection is generally used to adding noise to the input data to acquire proper output when we compare it with the estimated power. The RASP system and its mix give privacy of data and this approach is for the most part used to secure the multidimensional range of queries in secure mode furthermore with indexing efficient query processing will be carried out. RASP has some vital features. In RASP the utilization of matrix multiplication does not secure the dimensional values so no need to go through from the distribution based attack.

RASP does not preserve the distances between records, so it keeps the data that are perturbed from distance based attacks [8]. Furthermore it won't secure more troublesome structures it might be a matrix and different parts. The range queries can be send to the RASP perturbed data and this range query illustrate open limits in the multidimensional space. In Random space perturbation, the perturbation is utilized to do giving way this process will occur as indicated by the key value that is determined by the data owner. In this module the data owner need to enlist like owner and need to give owner name and also key value. And afterward the users have register and get 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 get their answer. We analyze and demonstrate the outcome with encrypted furthermore in decrypted format of the data for the query build by the user.

In Random space perturbation, the perturbation is utilized to do crumpling this process will occur as per the key value that is indicated by the owner. In this module the data owner need to register like owner and need to give owner name and key value. And afterward the users have register and get 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 get their answer. We analyze and demonstrate the outcome with encrypted and also in decrypted format of the data for the query build by the user.

B. Range Query

Range query is the normal database operation. It retrieves the data value from the database that values are in between upper bound & lower bound. The range query is not regular in light of the fact that user won't know ahead of time about the outcome for the query, what number of 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 code demonstrate one more sample of range query search it will give the passages of what are employee id that are store in EMP database with age over 40 and inside 60. So by utilizing range query user can essentially retrieve the data from records and this query process will be finished in secure way and additionally the speed of the query process will also increased.

C. kNN Query

kNN query stands k-Nearest Neighbor query. This type of query is usually used to recover the closest neighbor values of k. Here k is used to show positive number value. kNN calculation is primarily used for grouping and regression. The main use of kNN-R calculation is to process the range query to kNN query. This calculation comprises of two methods. This method is used to make communication between the end-user and the server. The end-user will send the query to the server with starting upper bound and lower bound limits. This upper bound range must be more than the k focuses and the lower bound range must be not exactly the k points. The below figure demonstrates the whole process of k- nearest neighbor query.

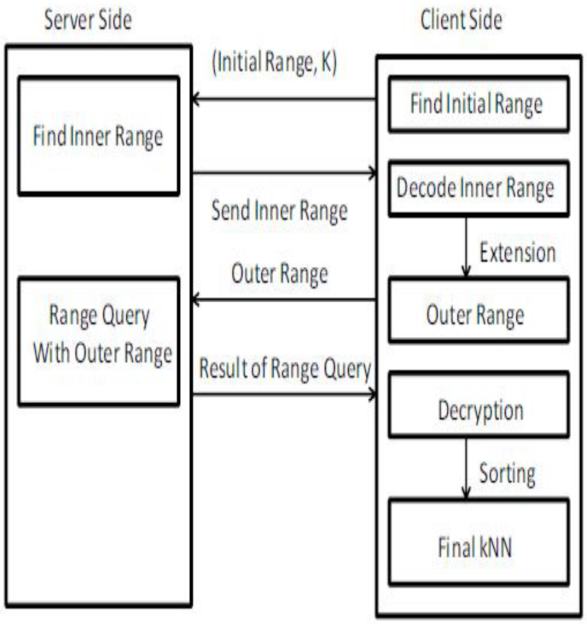

Fig.2 kNN query process

The above query process is used to give the inner range of the database by the server. With this internal query range the user will compute the outer range and send this outer range to the server. After that the server will search and discover the records in the outer range from the database and send it to customer and afterward the customer will decrypt the record and get the top k files to give the final result. This calculation is used to find the reduced inner square range for give high accuracy and it has two not simple courses of action in it.

They find many points that are present in the square range and update of the limit (i.e.) upper bound and lower bound is complex because t range queries are secured by using random space perturbation. The security of kNN query and range question is comparable.

The figure 2 shows entire query process involved in kNN query process. In server site they have to find inner range and result with outer range of query. In client side First, they find initial range and decode as inner range, again they extension as outer range. After finding outer range then decryption and soaring result has been done for getting final kNN.

V. STUDY OF EXISTING PROCESS

In this section we summarized the exiting process said in different studies.

A. OPE

OPE be a symbol of Order Preserving Encryption [1]. It is used for data which allows comparison between data, this comparison use to the encrypted original data; but not does decryption. This is used to build database indexing on encrypted data table. the downside of this process, having very large encryption key and this may increase time and space required for implementation.

B. Crypto-Index

This approach is used for giving security and confidentiality of data within cloud database and cloud environment. But it is more prone to the attack. The crypto-index approach has a lot of difficulties on providing secure encryption and maintains privacy.

C. Preserving Query privacy

In this technique multi keyword search is promoted based on the text search. In this the searching method mainly by ranking process. The drawback of ranking process is maximizing the processing time.

New Casper approach: In this method, clock boxes used for processing query. It affects efficiency of query processing and in-house workload.[8]Furthermore we had study about RASP technique, query security, query privacy, empowering search services benefits on out sourced information and many concepts.

VI. CONCLUTION

We have studied few methods that are used to give a security to data in the cloud. We proposed RASP method with range query and kNN query, this is mainly used to perturb the data given by the data owner and saved in cloud storage and also it combines random injection, order preserving encryption and random noise projection and also it has contains CPEL criteria in it. To satisfy the prerequisite on low in house workload, cloud computing give quality query services which is more efficient and extremely secure. This system fundamentally used to perturb the data given by the data holder and saved in distributed storage. By range query and kNN query end user can retrieve their required data in secured way and also processing time 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.