International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

Construction of Private Methodical Query Services in the Cloud with RASP Data Commotion

Mudduluru Avinash[1]& Ch.V.Swarna Kumari[2]

[1]B-Tech, Department: CSE, Narayana Engineering College, Nellore. Email Id: fromavi@gmail.com

[2]Associate Professor, Department: CSE, Narayana Engineering College, Nellore.
Email Id: swarnachinni@gmail.com

ABSTRACT

HOSTING input/data in the cloud is an alluring approach because it provides advantages like scalability and cost-saving. The service owner can increase or decrease the service of cloud infrastructure and pay as per uses. Due to its work the cloud features is very much attractive. It is highly cost and wasteful to serve such forceful workloads with in-house infrastructures. The data owners always desire to submit their quires after realizing the privacy assurance of the cloud. In this aspect researchers have introduced few techniques such as RASP, kNN Algorithm etc. Many a times, data into cloud is stored by maintaining confidentiality, query privacy, efficient query processing at low cost (CPEL Criteria). However, the data owners always desire to submit their quires after realizing the privacy assurance of the cloud. In this aspect researchers have introduced few techniques such as RASP (Random Space Perturbation), k-NN (k-Nearest Neighbour) Algorithm etc. The main problem across RASP technique is, generating the encryption key which is too large and its implementation makes the time and space overhead. Grate information bother technique along with k-NN algorithm is exploited to furnish privacy to the cloud. Wherein, issues such as categorical data and leaked query in the model are identified and addressed, by holding no change in designing the k-NN-R algorithm.

1. 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.

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 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 cryptoindex and order preserving encryption (OPE) are vulnerable to the attacks. The enhanced cryptoindex 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.

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 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. 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. RELATED WORK

Importance of RASP

Distributed computing is a rising technology. Posting information serious question administrations in the cloud is progressively prominent as a result of the one of a kind points of interest in versatility and expense sparing. But data privacy and confidentiality has become the major concerns. In summary, the prospective path has a number of individual contributions: The RASP perturbation technique is a unique combination of order preserving encryption, dimensionality expansion, random noise injection, and random projection, which provides the guarantee of strong confidentiality. The proposed approaches are able to reduce the processing workload because of the low commotion cost and high rigor query results. This is an important feature enabling practical cloud-based solutions.

RASP using kNN

The proposed methodology RASP with kNN will address all the four angles to keep up the CPEL criteria and mean to accomplish a decent adjust on them. The primary concern is to decide multidimensional information sets with a blend of request safeguarding encryption, dimensionality development, arbitrary commotion infusion, and irregular task. Scratch does not hold the separation between the records, and KNN question is utilized to decide the reach esteem. KNN calculation depends on the reach inquiries and uses the file in extent question handling and uses the file in reach inquiry preparing and subsequently quick handling of reach inquiries happens.

Processing of KNN algorithm

The main goal of KNN algorithm is to determine the KNN adjacent point in the spherical range that centred at the query point. It uses the spectrum value instead of spherical range. However it has to overcome the few issues such as whether the data privacy and query privacy are present, whether these are an increment in the service workload and to find the minimal square range that exactly contains the KNN-nearest points. The KNN algorithm consists of three rounds of interaction between service and the applicant.

1. First, the applicant will sent, the upper-bound range which contains greater value than K points and the lower- bound range which contains less than K-points to the server. The server finds the focal range and then returns to the applicant.
2. Second, the client finds the outer range depending upon the inner range and returns to the server. The server finds the outer range and then returns to the applicant.
3. Third, the client decrypts the records and finds the first K-point as the resulting server side.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

How kNN Works

The RASP perturbation technique is designed in such a way that the queried ranges are securely transformed into polyhedral in the commotion data space, which can be effectively prepared with the backing of indexing structures in the irritated space. Algorithm is a process or different types of formulas are used for calculations or other problem-solving operations, especially by a computer.

Details about K nearest neighbour algorithm:

K Nearest Neighbours' is a simple algorithm that used to store all available cases and predicts the actual target based on an affinity measure for e.g., distance function. kNN has been used to determine the nearestvalue. design acknowledgment which is as of now at the outset of 1970's as a non-parametric method. A straightforward usage of kNN relapse is to ascertain the normal of the numeric focus of the K closest neighbours. Another methodology utilizes a reverse separation weighted normal of the K closest neighbours. kNN relapse utilizes the same separation capacity as kNN characterization. For closest neighbour questions Knn calculation is utilized. Knn is a non-parametric strategy utilized for order and relapse. There will be n-training vectors; knn identifies k-nearest neighbour of the class regardless of labels.

Example for Knn algorithm

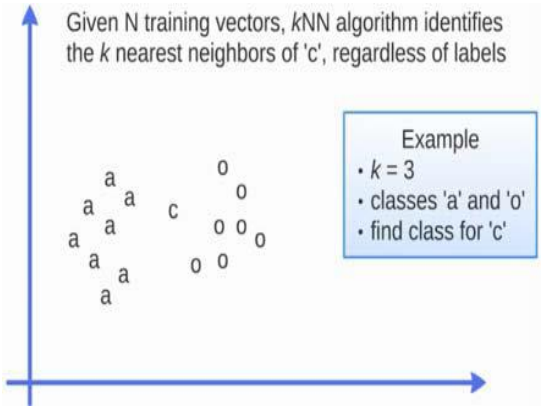

Fig 1: Initial State

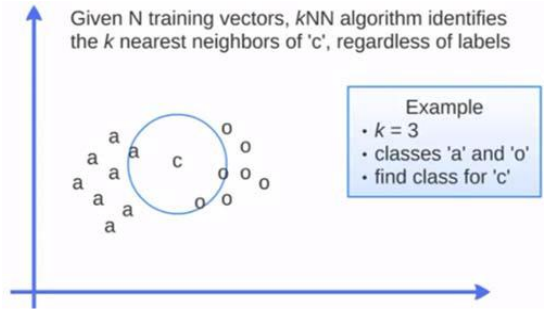

Fig 2: Identifying Nearest neighbors

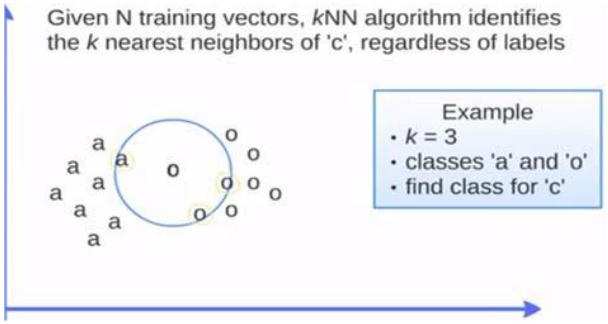

Fig 3: Assigning label to nearest Class

Two classes "a" and "o" are given with some elements, label "c" (fig 1) has to be assigned to the nearest neighbor class. In the next step,

nearest neighbor region has to find out based on K value ie, 3 (fig 2). The element "c" is assigned to the class based on the K elements as shown in fig (3). Apart from k nearest neighbor algorithm I will be using the following algorithms:

SHA1 - Secure Hash Algorithm for Digital Signature.

RSA - For Secret Key Generation.

AES - Cryptography technique (For Data Encryption and Decryption)

Query Service

Query is mainly used to search. Queries are constructed by using structured query language. It is mainly used to fetch the essential information from the database. 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 protected, fast storing and retrieving process of encryption and decryption of a data from database.

3. SYSTEM ARCHITECTURE

Cloud computing infrastructures used to store large datasets and query services. The architecture shows two main parts in it. The data's can be stored in the cloud by data owners d=n (d, k) here d represents data, n represents normal form of data, k represents key value given by the data owner. This format will be saved in the cloud as encrypted form D^1=e (D, k) here e represents encryption. When user request to cloud for any data, and while downloading the data, Decryption H, will performed with encrypted data D^1 along with key k. i.e., D=H(D^1, k);

All metadata stored in the cloud database are perturbated. Any application running on a legitimate client can transparently issue SQL operations (e.g., SELECT, INSERT, UPDATE and DELETE) to the cloud database through the perturbated database interface. Data transferred between the user application and the index engines are in plain format, whereas information is always encrypted before sending it to the cloud database. When an application issues a new SQL operation, the perturbated database interface contacts the perturbated engine that retrieves the metadata. In order to improve performance, the plain metadata are cached locally by the client as volatile information. After obtaining the metadata, the encryption engine is able to execute the SQL operation on encrypted data, and then to decrypt the results. The results are returned to the user application through the encrypted database interface. Do not reveal information about plain data, plain metadata.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

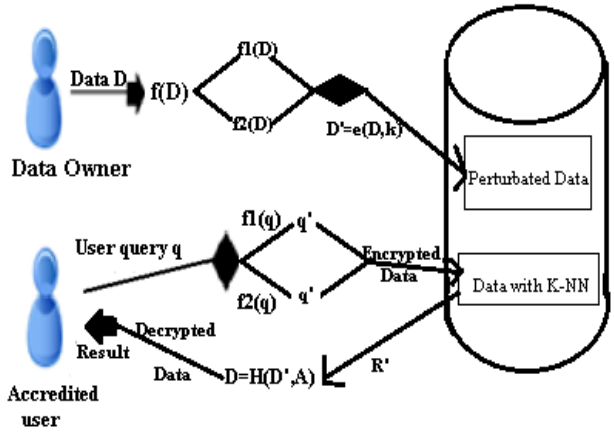

Fig. 4 System Architecture using RASP method

The above diagram shows two separate parties. They are customer who is the trusted party store their data in cloud and second are cloud provider who is storing the data in encrypted format. In the customer party it includes data and service owners, proxy server of in-house process and users. Here the owner can store their data in cloud while those data will encrypted in cloud and stored in the cloud database and also the data owner will provide key value by using that key value only cloud will encrypt the data by using random space perturbation method. The user will send query to retrieve the data from cloud, user can send range query and kNN query to get the data. In the cloud, the cloud providers have to host the user query services and have to protect the data stored in the cloud database. The basic procedure in the diagram is: (1) the owner sends the data to store in cloud that data will be encrypted by using random space perturbation method and stored in cloud database based on datatype (VARCHAR OR INT). (2) the user will send the range query or kNN query to retrieve the data that query will be encrypted and send to cloud storage. (3) the cloud storage will send the data for that query after processing the query inside cloud storage and it will be decrypted and finally the data will send to the user.

Security Analysis

The security analysis in the architecture shows the following i) Users have been authorized by using the key value provided by the owner. So an authorized user is not being a malicious and only those users can send the queries for retrieving the data. ii) The communication process between the user, owner and cloud and client system are well secured, the data and queries cannot be leaked from the cloud.

iii) RASP method is used to protect the query privacy and confidentiality of the data. Attacker Process: The main process of attacker is to hack the data from the database and they will try to find the perturbed data and they will try to find the queries.

4. MODULES

Three modules are used. They are RASP, range query and kNN query.

RASP:

RASP denotes Random Space Perturbation. It also combines OPE, random projection and random noise injection. Here OPE denotes 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. 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. In RASP the use of matrix multiplication does not protect the dimensional values so no need to suffer from the distribution based attack.

Range Query

Range query is the query used to retrieve the data from the database. It will retrieve the data value that is between the upper bound and lower bound. The range query is not usual because user won't know in advance about the result for the query, how much entries will come as result for the query. For example

```
SELECT id FROM table name
WHERE id (
        SELECT top 10*
        FROM United States
        WHERE age >50
        );
```

The above example shows the sample query for range query. Here the example query is to retrieve the entries from United States it will retrieve the persons who are above 50 years in the top 10 list from the record of United States. The range search is mainly used to return the values that are present between the two specified values given in the query. For example database name is AAAworkers2012 then Go

```
        SELECT product id
        FROM
AAAworkers2012.production
        WHERE price BETWEEN 40 and
60
```

The above example will show an another example of range query search it will provide the entries of what are product id that are present in production database with price above 40 and within 60. So by using range query user can easily retrieve the data's from records and this query process will be done in secure manner and

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

the speed of the query process will also increased.

kNN QUERY

kNN query represents k-Nearest Neighbor query. This query is mainly used to retrieve the nearest neighbor values of k. here k used to denote positive integer value. kNN algorithm is mainly used for classification and regression. In this it uses kNN-R algorithm 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. The above process is used to give 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. Then 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 find the top k files to provide the final result. This algorithm is used to find the compact inner square range for providing high precision and it has two difficult processes in it. They are to find the number of points that are present in the square range and updating 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 equal.

5. STUDY OF PROPOSED PROCESS

Problem: In Categorical Data, a set of data is sorted or divided into different categories, according to the attributes of the data.

Example: Number of user { set } ϵ Number of data

When there is a mixture of numerical and categorical variables present in the data set, Hamming Code Distance is used for standardization.

Manhattan\Hamming Code Distance

$$D_H = \sum_{i=1}^{K} | X_i - Y_i | \qquad \begin{matrix} x = y \Rightarrow D = 0 \\ x \neq y \Rightarrow D = 1 \end{matrix}$$

We describe the details of the adaptive layer mechanism by referring to an example. Let us consider a table T with columns id of type int and name of type string, and a tenant client preparing to issue the following statement to the encrypted cloud database: SELECT * FROM T WHERE id < 10. The client perturbation engine analyzes the SQL statement, and identifies that the operation id < 10 has to be pertubate on the

encrypted database. Then, the client reads the metadata and checks whether there is the INT attribute associated to the column id because this is the only INT data type.

For INT data type

Input: Select Data D;

Output: data D^1;

Let String S=null;

S=D;

for (char c: S)

c → byte b;

temp = b;

temp = temp-rand();

b=temp + b;

endfor;

D^1=rand(c) + b;

return D^1;

For varchar data type

Input: Select Data D;

Output: data D^1;

Let S=null;

b=0;

temp = Timestamp (Date and Time);

b=temp;

D^1=rand(char) + b;

return D^1;

Threat model: It is a process for optimizing network / application harmed the internet security by identifying objective and vulnerability a threat is a potential or actual undesirable event that may be malicious such as DOS attack or incidental failure of a storage device. Threat model is a panel activity for identifying and assessing application threats and vulnerabilities.

Leaked Query: It reveals that the application is vulnerable, then accidental leaking of sensitive information through data queries.

6. EXPERIMENTAL RESULT:

We do explore the effectiveness of the various perturbation techniques in detail here due to the extensive coverage of this facet of these perturbation techniques in our previous work. Nevertheless, we do highlight some general's trends with regard to search effectiveness.

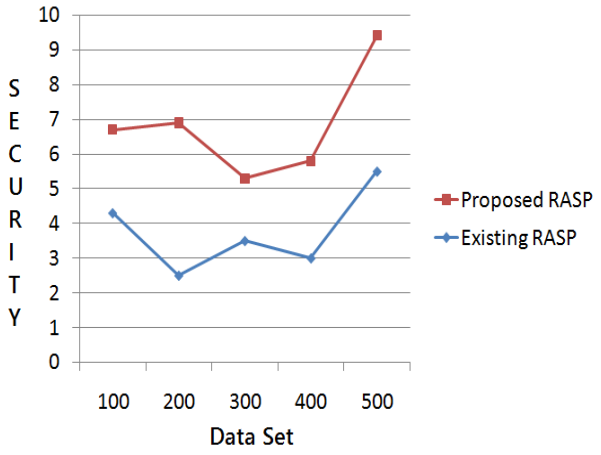

Fig: 5 Performance Graph

Fig. 5 graphs precision with two perturbation techniques. Graph is comparable to the number of datasets by a existing and proposed RASP;

this metric has been used in a number of previous evaluations. The figure reveals two interesting trends. First, the security levels of with existing RASP, and Second, the security levels of with Proposed RASP with two level of security.

7. CONCLUSIONS:

We proposed 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. And also we continue our studies to improve the effect of query.

8. REFERENCES

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599–616, 2009.

[2] T. Mather, S. Kumaraswamy, and S. Latif, Cloud security and privacy: an enterprise perspective on risks and compliance. O'Reilly Media, Incorporated, 2009.

[3] H.-L. Truong and S. Dustdar, "Composable cost estimation and monitoring for computational applications in cloud computing environments," Procedia Computer Science, vol. 1, no. 1, pp. 2175 – 2184, 2010, iCCS 2010.

[4] E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good, "The cost of doing science on the cloud: the montage example," in Proc. 2008 ACM/IEEE Conf. Supercomputing, ser. SC '08. Piscataway, NJ, USA: IEEE Press, 2008, pp. 50:1–50:12.

[5] H. Hacig ¨um¨us¸, B. Iyer, and S. Mehrotra, "Providing database as a service," in Proc. 18th IEEE Int'l Conf. Data Engineering, Feb. 2002.

[6] G. Wang, Q. Liu, and J. Wu, "Hierarchical attribute-based encryption for fine-grained access control in cloud storage services," in Proc. 17th ACM Conf. Computer and communications security. ACM, 2010, pp. 735–737.

[7] Google, "Google Cloud Platform Storage with server-side encryption," http://googlecloudplatform.blogspot.it/2013/08/google-cloud-storage-now-provides.html, Mar. 2014.

[8] H. Hacig ¨um¨us¸, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model," in Proc. ACM SIGMOD Int'l Conf. Management of data, June 2002.

[9] L. Ferretti, M. Colajanni, and M. Marchetti, "Distributed, concurrent, and independent access to encrypted cloud databases," IEEE Trans. Parallel and Distributed Systems, vol. 25, no. 2, Feb. 2014.

[10] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: protecting confidentiality with encrypted query processing," in Proc. 23rd ACM Symp. Operating Systems Principles, Oct. 2011.

[11] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st ACM Symp. Theory of computing, May 2009.

[12] A. Boldyreva, N. Chenette, and A. O'Neill, "Order-preserving encryption revisited: Improved security analysis and alternative solutions," in Proc. Advances in Cryptology – CRYPTO 2011. Springer, Aug. 2011.

[13] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proc. Advances in Cryptology – EUROCRYPT99. Springer, May 1999.

[14] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symposium on Security and Privacy., May 2000.

[15] L. Ferretti, F. Pierazzi, M. Colajanni, and M. Marchetti, "Security and confidentiality solutions for public cloud database services," in Proc. Seventh Int'l Conf. Emerging Security Information, Systems and Technologies, Aug. 2013.

[16] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 68–73, Jan. 2008.

[17] L. Popa, S. Ratnasamy, G. Iannaccone, A. Krishnamurthy , and I. Stoica, "A Cost Comparison of Data Center Network Architectures," in Proc. ACM Int'l Conf. Emerging Networking Experiments and Technologies, 2010.

[18] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, and R. Buyya, "Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011