International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 01
January 2015

An efficient Query Services in the Cloud Using kNN-R and RASP Data Perturbation

[1]Sivaprasad Guntakala, [2]Mohanbabu Choragudi
[1]Assistant Professor, Department of MCA, Narasaraopeta Engineering College, Narasaraopeta, Guntur Dist., AP. India
[2]Associate Professor,Department Of Computer Science and Engineering,Lingayas Institute of Management and Technology,Vijayawada,AP, India

Abstract: Now a day's cloud is more popular because in cloud users host the data and upload a large contained data. It has large databases to database service providers so database service providers maintain the services of range query services. In clouding process some users have a sensitive private data in that situation users can't move the data for hosting until we provide security, confidentiality, perfectness, query privacy are guaranteed to the hosted data. 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.

Key Words: query services in the cloud, kNN query, range query.

I. INTRODUCTION

In public cloud infrastructure is wide range of deployment using host data query services in host, data query services has become an best solution for the advantages on cost saving and efficiency. In cloud infrastructure the service owner can flexible rise up that is scale up or decrease down the services, user can only pay the service for servers based on hourly works. New methods are required for to protect the data query privacy and confidentiality for the data and query privacy the query service efficiency and the advantages of using cloud should be under secure. It will be not effective meaningful provide slow query services as a result output of as a security and assurance of privacy. 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. Now here constructing query service in CPEL criteria: data confidentiality that is secure data, in house processing query processing and query privacy in low in house processing, full fill these all requirements will it helps increase the complexity of developing services of query in cloud. Some methods are related have been constructing address some aspects of the problem. In that may be chance to don't full feel address of these all aspects. Now discuss for example crypto index and order preserving encryption (OPE) are not honorable attacks. In both of the encryptions methods is heavy burden on in house infrastructure that is improve the security and privacy. In this project we proposed a RASP random space perturbation method to construct improve the practical range queries kNN k-Nearest neighbor query process in cloud. The RSAP approach will satisfy all the four aspects those are data confidentiality, secure data in house query processing, balanced these aspects and The RASP query service uses with the kNN query services.

The Random space approach is unique combination of order preserving encryption expansion of dimensionality, not a OPE, expansion dimensionality and also random projection and random noise injection. This gives a more confidential security for the data to be provided guarantee. We have to find out our proposed approach RASP with synthetic and real data sets. In this process the result shows the best and unique advantage of CPCL aspects those are data confidentiality, query processing and query privacy, in house processing query. In our proposed approach RASP it is combination of data confidentiality and query process and it mainly help for protected the multidimensional range

queries in secure cloud manner, with efficient query processing and indexing. The range query data base queries help to retrieve the data from databases; it will retrieve records based on queries with conditions based on some boundaries between like upper and lower boundaries. The kNN query denotes k nearest neighbor query here k means a positive integer value nearest value of the positive integer of k. The RASP perturbation add multi-dimensional data into secret place that is secret higher dimensional space and make a more secure with random noise addition to protect the confidential of the data.

II. RELATED WORKS

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, $Xi < ai$ 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. PROPOSED METHOD

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.

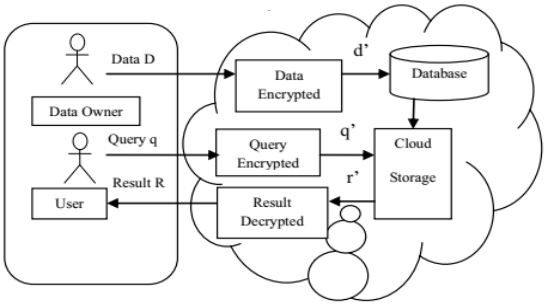

Fig. 1 System Architecture for RASP method.

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.

Five modules are used. They are user interface design, range query processing, kNN query processing, server protecting data, and 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 protects 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 protects 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
1: **RASP Perturb**(X,RNG,RIMG,Ko)
2: Input: X: k × 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
3: A ← 0;
4: A3 ← the last column of A;
5: v0 ← 4;
6: **while** A3 contains zero **do**
7: generate A with RIMG;
8: **end while**
9: **for** each record x in X **do**
10: v ← v0 - 1;
11: **while** v < v0 **do**
12: v ← RNG;
13: **end while**
14: y ← A((Eope(x, Kope))T , 1, v)T ;
15: submit y to the server;
16: **end for**
17: return A;

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

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. CONCLUSION

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] 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 It's 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.