ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Exploiting RASP Data Perturbation to Build Confidential Query Services in the Cloud

Reena D K

Final year M.Tech, Department of Computer Science, S.T.J.I.T. Ranebennur, Karnataka, India

ABSTRACT: With the advent of cloud computing technology, using clouds for hosting data query service has become increasingly popular because of the low cost of computation, hosting applications and content storage. As in cloud the data management and infrastructure management is provided by third-party security and privacy are the biggest concern. Until and unless data confidentiality and secure query processing are guaranteed it is always a risk for the data owner to move the sensitive data to the cloud. Workload must be reduced to fully realize the benefits of cloud computing. Therefore to meet the above said requirements RASP method is proposed where RASP stands for Random Space Perturbation. This data perturbation technique ensures that the data is not distorted and does not lead to a security breach by allowing users to ascertain key summary information. Cloud computing enables outsourcing the management of the data related to individuals and organizations to a service provider as the hardware cost and the maintenance cost is less. RASP provides exclusive security features for hosting query services in the cloud by satisfying the CPEL criteria where CPEL stands for data Confidentiality, query Privacy, Efficient query processing and Low working cost .KNN-R algorithm is used to process the range query and the KNN query. Here users have been authorized by the randomly generated product key value provided by the admin after successful registration followed by activation by admin thus maintaining confidentiality. User queries are retrieved within a very short span of time. Also analysed how the RASP method provides confidentiality of data and increases the working process of query.

KEYWORDS: query services in the cloud, low in-house processing, RASP perturbation, Range query, KNN query.

I. INTRODUCTION

Cloud computing refers to the manipulation, configuration and access of applications online providing dynamically scalable infrastructure for application, data and file storage. On-demand self-service is offered by cloud computing. It is not mandatory to have interaction with the cloud service provider for the usage of resources. Cloud computing operates at higher efficiencies with greater utilization and is thus cost effective. Only Internet connection is required for it. Cloud computing becomes more reliable as it offers load balancing. Pay as per usage technique is followed by the billing model. Maintenance is lowered as the infrastructure is not purchased. This is an important feature as in the cloud the working time of the query services is very high and expensive.

Data Confidentiality and query Privacy have become the major concerns as the service provider might lose the control over the data in the cloud. Adversaries such as curious service provider can possibly make a copy of the database or secretly hears the user queries which may be difficult to detect and prevent in the cloud infrastructures.
For the protection of data and query privacy there is a need for new methods in the cloud. But, if the new methods provide slow query processing than it will be not advantageous. The approach holds the CPEL criteria for submitting a query in cloud. The CPEL criteria stands for Confidentiality of data, query Privacy, Efficient query processing and Low working cost. The complexity of constructing the query service becomes complex by this method.

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, to improve the privacy and security the crypto-index [4] approach puts heavy burden on the infrastructure. The Order Preserving Encryption (OPE) [1] is vulnerable to distribution based attacks. Use of cloaking boxes to protect the data objects and queries is followed in the new Casper approach [4] which affects the efficiency of query processing and the in-house workload.

10.15680/ijircce.2015.0304046

For the construction of query services in the cloud Random Space Perturbation (RASP) method is proposed. Here the query is separated as the Range query and the KNN query. This method satisfies all the four CPEL criteria concepts. The transformation of the multidimensional data is done with the combination of order preserving encryption, dimensionality expansion, random projection and random noise injection.

- Data confidentiality is provided by the RASP method and its combination. It is mainly used to protect the multidimensional range of queries in secure manner and with efficient query processing.
- The range query is used for retrieving the data's stored in the database where range denotes some value between upper and lower boundary.
- The kNN query denotes K-Nearest Neighbor query where K refers to the positive integer.

II. RELATED WORK

Here is the summarization about the study of the existing process.
In (1) 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. A bucket-diffusion scheme was proposed to protect the access pattern which however has to sacrifice the precision of query results and thus increase the client's cost of filtering the query result. In (4) 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. In **(7)** 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. In (9) the New Casper approach is used to protect data and query but the efficiency of the query process is affected. Distances-Recoverable Encryption (DRE) is the most intuitive method for preserving the nearest neighbour relationship. Because of the exactly preserved distances many attacks can be applied Wong et al. suggest preserving dot products 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.

III. PROPOSED ALGORITHM

a. *kNN-R Algorithm:*

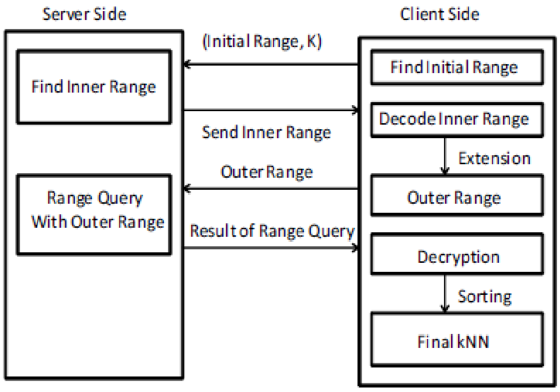


Fig.1 Procedure of kNN-R algorithm

Above figure shows two rounds of communication between Client and Server.

b. *Design Considerations:*

- The query processing time is minimized to an extent.
- User queries are answered in a short period i.e., 0.7 seconds.

- Only admin can activate the registered user and frequently checks his behavior and if found fraudulent practices can deactivate the user.
- Only with randomly generated product key provided by admin the user can access the cloud services thus data Confidentiality is guaranteed.

c. *Proposed Algorithm with description:*

Aim of the proposed algorithm is reduce the query processing time thus providing data confidentiality. The query processing finds the nearest k points in the *square range* that is centered at the query point. The proposed algorithm is consists of three main steps.

Step 1: Finding the inner range by the Server:

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. The *Inner Range* is the square range that contains at least k points, and the *Outer Range* encloses the square range that encloses the inner range.

Step 2: Providing records in the Outer range to the client:

The outer range surely contains the kNN results but it may also contain irrelevant points that need to be filtered out. 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.

Step 3: Finding top k candidates:
The client decrypts the records and finds the top k candidates as the final result.

IV. PSEUDO CODE

```
Step 1:  The client generates the initial range and sends its secure form to the server.
Step 2:  The server works on the secure range queries and finds the inner range covering at least $k$ points.
Step 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
Step 4:  The server returns the points in the outer range
Step 5: The client decrypts the points and extracts the $k$ nearest points.
Step 6: End.
```

V. EXPERIMENTAL RESULTS

The experimental studies involve the kNN-R algorithm when k=3 as shown in Fig.2. The kNN-R algorithm provides data confidentiality. There are at least k nearest neighbors to the query points with distances less than the radius r, as the inner range contains at least k points. Therefore, the k nearest neighbors must lie in the outer range. Proposed kNN-R algorithm illustrates this procedure. Fig.3 shows that the authorized users can retrieve their queries either by Range query or kNN query. Because the user won't know in advance about the result for the query that how many entries may come as a result for the query so the range query is not usual. The query processing time is close to the upper bound 0.7. Fig.5 shows the result of Advanced Encryption Standard (AES) algorithm for the text file in Fig.4. The proposed system uses AES algorithm for Encryption. The key consideration dealt in the proposal of implementing AES encryption for security over data that provides benefits of less memory consumption and less computation time as compared to other algorithms. Though each cloud infrastructure has its own security strengths, the user can choose infrastructure according to security constraints. AES provide security to cloud consumers as encrypted data which is safe from many attacks .There are no serious weak keys in AES and it supports any block sizes and key sizes that are multiples of 32 (greater than 128-bits). Statistical analysis of the cipher text has not been possible even after using huge number of test cases. No differential and linear cryptanalysis attacks have been yet proved on AES.

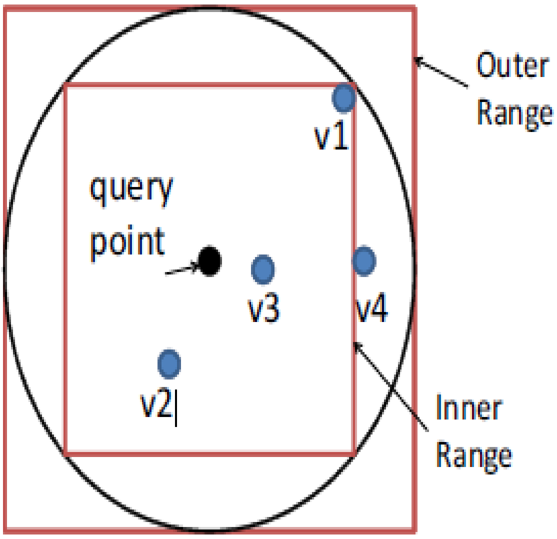
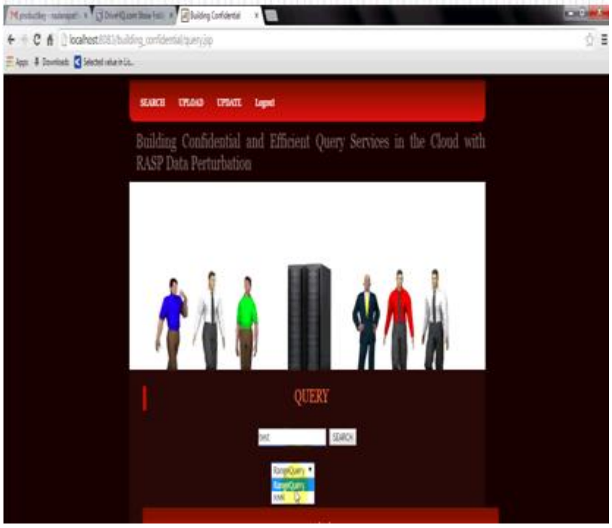
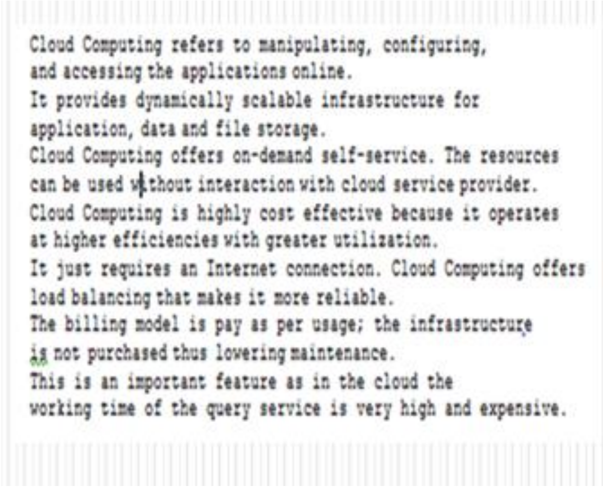


Fig.2 Illustration for kNN-R Algorithm when k=3

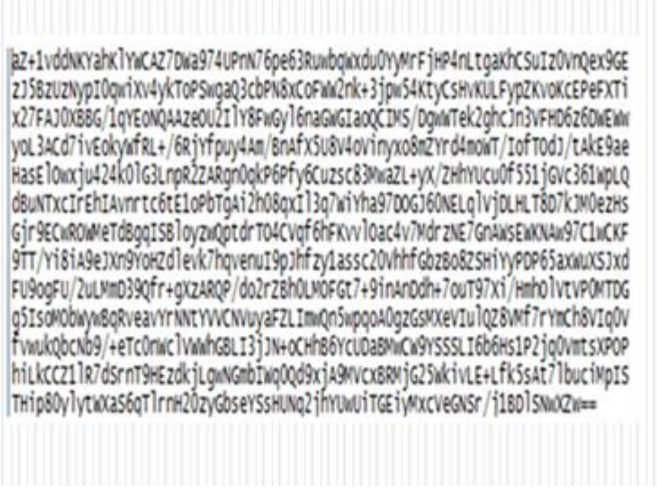

localhost:8080/building_confidential/query.jsp
SEARCH UPLOAD UPDATE Logout

Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation

Fig.3 User Query search options



Cloud Computing refers to manipulating, configuring,
and accessing the applications online.
It provides dynamically scalable infrastructure for
application, data and file storage.
Cloud Computing offers on-demand self-service. The resources
can be used without interaction with cloud service provider.
Cloud Computing is highly cost effective because it operates
at higher efficiencies with greater utilization.
It just requires an Internet connection. Cloud Computing offers
load balancing that makes it more reliable.
The billing model is pay as per usage; the infrastructure
is not purchased thus lowering maintenance.
This is an important feature as in the cloud the
working time of the query service is very high and expensive.



aZ+1vddNKYahKlYWCAZ7DWa974UPnN76pe63RuwbqwxduOYyMrFjHP4nLtgaKhCSuIz0VnQex9GE
zJ5BzUzNypI0qwiXv4ykToPSwgaQ3cbPN8xCoFWW2nk+3jpw54KtyCsHvKULFypZKvoKcEPeFXTi
x27FAJ0XBBG/1qYEoNQAAzeOU2IlY8FwGyl6naGWGIaoQCIMS/DgwwTek2ghcJn3VFHD6z6DwEWw
yoL3ACd7ivEokywfRL+/6RjYfpuy4Am/BnAfX5U8V4oVinyxo8mZYrd4moWT/IofTOdJ/tAkE9ae
HasElOwxju424k0lG3LnpR2ZARgnOqkP6Pfy6Cuzsc83MwaZL+yX/ZHhYUcu0f551jGVc361wpLQ
dBuNTxcIrEhIAvnrtc6tE1oPbTgAi2h08qxIl3q7wiYha97DOGJ60NELqlVjDLHLT8D7kJM0ezHs
Gjr9ECwROWMeTdBgqISBloyzwQptdrTO4CVqf6hFKvvl0ac4v7MdrzNE7GnAWsEWKNAw97C1wCKF
9TT/Yi8iA9eJXn9YoHZdlevk7hqvenuI9pJhfzy1assc20VhhfGbzBo8ZSHiYyPDP65axWuXSJxd
FU9ogFU/2uLMmD39Qfr+gXzARQP/do2rZBhOLMOFGt7+9inAnOdh+7ouT97Xi/HmhOlVtVPOMTDG
g5IsoMObwywBqRveaVYrNNtYVVCNVuyaFZLImwQn5wpqoA0gzGsMXeVIulQZ8VMf7rYmCh8VIq0V
fvwukQbcNb9/+eTcOnwclVWWhGBLI3jJN+oCHhB6YcUDaBMwCw9YSSSLI6b6Hs1P2jqOVmtsXPOP
hiLkCCZ1lR7dSrnT9HEzdkjLgwNGmbIwqOQd9xjA9MVCxBRMjG25wkivLE+Lfk5sAt7lbuciMpIS
THip80ylytwXaS6qTlrnH20zyGbseYSsHUNq2jhYUwUiTGEiyMxCVeGNSr/j1BDlSNwXZw==

Fig.4 Normal Text File algorithm.

Fig.5 Same Text file stored in perturbed form in cloud after encryption by AES

VI. Conclusion and Future Work

This project pursued the idea of using open source cloud namely DriveHQ as the storage of data submitted by the data owner and the trusted party in a perturbed form. Upload speed can be many times faster in certain cases by using DriveHQ magic upload technology. Before the files are uploaded to DriveHQ, they are encrypted locally. Thus it is extremely secure. Solutions to certain top issues in cloud computing are provided by the proposed work. Firstly, lack of security which results in data being intentionally corrupted by the unauthorized disclosure of information. Secondly, slow query services in the cloud. Processing time of query is minimized to a larger extent. The data will be encrypted by Advanced Encryption Standard (AES) algorithm and stored in the cloud database to avoid original data loss.

The project is a feasibility study that aimed to explore the feasibility and potential for utilizing Cloud capability to address data storage and processing needs. And can also continue studies to provide still better perturbation approach and can improve the effect of the query by using real time cloud.

REFERENCES

[1] 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.

[2] J. Bau and J.C. Mitchell, "Security Modeling and Analysis," IEEE Security and Privacy, vol. 9, no. 3, pp. 18-25, May/June 2011.

[3] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge Univ. Press, 2004.

[4] 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.

[5] 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.

[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer-Verlag, 2001.

[7] B. Hore, S. Mehrotra, and G. Tsudik, "A Privacy-Preserving Index for Range Queries," Proc. Very Large Databases Conf. (VLDB), 2004.

[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] 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.

BIOGRAPHY

Reena D K is a M.Tech Student in Computer Science and Engineering in STJIT Ranebennur affiliated to Visvesvaraya Technological University, Belgaum Karnataka. She completed her B.E in Information Science and Engineering from SDMCET Dharwad in the year 2012. Her research interests are Cloud computing, Data Mining, Networking etc.