International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
*ISSN (Online)*2349-6967
Volume 2 , Issue 1(Jan-Feb 2015), PP240-243

RASP Data Perturbation Method to Provide Secure and Efficient Query Services In the Cloud

Miss.Puja Bhaganagarkar

(Department of Computer ScienceDr.D.Y.Patil College of Engineering, Ambi, Pune, India)

.

Abstract :- To host day services in public cloud is an best solution for the solution to save operating expense. However users have concerns on data security to lost control of infrastructure. RASP used to provide security to the setting of cloud based computing while enabling much faster query processing compared to the encryption based approach. In RASP encryption data confidentiality and query privacy are guaranteed when appealing it for range query and kNN. In order to process the range query to kNN query here used kNN-R algorithm

Keywords:-Confidentiality, kNN-R Algorithm, RASP Algorithm, Query Processing

I. INTRODUCTION

The extensive exploitation of cloud infrastructures has made it possible to host services and big data in public clouds. This new paradigm is especially attractive for data intensive query and analysis services for its great scalability and significant cost savings. It is well known that maintaining and mining data incurs much higher cost than initial data acquisition. 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. On the other hand, a secured query service should still provide efficient query processing and significantly reduce the in-house workload to fully realize the benefits of cloud computing.By moving data services to the cloud, data owners can cut costs in almost every aspect of managing and mining data. However, data privacy is still haunting data owners' minds as the underlying infrastructure is out of their control. In particular, data owners may not be aware of information leakage, which can happen in all kinds of possibilities, if the cloud provider does not want to report the leakage. A straightforward method is to encrypt datasets before exporting them to the cloud. However, searchable encryption is very challenging, showing limited successes in some specific areas such as document search [4]. Boneh et al. [2] showed that it is possible to construct a public-key system for range query, which is one of the basic database queries,However, it requires a significant amount of storage and computational costs, only applicable tolinear scan of the entire database. Database queries such as range and kNN queries normally demand fast processing time (logarithmic or sublinear time complexity) with the support of indexing structures. However, if not impossible, there is no efficient indexing structure developed for encrypted data yet, which renders the current encryption schemes [2] unusable for search in large databases. Recently proposed the Random Space Perturbation(RASP) method [5] for the protection of tabular data, which is secure under the assumption of limited adversarial knowledge - only the perturbed data and the data distributions are known by adversaries. This assumption is appropriate in the context of cloud computing. The RASP perturbation is a unique combination of Order Preserving Encryption (OPE) [1], dimensionality expansion, noise injection, and random projection, which provides sufficient protection for the privacy of query services in the cloud. It has a number of unique features, such as preserving the topology of range query, non-deterministic results for duplicate records, and resilience to distributional attacks [5].

We develop the secure half-space query transformation method that casts any enclosed range in the original space to an irregularly shaped range in the perturbed space. Therefore, are able to use a two-stage range query processing method: an existing multidimensional index, such as R*-Tree in the perturbed space is used to find out the records in the bounding box of the irregularly shaped range, which is then filtered with the transformed query condition. This processing strategy is fast and secure under the security assumption. To allow the readers to fully appreciate the intuition and the ideas behind the RASP based perturbation and query processing, RASP Query Services (RASPQS) demonstration system. This system consists of the following major components: (1) the user interface for perturbation parameter generation that allows users to observe the details of RASP perturbation, (2) the visualization of the two-stage range query processing procedure to understand the transformed query ranges and the query results, (3)the visualization of the progressive steps in the kNN query processing that is based on RASP range query processing, and (4) the performance comparison on index-aided processing on non-encrypted data, linear-scan query processing on encrypted data [2], and the RASP query processing.

II. LITERATURE SURVEY

Using clouds to host data query services has become an appealing solution for the advantages on scalability and cost-saving. 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. In addition, a secured query service should still provide efficient query processing and significantly reduce the in-house workload for the purpose of cloud computing. Bearing these criteria in mind, The RASP data perturbation method to provide secure 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 multidimensional indexing techniques to be applied in range query processing. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the kNN queries. Carefully analyze the attacks on data and queries under a precisely defined threat model and realistic assumptions. Extensive experiments have been conducted to show the advantages of this approach on the balance of performance and security. Secure data intensive computing in the cloud is challenging, involving a complicated tradeoff among security, performance, extra costs, and cloud economics. Although fully homomorphic encryption is considered as the ultimate solution, it is still too expensive to be practical at the current stage. In contrast, methods that preserve special types of data utility, even with weaker security, might be acceptable in practice. The recently proposed RASP perturbationmethod falls into this category. It can provide practical solution for specific problems such as secure range queries, statistical analysis, and machine learning. The RASP perturbation embeds the multidimensional data into a secret higher dimensional space, enhanced with random noise addition to protect the confidentiality of data. It also provides a query perturbation method to transform half-space queries to a quadratic form and, meanwhile, preserving the results of half-space queries. The utility preserving property and wide application domains are appealing. However, since the security of this method is not thoroughly analyzed, the risk of using this method is unknown. The purpose of this paper is to investigate the security of the RASP perturbation method based on a specific threat model. The threat model defines three levels of adversarialpower and the concerned attacks. Show that although the RASP perturbed data and queries are secure on the lowest level of adversarial power, they do not satisfy the strong indistinguishability definition on higher levels of adversarial power. As noticed, the indistinguishability definition might not be too strong to be useful in the context of data intensive cloud computation. In addition, the noise component in the perturbation renders it impossible to exactly recover the plain data; thus, all attacks are essentially estimation attacks. A weaker security definition based on information theoretic measures to describe the effectiveness of estimation attacks, and then study the security under this weaker definition. This security analysis helps clearly identify the security weaknesses of the RASP perturbation and quantify the expected maintenance of confidentiality under different levels of adversarial power.

III.SYSTEM FEATURES

3.1 System Feature 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 thatcomparison 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 :potential and good performances.

3.2 System Feature 2 -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 a result for the query.

3.3 System Feature 3 – 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 integervalue. 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 clientand then the client will decrypt the record and find the top k files to provide the final result.

IV.SYSTEM DESIGN

4.1 System Architecture

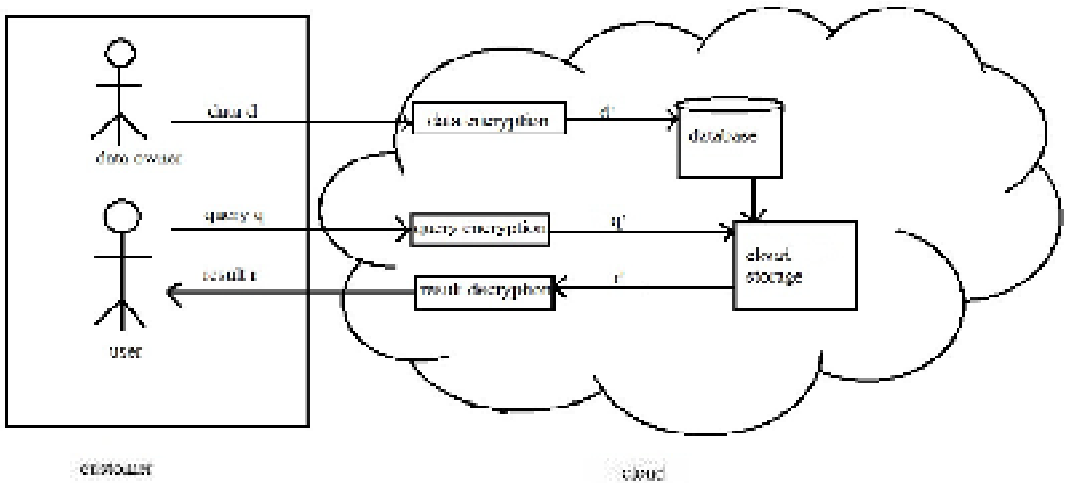

Fig.4.1 system architecture

Assume that a cloud computing infrastructure, is employed to host the question services and huge datasets. the aim of this design is to increase the proprietary info servers to the general public cloud, or use a hybrid private-public cloud to realizequantifiability and scale backpriceswhereas still maintaining confidentiality. Each record x within the outsourced info contains 2 parts: the RASP-processed attributes D′ = F (D, K) for assortment and questionprocess, and therefore the encrypted original records, Z = E(D, K′), for lossless record retrieval, wherever K and K′ ar keys for perturbation and cryptography, severally.Thehigher than figure shows the system design for each RASP-based varyquestion service and kNNquestionservice.Therear2 clearly separated groups: the sure parties and therefore the untrusted parties. The sure parties embrace the data/service owner, the in-house proxy server, and therefore theapproved users WHOwillsolely submit queries. knowledge|theinfo|the information} owner exports the hot and bothered data to the cloud. The approved users will submit vary queries or kNN queries to {findout|to be told} statistics or find some records. The RASP-perturbed knowledgeareaccustomed build indices to support question processing:Terms concerned in design are:

RASP Perturbation: RASP perturbation could be a novel combination of order conservingcryptography (OPE), dimension growth, random noise injection, and random projection. Let's think about the three-dinformationsquare measure numeric and in multidimensionalvectorhouse. The info has d searchable dimensions,which can be employed in queries, and n records, that makes a d × n matrix X. Let x represent a d-dimensional record, x ∈ R d. Note that within the d-dimensional vector house R d, a spreadquestion is pictured as AN intersection of [*fr1]-space functions and a spreadquestion is translated to finding the purpose set in corresponding solidspacerepresented by the half areas. in an exceedinglytraditional setting, the searchable dimensions are indexed with techniques like R-Tree for quickquestionprocess.

Two-Stage Fast Range Query Processing: Because the OPE transformation is often non-linear, an indoorvaryoutlined by half-space conditions is reworked to a nonlinear manifold with some unknown form. However, we'vetested that the form is broken-backed, thatpermitsU.S. to expeditiouslynotice its bounding box. Therefore, to use the subsequent two-stage process strategy to expeditiouslynotice the question results. Specifically, the proxy within theshopperaspect finds the most bounding box (MBR) of the form submitted reworked query), then submits the MBR and a group of reworkedquestion conditions to the server. The server uses the multidimensional tree index to search out the set of records fenced in by the MBR, thatarea unit then

filtered by the conditions.Theresult'sthe preciseresults of the varyquestion, thatconsiderably reduces the post-processing value that the proxy server has to take.

KNN-R Query Processing: The kNN-R rule uses sq. ranges round thequestionpurposeto seek out the candidate nearest records. The inner sq.vary starts from the questionpurpose and expands till k points square measureenclosed. the precisekNN result ought to be within the bounding sphere of the inner vary, thatsuccessively is approximated by the bounding box of the sphere. The on top of figure shows the state of affairs of finding the candidate set for a 3-NN questionsupportedsq.ranges.The inner varyenlargementis achieved by a binary vary search rule. The user will set the initial outer sq.vary with an explicit distance from the questionpurpose.Inevery iteration, the rule finds the centervary between the inner varyand therefore the outer vary, within which if the amount of boxed points is larger than k, the outer vary is replaced by the center range; otherwise, the inner vary is replaced by the outer vary. This repetitivemethodwill exponentially scale back the search vary and notice the result quickly. The records within the final varyis distributed back to the shopper for final kNN filtering. Note that this method utilizes the linear property of the remodeled queries to derive the queries for the centervary, thatdoesn'tneed the client's participation

V. CONCLUSION

The proposed system is 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. 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.Thusalgorithm focuses to further improve the performance of query processing for both range queries and kNN queries and formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality

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. 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)

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

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

[8]. 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.

[9]. K. Chen and L. Liu, "Geometric data perturbation for outsourced data mining," Knowledge and Information Systems, 2011.

[10]. M. L. Liu, G. Ghinita, C. S.Jensen, and P. Kalnis, "Enabling search services on outsourced private spatial data," The International Journal of on Very Large Data Base, vol. 19, no. 3, 2010