Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 50 (2015) 122 – 127

Procedia Computer Science

Building an Authentication and Quality of Query Services in the Cloud

J.Sunitha[a], Mrs. A.M.Sermakani[b]

[a]PG Student, S. A. Engineering College, Computer Science and Engineering, Chennai, India
Email: sunithaj04@gmail.com
[b]Assistant Professor, S. A. Engineering College, Information and Technology, Chennai, India
Email: sermakani@gmail.com

Abstract

Cloud outpouring is careful when an information distributor has given sensitive data to a set of trusted agents and few of the information is leaked and found in an unauthorized place. An enterprise data leak may be a scary proposition. Security practitioners always deal with data cloud leakage issues that arise from various ways like e-mail and different net channels. In case of information cloud leakage from trusted agents, the distributor should assess the probability that the leaked information came from one or more agents.

The proposed system can identify those parties who are guilty for such cloud leakage even once the data is altered. For this the system will use data allocation ways are also can inject "realistic but fake" information records improve the identification of cloud leakage. Moreover, data can also be leaked from inside an organization through e-mail. Hence there's also a need to filter these e-mails, may be done by blocking e-mails that contain pictures, videos or sensitive data in an organization. A principle utilized in e-mail filtering is classify e-mail based mostly the fingerprints of message bodies, the white and black lists of mail addresses and the words specified to spam.

Keywords: Sensitive data; Unauthorized access; E-mail filtering; Query privacy.

1. Introduction

Nowadays Peoples use cloud computing infrastructure because of its unique features in scalability and cost-saving. People use cloud because of its effective features such as security, infinite storage, low cost and multiple users can access the files applications in the cloud infrastructure. The owners in the cloud only pay for the time of using the server. This is an important feature because the workload of query services in the cloud is highly dynamic and it is dynamic. Even though the security is increased, the data confidentiality and query privacy are the major issue in cloud. The organization information's are also leaked through the e-mail within the organization.

To secure the data and query privacy new approaches are needed in the cloud. But it is not an advantage, if the new approaches for providing security and privacy will provide the slow query processing. The CPEL criteria are examined for submitting a query in the cloud. A CPEL criterion denotes confidentiality of data, query privacy, efficiency in query processing and low in-house working cost. This method increases the complexity of query services. Some related methods have been elaborated to identify some particular aspects of the problem. But they do not identify all these aspects. For example, the Order Preserving Encryption (OPE)[1] and crypto-index[8] methods are

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
doi:10.1016/j.procs.2015.04.073

vulnerable to the attacks. Enhanced crypto-index method[9] gives heavy load on the in-house infrastructure to improve the data security and query privacy. New Casper approach[13] uses cloaking boxes to secure data objects and query, which affects query processing efficiency and the in-house workload.

We study techniques for detecting leakage of a set of records. This paper develops a sample that can be estimated the "guilt" of agents. We also introduce algorithms for distribution of objects to agents that improves the chances of identifying leakier. This system also considers the method of adding fake objects to the distributed set. If it turns out an agent was given one or more fake objects that are leaking, then the distributor is more confident that the agent was guilty. The proposed system provides a method for calculating the guilt probabilities in case of information leakage. The next part provides a method for data allocation agents. Finally, evaluate the methods in different data leakage scenarios, and check whether they help to identify the leaker.

The proposed Random Space Perturbation (RASP) method is used to construct the practical range query and k-nearest-neighbor (kNN) query services in the cloud. The proposed system will satisfy all the four aspects of CPEL criteria. RASP method also transforms the multidimensional data with the combination of order preserving encryption, random noise injection and random projection. The RASP approach and its combination provide the data confidentiality and protect the multidimensional range of queries and efficiently processing the query with indexing. The range query is used to retrieve the stored data from the database. It uses the upper and lower bounds to retrieve the data. K-nearest-neighbor query is to find the nearest record to the query point.

The proposed system also uses the e-mail filtering method to filter the mail that sends to the unauthorized user. The main aim of this proposed work is to identify the guilty agents.

1.1 Organization

This paper has organized into VI sections. Section II gives literature review of the work done related to data security. Section III describes about the proposed system. Section IV defines the algorithm used in the proposed system. Section V gives the architecture of the proposed method. Finally, section VI gives a conclusion of the work.

2. Related work

The following related works are referred for preparing proposed work.

2.1 Order Preserving Encryption

OPE represents Order Preserving Encryption is used for data that allows any comparison. And that comparison will be applied to the encrypted data; this will be done without decryption. It allows database indexes to be built over an encrypted table. The main drawback of this technique is the encryption key is too large and the implementation takes more time and space.

2.2 *Crypto-index method*

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 the New Casper approach is used to protect data and query but the efficiency of the query process will be affected.

2.3 Distance recoverable encryption

This method used for preserving the relationship between the nearest neighbors. Many attacks can be applied because the distances are exactly preserved[17, 10, 5.] Wong et al.[17] suggest that dot products are preserved instead of distances to find k- nearest neighbor which is more resilient to distance-targeted attacks. The drawback of DRE is, in

linear scan the search algorithm is limited and also there is no indexing methods are used.

2.4 Preserving query privacy

Papadopoulos et al.[20] use PIR methods to enhance the privacy of the location. However, their technique does not consider protecting the data confidentiality. Space Twist[19] proposes an approach to query kNN by providing a duplicate user's location for preserving the privacy of the location. But this method does not consider the confidentiality of data. The Casper method considers both the query privacy and data confidentiality.

2.5 Multidimensional range query

It requires the owner of the data provides the indices and keys for the server, and only the authorized users can access the data on the server. While in cloud database, the cloud servers are taking more responsibilities of indexing and processing the query.

3. Proposed system

The main aim of this work is to identify the parties who are guilty for cloud leakage and protect the sensitive data from unauthorized access. It also provides the confidentiality of the data and the query privacy. The RASP method is used to provide the confidentiality and the query privacy. The proposed system also uses e-mail filtering method. This method filters the leaked information by blocking mails that contains the sensitive data or videos, pictures in an organization.

The following are the modules of the project, which is planned in aid to complete the project with respect to the proposed system, while overcoming existing system and also providing support for the future enhancement. It consists of six modules.

3.1 Query authentication

In Server Section authorized clients are added with respective MAC & IP address. In this section server maintains the log of all query processed. RASP Implementation will be controlled in this section. Detected Clone Node logs are maintained in this section. It is done by the administrator. Here every client will give their IP address and MAC address for registration. Authenticated client only can able to transfer the data. Every authorized client should have an individual IP address and MAC address. This specific address is used to identify the clone node detection.

3.2 Securing the data with false objects

The server node will send some fake data in addition to the original data. The clone node will be unaware of these fake data. Only the owner of the node knows where and how many fake objects inserted into the original data.

3.2.1 E-Random implementation

The agent receives the entire data object that satisfies the condition of the agents' data request. If the explicit data request with fake allowed, then the data distributor cannot modify or remove the requests from the agents. However distributor can insert the fake object. The e-optimal algorithm reduces each term of the objective summary by adding the maximum number of fake data to every set giving optimal solution. The E-optimal solution is

$O(n+n2F) = O(n2F)$

Where, n= number of agents

F= number of fake objects

3.2.2 S-Random implementation

In this module the extra data object the agents request in total, the more recipients on average an object has; and the more objects are used among different agents, the more difficult it is to discover a guilty agent. In this algorithm, the agent receives only the subset of data object that can be given to the agent. The working of the Sample Data Request algorithm is same as the working of Explicit Data Request.

3.3 High level query processing

Private Information Retrieval (PIR) tries to keep the privacy of the data access pattern, while the information may not be encrypted. This PIR scheme is normally very costly. The efficiency side PIR, uses a pyramid hash index to perform efficiently in privacy preserving data-block execution based on the conception of Oblivious RAM. It is different from the setting of high throughput, range query processing. It addresses the query privacy issues and requires the authorized query users, the owner of the data, and the cloud to the collaborative process kNN queries. However, most computing operations are performed in the user's local system with strong interaction with the cloud server. The cloud server only gives query processing, which does not meet the fundamental of moving computing to the cloud.

3.4 Distribution of sensitive data

A data distributor has given sensitive data to the set of supposedly trusted agents. Some of the data are leaked and found in an unauthorized place. The distributor must assess the probability that the leaked data came from one or more agents, as opposed to having been separately gathered by other means.

3.5 Clone node formation

Clone node is detected once the client sent the data to the unauthorized person. The clone node is not aware about the fake objects created by the server.

3.6 E-mail filtering with organization sensitivity

This module filters the e-mail data which are sent to the clone node. This module involves 6 steps.

- Identify the data.
- Remove stopping words such as this, is, a, etc.
- Remove or change the synonyms.
- Determine the priority of the word depending upon the sensitivity of the data.
- Compare data with predefined company datasets.
- Filter the data if it has company's important data sets.

4. Proposed protocol

4.1 RASP

It provides the protection for the privacy of query services. The proposed Random Space Perturbation (RASP) method is used to construct the practical range query and k-nearest-neighbor (kNN) query services in the cloud. The proposed system will satisfy all the four aspects of CPEL criteria. RASP method also transforms the multidimensional data with the combination of order preserving encryption, random noise injection and random projection. The RASP approach and its combination provide the data confidentiality and protect the multidimensional range of queries and efficiently processing the query with indexing. The range query is used to retrieve the stored data from the database. It uses the upper and lower bounds to retrieve the data. K-nearest-neighbor query is to find the nearest record to the query point.

4.2 kNN-R algorithm

kNN is the k- nearest neighbor algorithm. This algorithm finds the nearest k samples in the spherical range, which is nearest to the query point. The kNN algorithm uses the following steps to find the nearest neighbor.

- Select the parameter k= number of nearest neighbors.
- Find the distance between the elements and the query point.
- Sort the distance.
- Determine the nearest neighbor having minimum distances within the parameter k.
- Select the majority of the category of nearest neighbors.

5. System implementation

Here we focus on the design phase of the proposed system. It describes the overall view of this work. The design phase provides the clearest view about how the proposed system provides confidentiality and security of the data. It also describes how the system is to filter the sensitive data from the unauthorized user. This phase helps the user easily to understand about the project. The architecture consists of six modules that are explained previously.

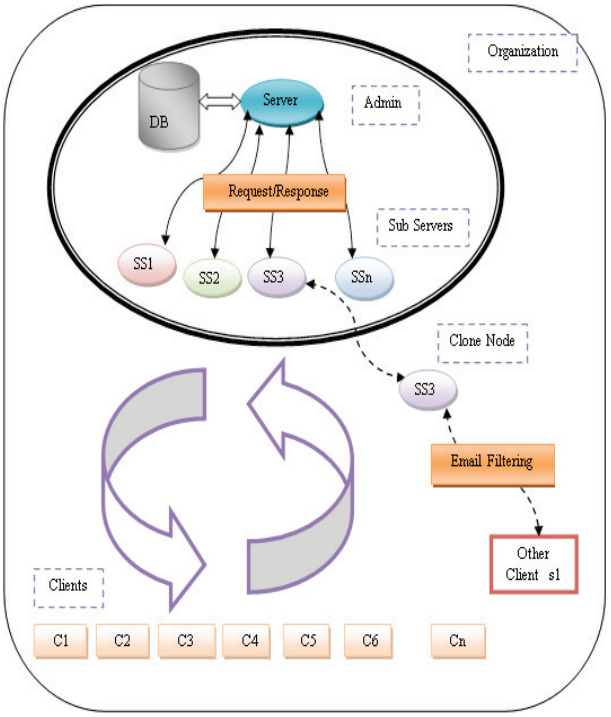

Fig. 1. The System Architecture

Fig.1 describes the structure of the proposed system. The server will maintain all the login of the system. The clients and sub servers are registered in the cloud to get the authorization to access the cloud information. During the registration the server automatically added the IP and MAC address of the user. These addresses are used to detect the clone node. The users or sub servers send the request to the server to access the cloud information. After getting the request, the server checks the database, whether the users or sub servers are the authorized party or not. The server creates the fake objects based on the sub server request. And then the server sends the response to the user or sub server. Only authorized user can able to transfer the data.

Once the clone node is detected, the e-mail filtering method is used to filter the e-mail from the clone node. The principle used in e-mail filtering methods is blocking the information that contains company's videos, pictures etc. The admin can view all the data transformation and admin gets the e-mail alerts.

6. Conclusion

We proposed RASP approach, which satisfies CPEL criteria that is confidentiality of data, privacy of query, efficient query processing and low in-house workload. This approach also identifies which part of intermediate data sets need to be encrypted in order to save the privacy preserving cost. It also reduces the privacy preserving cost comparison with existing approaches. RASP perturbation is the combination of order preserving encryption (OPE), dimensionality expansion, random projection and random noise injection, which provides the security feature. The proposed method increases the identification of cloud leakage and provides the security to the cloud data. The e-mail filtering method is used to protect the sensitive data from unauthorized access.

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. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in *INFOCOMM*, 2011.
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. K. Chen and L. Liu, "Geometric data perturbation for outsourced data mining," *Knowledge and Information Systems*, 2011.
5. K. Chen, L. Liu, and G. Sun, "Towards attack-resilient geometric data perturbation," in *SIAM Data Mining Conference*, 2007.
6. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *ACM Computer Survey*, vol. 45, no. 6, pp. 965–981, 1998.
7. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable 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.
8. 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.
9. B. Hore, S. Mehrotra, and G. Tsudik, "A privacy-preserving index for range queries," in *Proceedings of Very Large Databases Conference (VLDB)*, 2004.
10. F. Li, M. Hadjieleftheriou, G. Kollios, and L. Reyzin, "Dynamic authenticated index structures for outsourced databases," in *Proceedings of ACM SIGMOD Conference*, 2006.
11. K. Liu, C. Giannella, and H. Kargupta, "An attacker 's view of distance preserving maps for privacy preserving data mining," in *Proceedings of PKDD*, Berlin, Germany, September 2006.
12. R. Marimont and M. Shapiro, "Nearest neighbour searches and the curse of dimensionality," *Journal of the Institute of Mathematics and its Applications*, vol. 24, pp. 59–70, 1979.
13. M. F. Mokbel, C. yin Chow, and W. G. Aref, "The new Casper: Query processing for location services without compromising privacy," in *Proceedings of Very Large Databases Conference (VLDB)*, 2006, pp. 763–774.
14. S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbor search with strong location privacy," in *Proceedings of Very Large Databases Conference (VLDB)*, 2010.
15. E. Shi, J. Bethencourt, T.-H. H. Chan, D. Song, and A. Perrig, "Multi-dimensional range query over encrypted data," in *IEEE Symposium on Security and Privacy*, 2007.
16. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2010.
17. P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage," in *ACM Conference on Computer and Communications Security*, 2008.
18. W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in *Proceedings of ACM SIGMOD Conference*. New York, NY, USA: ACM, 2009, pp. 139–152.
19. M. L. Yiu, C. S. Jensen, X. Huang, and H. Lu, "Space twist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services," in *Proceedings of IEEE International Conference on Data Engineering (ICDE)*, Washington, DC, USA, 2008, pp. 366–375.
20. P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *EUROCRYPT*. Springer-Verlag, 1999, pp. 223–238.