International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
Dec 2016

Secure And Efficient Query Services In The Cloud With Random Space Perturbation

[1]R.Mounika, [2]Dr.Ravindar Reddy Thokala

[1] M.Tech Student, Dept of CSE, Brilliant grammar school educational institutions group of institutions integrated campus, T.S, India

[2]Associate Professor, Dept of CSE, Brilliant grammar school educational institutions group of institutions integrated campus, T.S, India

Abstract: 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. However, some data might be sensitive that the data owner does not want to move to the cloud unless the data confidentiality and query privacy are guaranteed. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. The random space perturbation (RASP) data perturbation method to provide secure and efficient range query and kNN query services for protected data in the cloud. 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.

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

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. [1]

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 low 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.[1]

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.

Related Work:

We review 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.

RASPQS ARCHITECTURE:

We assume that a cloud computing infrastructure, such as Amazon EC2, is used to host the query ervices and large datasets. The purpose of this architecture is to extend the proprietary database servers to the public cloud, or use a hybrid private-public cloud to achieve scalability and reduce costs while till maintaining confidentiality.

Each record x in the outsourced database contains two parts: the RASP-processed attributes $D' = F(D,K)$ for indexing and query processing, and the encrypted original records, $Z = E(D,K')$, for lossless record retrieval, where K and K′ are keys for perturbation and encryption, respectively. Figure 1 shows the system architecture for both RASP-based range query service and kNN query service. 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. The authorized users can submit range queries or kNN queries to learn statistics or find some records. The untrusted parties include the

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
Dec 2016

curious cloud provider who hosts the query services and the protected database.

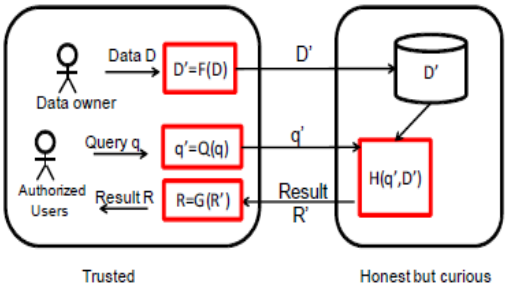

Figure 1: The RASP-QS system architecture.

The RASP-perturbed data will be used to build indices to support query processing. 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 works on D′ and q′ and returns the result R′. It is also possible to securely compute some statistics such as SUM or AVG of a specific dimension, if the original data records Z = E(D,K′) are encrypted by dimension and with a partial homomorphic encryption such as Paillier encryption. The result is recovered with the procedure G(R′).

KNNR Query Processing:

The kNN-R algorithm uses square ranges around the query point to find the candidate nearest records. The inner square range starts from the query point and expands until k points are included. The exact kNN result should be in the bounding sphere of the inner range, which in turn is approximated by the bounding box of the sphere. Figure 2 shows the scenario of finding the candidate set for a 3-NN query based on square ranges. The inner range expansion can be achieved by a binary range search algorithm. The user can set the initial outer square range with a certain distance from the query point. In each iteration, the algorithm finds the middle range between the inner range and the outer range, in which if the number of enclosed points is larger than k, the outer range is replaced by the middle range; otherwise, the inner range is replaced by the outer range. This iterative process can exponentially reduce the search range and find the result quickly. The records in the final range is sent back to the client for final kNN filtering. Note that this process utilizes the linear property of the transformed queries to derive the queries for the middle range, which does not require the client's

participation [5]. Experiments show that this algorithm is very efficient.

Performance Comparison:

To further understand the advantages of the RASP approach, we want to show the comparison with the R*-Tree supported query processing on the original data, and the sequential scan on the encrypted data (e.g., the work on range query [2]). These methods will also be implemented with C++ for fair performance evaluation. We will let the user generate a batch of random queries with a specific size of range for a selected dataset. All the queries will be sequentially submitted by the client side and processed by the server. We will show the average query processing time cost in the server side, the server storage cost, and the client-side pre-processing and post-processing costs. We expect that the RASP approach will have much lower storage cost, query processing time, and client-side processing costs, than other methods that depend on encryption and linear scan.

SUMMARY

The purpose of this demonstration is to show the key ideas of the RASP-based query processing approach for efficiently and confidentially hosting query services in public clouds. This demonstration system will be highly interactive and visual, allowing the users to easily understand the technical details and appreciate the advantages of this approach. Users of the demonstration system can manipulate the system to generate perturbation parameters, observe the key steps in query processing, and evaluate the performance of several related approaches. The technical details of the RASP approach have been published recently in the journal paper [5], for which this demonstration system will be a valuable addition. We believe that the RASP approach will be a significant step towards practical confidential query services in public clouds. This work is partially supported by NSF Award 1245847.

5. REFERENCES:

[1] Agrawal, R., Kiernan, J., Srikant, R., and Xu,Y. Order preserving encryption for numeric data. In Proceedings of ACM SIGMOD Conference (2004).\

[2] Boneh, D., and Waters, B. Conjunctive, subset, and range queries on encrypted data. In the Theory of Cryptography Conference (TCC (2007), Springer, pp. 535–554.

[3] Chen, K., and Liu, L. VISTA: Validating and refining clusters via visualization.

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
Dec 2016

Information Visualization 3, 4 (2004), 257–270.

[4] Curtmola, R., Garay, J., Kamara, S., and Ostrovsky, R. Searchable symmetric encryption: improved definitions and efficient constructions. In ACM CCS (2006), pp. 79–88.

[5] 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).