ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING

National Conference on Advanced Innovation in Engineering and Technology (NCAIET-2015)

Alva's Institute of Engineering and Technology, Moodbidri

Vol. 3, Special Issue 1, April 2015

A Review Paper on RASP Data Perturbation for Confidential and Efficient Queries in the Cloud

Jayalakshmi S[1], Harish kunder[2]

M.Tech Scholar, Department of Computer Science, Alvas Institute Of Engineering And Technology, Moodbidri, Karnataka, India[1]

Senior Assistant Professor, Department of Computer Science, Alvas Institute Of Engineering And Technology, Moodbidri, Karnataka, India[2]

Abstract: Cloud computing is an emerging technology, Privacy and confidentiality has become the major concern in the public cloud. Data owners do not want to move their data to the cloud until and unless the confidentiality and the query privacy are preserved. On the other hand a secured query services should provide efficient query processing and reduce the in-house workload to get the total benefit of cloud computing. This paper presents a Random space perturbation method to provide secure and efficient range query and K nearest neighbor query services for protecting 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. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the kNN queries.

Keywords: Privacy, Confidentiality, Range query, Knn query.

I. INTRODUCTION

Cloud computing is an emerging technology. Posting data-intensive query services in the cloud is increasingly popular because of the unique advantages in scalability and cost saving. With the cloud infrastructure data owner can scale up or down the services and pay as they use the server. Work load of query services in the public cloud is highly dynamic and it will be expensive to serve the queries with the in-house infrastructure. However service provider may lose the control over the data privacy and confidentiality has become the major concerns. Curious service providers can possibly make a copy of the database or eavesdrop users' queries, which will be difficult to detect and prevent in the cloud infrastructures.

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.

To address these issues a criteria called CPEL is constructed, 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 [2] and order preserving encryption (OPE) [1] are vulnerable to the attacks. The enhanced cryptoindex approach [2] puts heavy burden on the in-house infrastructure to improve the security and privacy. The New Casper approach [8] uses cloaking boxes to protect data objects and queries, which affects the efficiency of query processing and the in house workload.

The random space perturbation (RASP) approach for 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 data sets 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 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. In summary, the proposed approach has a number of unique contributions:

The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and

DOI 10.17148/IJIREEICE

random projection which provides strong confidentiality guarantee.

The RASP approach preserves the topology of multi-dimensional range in secure transformation, which allows indexing and efficiently query processing.

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.

II. LITERATURE SURVEY

A. *Order preserving encryption*

Order preserving encryption technique[1], Encryption is a well-known technology for protecting sensitive data Integration of encryption method with database system cause performance reduction example if a column of contain sensitive information is encrypted, and is used in query predicate with a comparison operator an entire table scan would be needed to evaluate the query. Reason is that current encryption techniques do not preserve order and there data base indices such as B-tree cannot be used.
Order preserving encryption allows comparison operation to be applied directly on encrypted data without decrypting the operands MAX, MIN, and COUNT queries can be directly processed over encrypted data. “Groupby” and “order by” operations can also be applied. SUM, AVG and “group by” values need to be decrypted. A value in the column can be modified or a new value can be inserted in a column without requiring changes in the encryption of other values. OPES can easily be integrated with the existing database system as it has been designed to work with the existing indexing structure such as B-trees.

B. *Crypto index*

A Privacy-Preserving Index for Range Queries,[2]In the database-as-a-service (DAS) model since data is stored at the service provider, many security and privacy challenges arise. Most approaches to DAS define a notion of a security perimeter around the data owner. Environment within the perimeter is trusted (client/data owners) whereas environment outside the perimeter is not (service provider).

Query processing: Data is stored in an encrypted form outside the perimeter but accessed within. Split the query Q into two components Q_{sec}+Q_{insec} where Q_{insec} executes at the server on the encrypted representation to compute a result for Q and Q_{sec} executes within the security perimeter to filter out the false positive.

For the purpose of splitting the queries Bucketization approach is used, each of the bucket is identified by a tag. These bucket tags are maintained as an index and are utilized by server to process the queries.

C. *Drawback of OPE and crypto-index*

OPE and crypto-index assumes the attacker knows only the ciphertext. However, If the attacker has some prior knowledge, such as the attribute domains (maximum and minimum values), the attribute distributions, and even a few pairs of plaintext and ciphertext, these encryption methods will be vulnerable to attacks.

D. *RASP efficient multi-dimensional range queries*

RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases [3] Range query is the most frequently used query in online data analytics (OLAP) that requires the service provider to quickly respond to concurrent user queries. Most existing encryption based approaches require linear scan over the entire database, which is inappropriate for online data analytics on large databases. It was reported that maintaining data and supporting query-based services incur much higher cost than initial data acquisition.

RAndom SPace encryption (RASP) approach for efficient range query processing on encrypted data, assume the outsourced data are multidimensional data and thus the data records can be treated as vectors (or points) in the multidimensional space. The RASP method randomly transforms the multidimensional space, while preserving the convexity of datasets.

The framework assumes a secure proxy server at the client side that handles data encryption/decryption and query encryption. The data owner and authorized users submit the original data and queries to the proxy server; the proxy server then sends the encrypted data/queries to the service provider. The service provider is able to index the encrypted data and use it to efficiently process encrypted queries.

Drawback: client need to take care of most of the things like encryption, decryption, indexing etc

Private information retrieval

There is a significant risk to the privacy of the user, since a curious database operator can follow the user's queries and infer what the user is after. One thing a user can do to preserve his privacy is to ask for a copy of the whole database[4].

Same database is replicated at several sites, viewing the database as binary string $x=x_1,x_2.........x_n$ of length n. identical copies of strings are stored by k>=2 servers. The user has some index i, and he is interested in obtaining the value of bit $x_{i.}$ To achieve this goal, the user queries each of the server and gets replies from which the direct bit x_i can be computed.
Drawback: cost increases because of creating more than one copy of database.

DOI 10.17148/IJIREEICE

E. Nearest neighbor search with strong location privacy

Nearest Neighbor Search with Strong Location Privacy [5] Applications like GPS in mobile devices facilitate the location based services which is an emerging application in the wireless market. Special queries pose an additional threat to privacy because location of a query may be sufficient to reveal sensitive information about the queries. Location dependent queries may disclose sensitive information about an individual health, financial information, political affiliation etc. for example, user wishes to find the nearest restaurant, to get the information user may choose to transmit the query through an anonymous network that hides his/her IP address. It is not sufficient to hide IP address, if the informations like co-ordinates of queries and background knowledge is known then information can be easily be hacked.

The solution can be classified as Location obfuscation. Data transformation. Private information retrieval.

III. EXISTING SYSTEM

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. 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.

Disadvantages

Adversaries, such as curious service providers, can possibly make a copy of the database or eavesdrop user's queries, which will be difficult to detect and prevent in the cloud infrastructures.

IV. PROPOSED SYSTEM

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 2 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 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. The key components in the RASP framework include (1) the definition and properties of RASP perturbation; (2) the construction of the privacy-preserving range query services; (3) the construction of privacy-preserving kNN query services; and (4) an analysis of the attacks on the RASP-protected data and queries.

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.

Challenges and expected outcome

Challenges and the expected outcome in creating the secured cloud environment is as follows,

Challenges

Preserving data confidentiality: the main challenge is to preserve the privacy and confidentiality for the data.

Should not provide slow query service as a result of security & privacy assurance: slow response should not be provided as a result of security as it increases the uses bandwidth.

Bandwidth should be decreased.

Expected outcome

A secured query service: the expected outcome is to provide an environment that makes the data owner to feel that his/her data is secured in the cloud and bandwidth is reduced.

V. STATEMENT OF PROBLEM

The attractive feature of cloud infrastructure like convenient scale up or scale down, pay as we use has made cloud popular.

As the workloads are dynamic and service providers lose the control over the data, it is important to protect data.

Some curious service providers can store a copy of database or eavesdrop users queries which will be difficult to detect and prevent in cloud infrastructure because of this data owners does not want to move to the cloud unless the data confidentiality is preserved.

Objectives

The main objective is to design a structured approach for providing a secured environment for data owners and ensuring the data confidentiality avoiding coping od database or eavesdrop from curious service providers.

DOI 10.17148/IJIREEICE

Methodology

The methodology has four modules

A. User Module :

In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

B. Multidimensional Index Tree :

Most multidimensional indexing algorithms are derived from R-tree like algorithms , where the axis-aligned minimum bounding region (MBR) is the construction block for indexing the multidimensional data.

For 2D data, an MBR is a rectangle. For higher dimensions, the shape of MBR is extended to hyper-cube. the MBRs in the R-tree for a 2D dataset, where each node is bounded by a node MBR. The R-tree range query algorithm compares the MBR and the queried range to find the answers.

C. Performance of kNN-R Query Processing :

In this set of experiments, investigating several aspects of kNN query processing.

1) studying the cost of (k, δ)-Range algorithm, which mainly contributes to the server-side cost.
2) showing the overall cost distribution over the cloud side and the proxy server.
3) showing the advantages of kNN-R over another popular approach: the Casper approach for privacy-preserving kNN search.

D. Preserving Query Privacy :

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. Focusing on the efficiency side of PIR, Williams et al. use a pyramid hash index to implement efficient privacy preserving data-block operations based on the idea of Oblivious RAM. It is different from our setting of high throughput range query processing.

Hu et al. addresses the query privacy problem and requires the authorized query users, the data owner, and the cloud to collaboratively process kNN queries. However, most computing tasks are done in the user's local system with heavy interactions with the cloud server. The cloud server only aids query processing, which does not meet the principle of moving computing to the cloud.

Architecture

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 maintaining confidentiality.

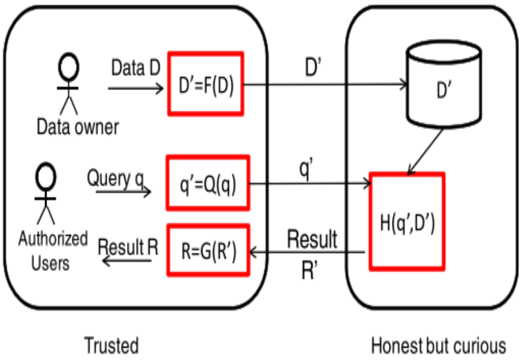


Fig 1: The system architecture for RASP-based query services

The system architecture contains following steps,

A. Exporting the perturbed data

Initially the data D is transformed D^1using the transformation function F, D^1=F(D,K) and the perturbed data is sent to the database.

B. Multidimensional Index Tree

The perturbed data D^1 is used for indexing, indexes and the perturbed data is stored in the database.

C. Transforming the user query

The user query Q is transformed using the function q^1=Q(q) and the query is sent to server for processing.

D. Appling the two-stage query processing at the receiver side

The query is processed using the query processing algorithm H,H(q^1,D^1). Result R^1 is sent to sender, sender transform the result R^1 using the function R=G(R^1).

VI. SYSTEM DESIGN

UML has intentionally been developed as a langue for modeling object- oriented systems. Its use has however widely been spread out. Today UML is used for system specifications. In different domains UML is used for specification and standardization of different systems or parts of the systems.

This standardization makes it possible that different vendors produce products that comply with the standard specification. From UML it is possible to automatically generate different types of descriptions (for example specifications in XML), or even automatic creation of software code. UML is becoming a standard tool for software and system engineers.

A. Data flow diagram

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated.

DOI 10.17148/IJIREEICE

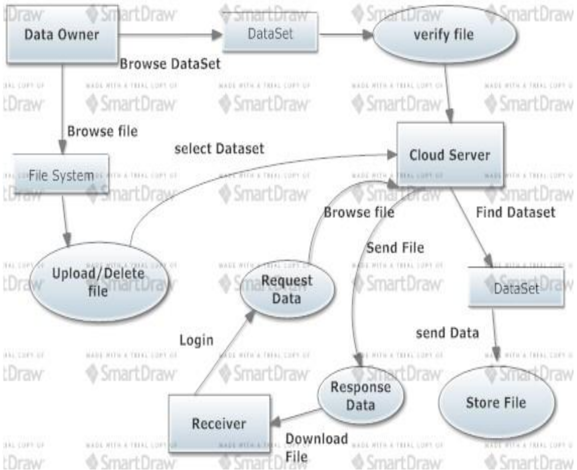
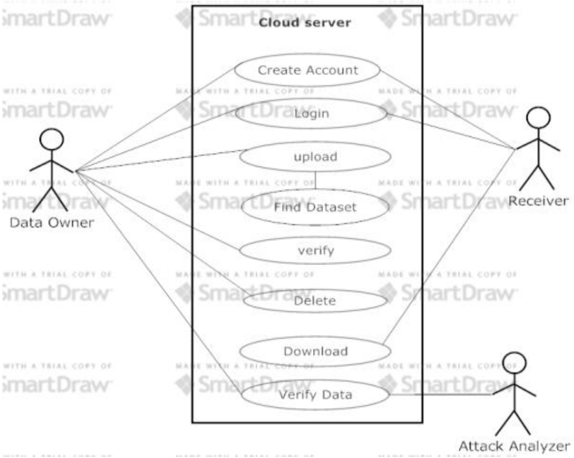


Fig 6.1: Data Flow Diagram

Intially, Data owner will register to the cloud server, he can browse the file and upload the file with file key. Data owner can verify the files in the data set to check whether it has been attacked or not. Receiver will register to the cloud server and when ever they need a file to download they will request the data by browsing the file and get the response.

B. Use case diagram

Use case diagrams are a set of use cases, actors and their relationships. They represent the use case view of a system. A use case represents a particular functionality of a system.



Use case diagram

In the use case diagram there are 3 actors namely Data owner, receiver and Attack analyzer, Cloud server act as a sub system, and there are eight use cases. The actors Data owner and Receiver can create the account and login to the account. Data owner will select the cloud server and data set and upload the file. Receiver will download the file by providing the secret file key.

Attack analyzer will periodically check whether the dataset is healthy or attacked.

VII. ALGORITHMS

Algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

K nearest neighbor algorithm

For nearest neighbor queries Knn algorithm is used. Knn is a non parametric method used for classification and regression.

Random Space perturbation method

Random space perturbation (RASP) approach to constructing practical range query

Other algorithmsd
SHA1 - Secure Hash Algorithm for Digital Signature.
RSA - For Secret Key Generation.
AES - Cryptography technique(For Data Encryption and Decryption)

VIII. CONCLUSION

Using Random Space Perturbation and K Nearest Neighbor method it is possible to process the query quickly and in the other words we can reduce the cost of the cloud usage and cloud provider can provide an secured environment for the data owner.

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

[3]. K. Chen, R. Kavuluru, and S. Guo, "RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases," Proc. ACM Conf. Data and Application Security and Privacy,pp. 249-260, 2011.

[4]. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private Information Retrieval," ACM Computer Survey, vol. 45, no. 6, pp. 965-981, 1998.

[5]. S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest Neighbor Search with Strong Location Privacy," Proc. Very Large Databases Conf. (VLDB), 2010.

[6]. Marimont and M. Shapiro, "Nearest Neighbour Searches and the Curse of Dimensionality," J. Inst. of Math. and Its Applications, vol. 24, pp. 59-70, 1979.

[7]. W.K. K, D.W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure KNN Computation on Encrypted Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 139-152, 2009.

[8]. M.F. Mokbel, C. yin Chow, and W.G. Aref, "The New Casper: Query Processing for Location Services without Compromising Privacy," Proc. 32nd Int'l Conf. Very Large Databases Conf. (VLDB), pp. 763-774, 2006.