International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Erection Trusted and Effective Request Services in the Cloud with RASP Data Perturbation

Rashmi Kadu[1], J. L. Chaudhari[2]

[1]ME CSE- Final Year, BSIOTR Wagholi (Pune), Maharashtra, India

[2]Assistant Professor, CSE Dept, BSIOTR Wagholi (Pune), Maharashtra, India

Abstract: ***Range query is one of the most frequently used queries for online data analytics. Providing such a query service could be expensive for the data owner. With the development of services computing and cloud computing, it has become possible to outsource large databases to database service providers and let the providers maintain the range-query service. With outsourced services, the data owner can greatly reduce the cost in maintaining computing infrastructure and data-rich applications. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. 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. We use RASP to generate indexable auxiliary data that is resilient to prior knowledge enhanced attacks. Range queries are securely transformed to the encrypted data space and then efficiently processed with a two-stage processing algorithm.***

Keywords: RASP, query services in the cloud, privacy, range query, kNN query

1. Introduction

Cloud computing infrastructures are popularly used by peoples now a day. Cloud computing is the delivery of computing services over the Internet. Examples of cloud services include online file storage, social networking sites, webmail, and online business applications. The use of cloud services is popular in world because of its attractive features like secure service, infinite of storage, it will satisfy the user experience, low cost and multiple user can access the files and applications. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. Cloud computing provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications. In cloud, the query service process are frequently used because, the user can save their cost. [2]. The owners in the cloud will pay the amount only for their using time of server.

To prevent any system or data from attack new approaches should be develop. While new approaches are needed to fulfill some requirements. 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

While using the exiting services of cloud computing, the growing concern is how to store, manage, and analyze a large volume of data while preserving the privacy. 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 basic idea is to randomly transform the multidimensional datasets with a combination of order preserving encryption, dimensionality expansion, random noise injection, and random project, so that the utility for processing range queries is preserved.

The proposed approach has a number of unique contributions:

- 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 [2].
- 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.

2. System Architecture

Cloud computing infrastructures used to store large datasets and query services. It consists of two parts 1. Customer and 2. Cloud. 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 [2]. 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.

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 [1]. 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;
(3) H(q′,D′) is the query processing algorithm that returns the result R′.
Figure 1 shows the system architecture for both RASP-base range query service and kNN service.

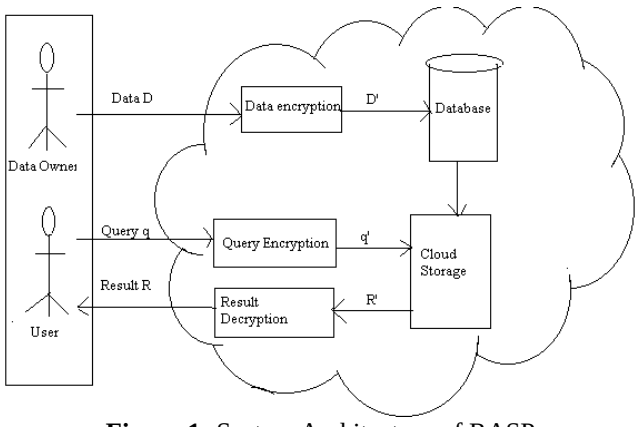

Figure 1: System Architecture of RASP

3. Existing System

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

3.1 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

OPE is a remote un-trusted database server is able to index the (sensitive) data it receives, in encrypted form, in a data structure that permits efficient range queries (asking the server to return ciphertexts in the database whose decryptions fall within a given range, say [a;b]). OPE not only allows efficient range queries, but allows indexing and query processing to be done exactly and as efficiently as for un-encrypted data, since a query just consists of the encryptions of a and b and the server can locate the desired ciphertexts in logarithmic-time via standard tree-based data structures.

In random projection (RP), the original high-dimensional data is projected onto a lower-dimensional subspace using a random matrix whose columns have unit lengths. Random Projection has been found to be a computationally efficient, yet sufficiently accurate method for dimensionality reduction of high-dimensional data sets. While this method has attracted lots of interest, empirical results are sparse.

Noise injection consists of adding noise to the inputs during query construction. Experimental results suggest that it might improve the generalization ability of the resulting network [4]. 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.

3.2 Range Query

Range query is the query used to retrieve the data from the database. As name suggests, user has to provide certain range that is upper bound and lower bound It will retrieve the data value which 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. Consider an example:

User need to fire a range query which can retrieve the entries from India where persons who are above 50 years in the top 10 list from the record of India. So the query will be:

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

Another example of range query is; choosing students from Student table who has marks between 40 to 60 in the examination. Query will be like:

```
SELECT Student_Names, Marks
FROM Student
WHERE Marks BETWEEN 40 AND 60;
```

3.3 KNN Query Processing

In this section, we design a kNN query processing algorithm based on range queries (the kNN-R algorithm).

3.3.1 Overview

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 [3].

Features:

- All instances correspond to points in an n-dimensional Euclidean space
- Classification is delayed till a new instance arrives
- Classification done by comparing feature vectors of the different points
- Target function may be discrete or real-valued

An arbitrary instance is represented by (a1(x), a2(x), a3(x),.., an(x)), where ai(x) denotes features
Euclidean distance between two instances

$$\text{Dist}(X,Y) = \sqrt{\sum_{i=1}^{D}(Xi - Yi)^2} \quad (1)$$

3.3.2 Examples:

1. Nearest Neighbor

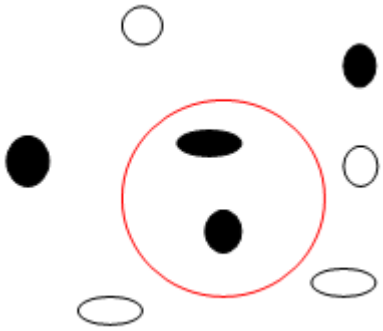

Figure 2: 1-Nearest Neighbor

2. K-Nearest Neighbor

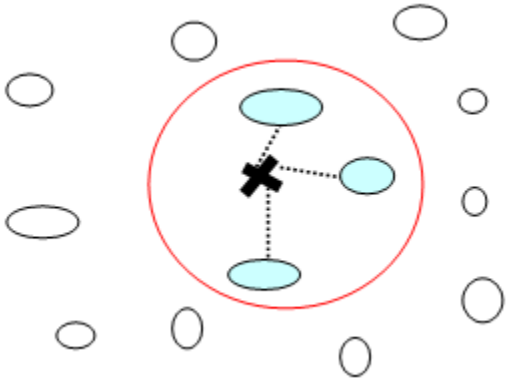

Figure 3: 1-Nearest Neighbor

Stored training set patterns

Input pattern for classification

----- Euclidean distance measure to the nearest three patterns

Number of nearest neighbors

- The numbers of nearest neighbors (K) should be based on cross validation over a number of K setting.
- When k=1 is a good baseline model to benchmark against.
- A good rule-of-thumb numbers is k should be less than the square root of the total number of training patterns.

Proposition: The kNN-R algorithm returns results with 100% recall.
Proof: The sphere in Figure 4 between the outer range and the inner range covers all points with distances less than the radius r. Because the inner range contains at least k points, there are at least k nearest neighbors to the query points with distances less than the radius r. Therefore, the k nearest neighbors must be in the outer range. The kNN-R algorithm consists of two rounds of interactions between the client and the server.

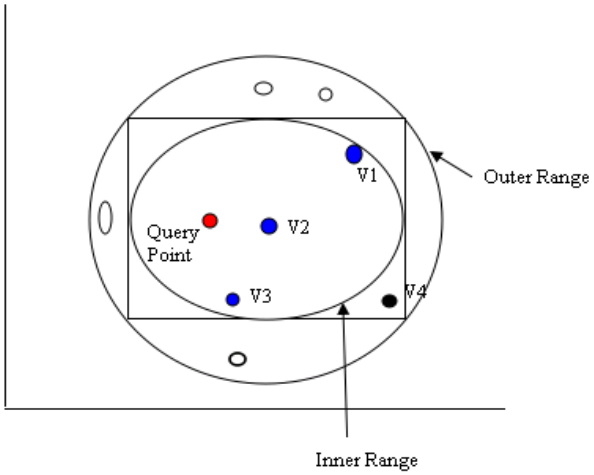

Figure 4: Illustration for kNN-R Algorithm when k=3

1) 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.
2) 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.
3) The client decrypts the records and find the top k candidates as the final result.

4. Proposed System

Here we summarized about the study of the existing process.

- **OPE**: 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.
- **Privacy Preserving**: 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.
- **Crypto index**: 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. And also we had study about RASP method, query privacy, enabling search services on out sourced data and many concepts.

5. Conclusion

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 will continue our studies

on two aspects: (1) further improve the performance of query processing for both range queries and kNN queries; (2) formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality

References

[1] Huiqi Xu, Shumin Guo, Keke Chen, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL:26 NO:2 YEAR 2014*

[2] E.Saral Elizabeth, Ms.K. Padmaveni, "Confidential and Efficient Query Services in the Cloud", *IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 1, Feb-Mar, 2014 ISSN: 2320 - 8791*

[3] 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.*

[4] Shaozhi Ye, Felix Wu, Raju Pandey,Hao Chen, "Noise Injection for Search Privacy Protection", *Department of Computer Science University of California, Davis Davis, CA 95616*

[5] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in *Proceedings of ACM SIGMOD Conference*, 2004.

[6] 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.

[7] J. Bau and J. C. Mitchell, "Security modeling and analysis,"I*EEE Security and Privacy*, vol. 9, no. 3, pp. 18–25, 2011.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.