IJCSEC- International Journal of Computer Science and Engineering Communications.
Vol.3, Issue 2, 2015, Page.757-762, ISSN: 2347–8586

Build Efficient Query Services in the Cloud with RASP and Range Queries

Brindha.P[1], M.Sivanathan[2]
Department of Computer Science and Engineering [1, 2]
EBET Group of Institutions, Nathakadiyur, Kangayam, India[1, 2]
brindhassk@gmail.com[1] ,ms.cse@ebet.edu.in[2]

***ABSTRACT**:* Cloud computing is mainly used to store and retrieve data and also to host data query services has become an appealing solution for the advantages on scalability and cost-saving. However, Encryption is a well-established technology for protecting sensitive data. When data encrypted once, data can no longer be easily queried aside from exact matches .A secured query service reduce the in-house workload to fully realize the benefits of cloud computing. The RASP data perturbation and kNN query services method provide secure range query and kNN query services for protected data in the cloud. Continuous K nearest neighbor queries (C-KNN) is defined as the nearest points of interest to all the points on a path. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the queries. The attacks on data and queries under a precisely defined threat model and realistic security assumptions are analyzed. Extensive experiments have been conducted to show the advantages of this approach on efficiency and security.
***Keywords*:** query services in the cloud, privacy, range query, kNN query

I. INTRODUCTION

Hosting data-intensive query services in the cloud is increasing now a day. The unique advantages in cloud is scalability and cost-saving. With the cloud Infrastructures, the service owners can only pay for the hours of using the servers. It will be expensive and inefficient to serve dynamic workloads with in-house infrastructures [2]. The service providers can lose the control over the data in the cloud. Adversaries, such as curious service Providers can possibly make a copy of the database or eaves drop users' queries, which will be difficult to detect and prevent in the cloud infrastructures. 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, we propose the Random Space Perturbation (RASP) approach to constructing practical range query and k-nearest-neighbor (kNN) query services in the cloud. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries. RASP has several important features. RASP does not preserve the order of dimensional values and thus does not suffer from the distribution-based attack. The Knn query service is to find the nearest places. Order preserving encryption (OPE) scheme maps a set of single-dimensional values to another, while keeping the value order unchanged. The RASP perturbation is a unique combination of OPE. The main scope of the project is to host data query services using clouds and build efficient query services with data perturbation method. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the queries that stored in the cloud. The rest of the paper is organized as follows- Section II is about some existing methodologies proposed for partially observable system. In Section III, the proposed Color Pass scheme has been discussed in detail. The user interface for Color Pass has been described in Section IV. Finally we conclude in Section VI and give future direction of our work.

II. RELATED WORK

Another line of research facilitates authorized users to access only the authorized portion of data, e.g., a certain range, with a public key scheme. The most relevant work about perturbation techniques includes the random noise addition methods and the condensation-based perturbation technique.

A. Authorized users with keywords:

Their approach requires that the data owner provides the indices and keys for the server, and authorized Users use the data in the server. While in the cloud database scenario, the cloud server takes more responsibilities of indexing and query processing. Secure keyword search on encrypted documents scans each encrypted document in the database and finds the documents containing the keyword, which his more like point search in database. The research on privacy preserving data mining has discussed multiplicative perturbation methods [7], which are similar to the RASP encryption, but with more emphasis on preserving the utility for data mining.

B.Private information retrieval

(PIR) tries to fully preserve the privacy of access pattern, while the data may not be encrypted. PIR schemes are normally very costly. Use a pyramid hash index to implement efficient privacy preserving data block operations based on the idea of Oblivious RAM.

Another line of research facilitates authorized users to access only the portion of data in the authorized range with a public key scheme. The underlying identity based encryption used in these schemes does not produce indexable encrypted data. The untrusted service provider in our setting is responsible for both indexing and query processing. Secure keyword search on encrypted documents scans each encrypted document in the database and finds the documents containing the keyword, which is more like point search in database. The research on privacy preserving data mining has discussed multiplicative perturbation methods which are similar to the RASP encryption, but with more emphasis on preserving the utility for data mining

A. Random Noise Addition Approach

The random noise addition approach can be briefly described as follows: A new decision-tree algorithm for the randomization approach is developed in order to build the decision tree from the perturbed data. Randomization approach is also used in privacy-preserving association-rule mining.

While the randomization approach is intuitive, several researchers have recently identified privacy breaches as one of the major problems with the randomization approach. The authors demonstrated that the randomization approach preserves little privacy in many cases.

Furthermore, there has been research addressing other weaknesses associated with the value based randomization approach. For example, most of existing randomization and distribution reconstruction algorithms only concern about preserving the distribution of single columns. There has been surprisingly little attention paid on preserving value distributions over multiple correlated dimensions.

Second, value-based randomization approach needs to develop new distribution-based classification algorithms. In contrast, our random rotation perturbation approach does not require modify existing data classification algorithms when applied to perturbed datasets. The randomization approach is also generalized to improve the balance between the privacy and accuracy.

C. Condensation-based perturbation approach

The condensation approach aims at preserving the covariance matrix for multiple columns. Different from the randomization approach, it perturbs multiple columns as a whole to generate entire "perturbed dataset".

The condensation approach can be briefly described as follows. The authors demonstrated that the condensation approach can preserve data covariance well, and thus will not significantly sacrifice the accuracy of classifiers if the classifiers are trained with the perturbed data. However, we have observed that the condensation approach is weak in protecting the private data. The *KNN* based data groups result in some serious conflicts between preserving covariance information and preserving privacy.

As the authors claim, the smaller the size of the locality in each group, the better the quality of preserving the covariance with the regenerated *k* records is. We design an algorithm that tries to find the nearest neighbor in the original data for each regenerated record. The result shows that the difference between the regenerated records and the nearest neighbor in original data is very small, and thus, the original data records can be estimated from the perturbed data with high confidence

III. PROPOSED METHODOLOGY

The proposed approach will address all the aspects of the criteria and aim to achieve a good balance on them. The RASP perturbation is designed in such a way that the queried ranges are securely transformed into polyhedral in the RASP-perturbed data space, which can be efficiently processed with the support of indexing structures in the perturbed space. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries.

Proposed System Advantages

The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee.

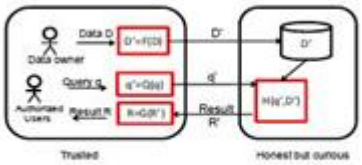
The proposed service constructions are able to minimize the in-house processing workload because of the low perturbation cost and high precision query results. This is an important feature enabling practical cloud-based solutions.

Order Preserving Encryption schemes, does not suffer from the distribution-based attack so threats are reduced.

The transformed range queries are secure under the assumptions, so the kNN queries are also secure.

RASP Based Query

The proposed system use new authentication technique that consists of phases: registration phase, login phase during the registration phase user register their details. In login phase user enters with their username and password, if it is valid then it enter in to our application and upload or download our data in the cloud. The administrator can view how many number of visitors, visited our files. At last we can prevent the query leakage and access pattern mechanism. The query leakage can be done using KNN and RASP algorithms.



I. The system architecture for RASP-based query

IV. KNN QUERY PROCESSING WITH RASP

RASP is one type of multiplicative perturbation, with a novel combination of OPE, The RASP data perturbation method to provide secure and efficient range query and kNN query services for protected data in the cloud. The RASP range query algorithm is mainly to process the kNN queries. He RASP perturbation does not preserve distances (and distance orders), kNN query cannot be directly processed with the RASP perturbed data. The RASP-perturbed data records are only used for indexing and helping query processing, there is no need to recover the perturbed data.

Algorithm 1 RASP Data Perturbation

1: RASP Perturb(X, RN G, RIM G, K_O)
2: Input: X : k × n data records, RN G: random real value generator that draws values from the standard normal distribution, RIM G : random invertible matrix generator,
K_{ope} : key for OPE E_{ope}; Output: the matrix A
3: A ← 0;
4: A_3 ← the last column of A;
5: v_0 ← 4;
6: ***while*** *A_3 contains zero* ***do***
7: generate A with RIM G;
8: ***end while***
9: ***for*** *each record x in X* ***do***
10: $v \leftarrow v_0 - 1$;
11: ***while*** *$v < v_0$* ***do***
12: $v \leftarrow RNG$;
13: ***end while***
14: $y \leftarrow A((E_{ope}(x, K_{ope}))^T, 1, v)^T$;
15: submit y to the server;
16: ***end for***
17: return A;

RASP has several important features.

First, RASP does not preserve the order of dimensional values Second, RASP does not preserve the distances between records, which prevents the perturbed data from distance-based attacks RASP does not preserve other more sophisticated structures such as covariance matrix and principal components. Therefore, the PCA-based attacks do not work as well. Third, the original range queries can be transformed to the RASP perturbed data space, which is the basis of our query processing strategy. A range query describes a hyper-cubic area (with possibly open bounds) in the multidimensional space.

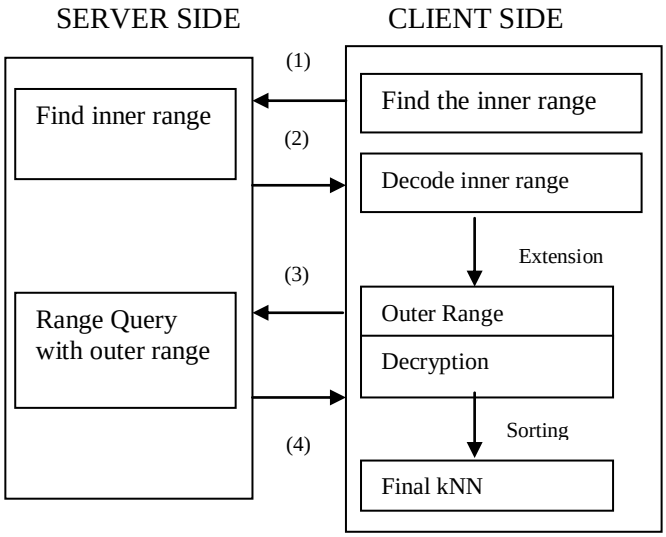
Overview of the kNN-R Algorithm:

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. There are a number of key problems to make this work securely and efficiently.

The algorithm is based on square ranges to approximately find the kNN candidates for a query point, which are defined as follows;

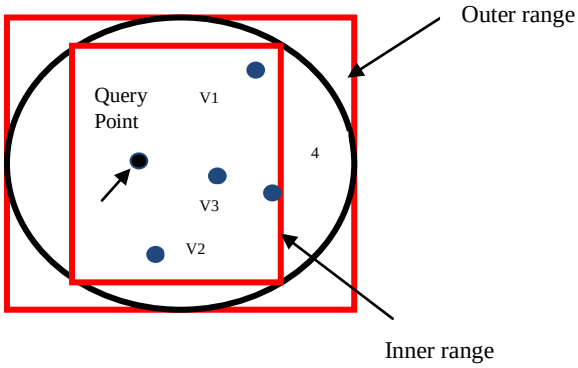
Definition 1: A square range is a hyper-cube that is centered at the query point and with equal-length edges.

PROCEDURE FOR KNN ALGORITHM



(1) Initial Range k
(2) Send Inner Range
(3) Outer Range
(4) Result of Range Query

ILLUSRATION FOR KNN ALGORITHM



V. CONCLUSION

We propose the RASP perturbation approach to hosting query services in the cloud, which satisfies the CPEL criteria: data Confidentiality, query Privacy, Efficient query processing, and Low in-house work- load. The requirement on low in-house workload is a critical feature to fully realize the benefits of cloud computing and efficient query processing is a key measure of the quality of query services.

RASP perturbation is a unique composition of OPE, dimensionality expansion, random noise injection, and random projection, which provides unique security features. It aims to preserve the topology of the queried range in the perturbed space, and allows using indices for efficient range query processing. With the topology-preserving features, we are able to develop efficient range query services to achieve sub- linear time complexity of processing queries. We then develop the kNN query service based on the range query service. The security of both the perturbed data and the protected queries is carefully analyzed under a precisely defined threat model. We also conduct several sets of experiments to show the efficiency of query processing and the low cost of in-house processing.

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

[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 Berkeley*, 2009.

[3] J. Bau and J. C. Mitchell, "Security modeling and an IEEE Security and Privacy, vol. 9, no. 3, pp. 18–25, 2011.

[4] S. Boyd and L. Vandenberg he, *Convex Optimization*. Cam- bridge University Press, 2004.

[5] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy preserving multi-keyword ranked search over encrypted cloud data," in *INFOCOMM*, 2011.

[6] 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

[7] K. Chen and L. Liu, "Geometric data perturbation for out- sourced data mining," *Knowledge and Information Systems*, 2011.

[8] K. Chen, L. Liu, and G. Sun, "Towards attack-resilient geometric data perturbation," in *SIAM Data Mining Conference*, 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, "Search- able symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security*. New York, NY, USA: ACM, 2006, pp. 79–88.

[11] N.R.Draper and H. Smith, *Applied Regression Analysis*. Wiley, 1998.

[12] 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.