International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 09 | Dec-2015 www.irjet.net p-ISSN: 2395-0072

A Framework for Privacy preserving of Intermediate data sets in cloud

Sujeet Shinde[1],Swapnil Bhand[2], Karan Gophane[3], Prof. Lagad Jyoti U.[4]

[1] *Student, Shri Chhatrapati Shivaji Maharaj College of Engineering,Ahmednagar.*
[2] *Student, Shri Chhatrapati Shivaji Maharaj College of Engineering,Ahmednagar.*
[3] *Student, Shri Chhatrapati Shivaji Maharaj College of Engineering,Ahmednagar.*
[4] *Assistant Professor, Shri Chhatrapati Shivaji Maharaj College of Engineering,Ahmednagar.*

Abstract - ***In recent year the paradigm of cloud computing has gained very appalling concept for both provider and user .cloud computing can save huge capital investment of IT industry and concentrate on their own core business .cloud computing provide massive computation power and storage capacity which enable user to display computation and data intensive application without infrastructure investment ,because of that many company or Organization have been migrating or building their business into cloud. However numerous potential customer are still hesitant to take advantages of cloud due to security and privacy concern .This drawback of cloud are somewhat overcome by providing security to cloud using encryption and decryption, but nowadays the application which providing an security to the cloud are en/decrypt all the data or information at the time of storage or retrieval to/from the cloud. This technique is neither efficient nor cost effective ,because it is very time consuming and costly for data intensive application to en/decrypt data sets frequently while performing any operation on them. we provide an framework for privacy leakage upper bound constraint based on the identifying which intermediate dataset need on the cloud need to encrypt or which not.so that it help to save the time as well as cost of privacy preserving ,so that data holder can still be satisfied . Evaluation results demonstrate that the privacy preserving cost of intermediate data sets can be significantly reduced with our approach over existing ones where all data sets are encrypted.***

Key Words: *Cloud computing, data storage privacy, privacy preserving, intermediate data set, privacy upper bound.*

1. INTRODUCTION

Cloud computing is regarded as an ingenious combination of series of technology .It establishing a novel approach by offering IT services .cloud computing provide massive computation power and storage capacity which enable user to deploy computation and data-intensive application without investment. Therefore many company or organization has been migrating or building their business into cloud .however numerous potential customer are still hesitant to take an advantages of cloud due to its security problem and privacy concern.

Providing a security and privacy to the cloud is very challenging and essential task in recent days. Some existing application which provide security to data on cloud they are work only on the non-encrypted data, these system en/decrypt all the intermediate data sets at the time of retrial or storing it to/from database of cloud using third party encrypted ,which is charge in proportion to their usage. At the time of retrial of information all the intermediate data sets will be decrypted totally , however the storage of intermediate datasets enlarge attack surface ,so that privacy requirement of data holders are at risk of being violated .Usually ,intermediate data set in cloud are accessed and processed by multiple parties but rarely control by original data sets holders .This enables an adversary to collect intermediate datasets together and menace privacy-sensitive information from them ,brining considerable economies loss or several social reputation important to data owners . But, little attention has been paid to such a cloud-specific privacy issue.

Existing technical approach for preserving the privacy of data sets stored in cloud mainly include encryption and anonymization. On other hand encrypting all the data sets, a straightforward and effective approach, is wildly adopted in current research. However processing

on encrypted data is challenging task because most existing application only runs on unencrypted data sets. Current privacy-preserving techniques like generalization can withstand most privacy attacks on one single data sets, while preserving the privacy for multiple data sets is still a challenging problem .Thus, for preserving the privacy of multiple data sets is promising to anonymize all data set first and then encrypt them before storing or sharing them in cloud. Usually the volume of intermediate data sets is huge. Hence, we argue that encrypting all data sets lead high overhead and low efficiency when they frequently accessed or processed.

In this paper, we propose a novel approach to identify which intermediate datasets need to encrypted while other do not ,in order to satisfy privacy preserving requirement of data holders .finally we design a practical heuristic algorithm according to identifying data set that need to be encrypted. First ,we formally demonstrate the possibility of ensuring privacy leakage requirement without encrypting all intermediate data when encrypting is incorporated with anonymization to preserve the privacy .second ,we design a practical heuristic algorithm to identify which data set need to encrypt and which not .third, experimental result demonstrate that our approach can significantly reduce privacy preserving cost over existing approaches which is quite beneficial for cloud users who utilize cloud service in a pay-as-go fashion . We mathematically prove that our approach can ensure privacy-preserving requirement. Further, the heuristic algorithm is design considering more factors. We extend experiments over all real data sets.

2.Related Work

Currently encryption is exploited by most existing research to ensure the data privacy in cloud Although encryption work well in data privacy approaches ,it is necessary to encrypt and decrypt the data sets frequently in a many application. But now a days so many application are work only on the non-encrypted data to provide the security on the cloud investigated the data privacy problem cause by map-reduce and presented a system name *Airavat* which incorporate mandatory access control which differential privacy. *Ciriani* et al .proposed an approach that combine encryption and data fragmentation to achieve privacy protection for distributed data storage with encrypting only part of data sets . we follow this line but integrate data anonymization and encryption together to fulfill cost-effective privacy preserving .

The importance of retaining intermediate data sets in cloud has been widely recognize ,but the research on the privacy issue incurred by such a data sets just commences .*Davidson* et al . Studied that privacy issues in workflow province and high utility of provenance information via carefully hiding a subset of intermediate data. This general idea is similar to ours, yet our research mainly focus on data privacy from an economical cost perspective while their concentrate majorly on functionality privacy of workflow models. Our research is many way different from the existing system which provide the privacy and security on cloud, i.e. unlike to the existing system our system work on the encrypted as well as decrypted data sets, in does not require to encrypt all the data while storing in cloud database .so that it will reduce the doth the time and economical cost of en/decryption of intermediate data sets.

3. Propose Work

One of the most biggest drawback of the cloud is it has very less security ,so providing the security on the cloud is most important and challenging task . Many existing application work on this issue but some of them work on only non-encrypted data sets. We propose novel approach for security and privacy preserving of intermediate data set on cloud .When data provider send data on the cloud it will accept by third party Encrypted which encrypt the data sets and stored that on the cloud .at the time of receiving or accepting the data from that intermediate data set are retrieve from the database and given to the third party encrypted to decrypt that intermediated data sets, after decrypting the intermediate data sets that intermediate data sets are not re-encrypted again because of this another user or server administrator can use that data and modify the data sets .

Consider an example of database which contains large amount of information about law and order (crime) when the data provider save or upload the data about the law cases in the cloud for economic benefit .

Original data sets are encrypted for confidentiality by using third party encrypter. At the time of retrieving data intermediate data sets will decrypted and send to the data requestor, this decrypted data can be access by other user also like government Canter .i.e. court , police department ,etc ,or Adversary.

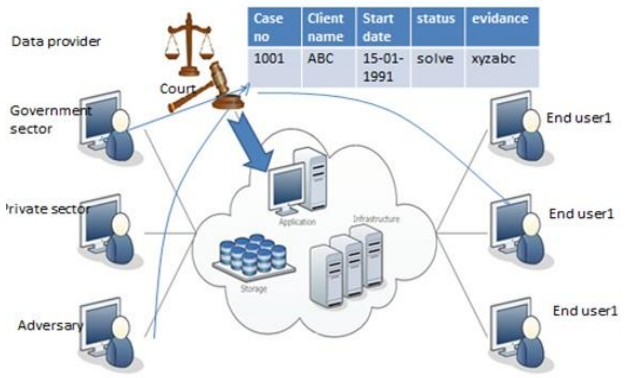

Fig-1: scenario showing privacy threats due to intermediate data sets.

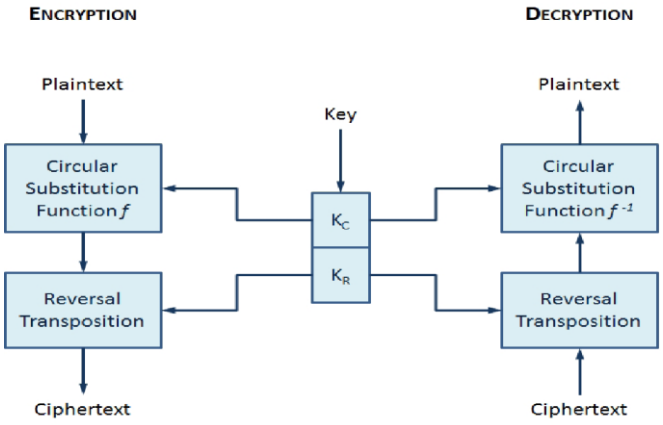

Fig-2: Simple Representation of the Reverse Circle Cipher

As shown in above fig data provider store the data on the cloud through third party encrypter this data can be access by other data provider, adversary as well as end users and they can modify the database about particular client. To avoid this scenario we are implementing the framework for privacy preserving for intermediate data sets in cloud. In this system when intermediate data sets are decrypted and given to the data requester. In our system at the same time upper bound constraint based plug-in approach work continuously and identifying the which intermediate data sets are decrypted and make tree structure of that decrypted data sets

As we have discussed in above section it is clear that the existing application that provide the security to the cloud is time consuming as well as costly. So that we are design frame work for privacy preserving intermediate data set in cloud .Which overcome the Drawbacks of existing application.

3.1 ARCHITECTURE OF THE CRYPTOSYSTEM

The simple model of the proposed model is as shown in Fig. 2. As it is a symmetric cipher, the same key is used for both encryption and decryption.

The input key is a tuple of the circular character key KC and reversal length integer key KR. During encryption, the circular substitution first takes place with the plaintext and the circular key as input. This output of this operation goes through reversal transposition with the reversal length. The decryption process is the reverse of the encryption process. The reversal algorithm is the same while the circular substitution function is the arithmetic converse of the function used for the encryption process. Thus through successful operation, the plaintext obtained after the decryption and before the encryption will be the same.

4. CONCLUSIONS

In this paper, we have proposed an approach that identifies which part of intermediate data sets needs to be encrypted while the other does not, in order to privacy preserving cost. A tree structure has been modeled from the generation relationship of intermediate data sets to analyze privacy propagation among the data intermediate datasets. We have illustrate the problem which is addressed by decomposing privacy leakage constraints. A practical heuristic algorithms has been designed according to problem. Evaluation results on real world data sets and large extensive data sets have demonstrated cost of preserving privacy in cloud can be reduced significantly with our approach over existing ones where all data sets are encrypted. In according to various data and computation intensive application on cloud, intermediate data sets management is becoming an important research

area. Privacy preserving of intermediate data set is one of the intensive investigation. With the help of this paper We are planning to the further investigate privacy aware efficient scheduling of intermediate data sets in cloud by taking privacy preserving as metric together with other metric such as storage and computation. Optimize balanced scheduling strategies are expected to be developed toward overall highly efficient privacy aware data sets scheduling.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud Computing ", comm. ACM, Vol. 53,no. 4 pp. 50-58 2010.

[2] R. Buyya, C.S Yeo, S. Venugopal, J. Broberg, and I. Brandic,"Cloud Computing and Emerging It Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility", Future Generation Computer System ,Vol.25,no. 6,pp.599-616,2009.

[3] L. Wang, J. Zhan, W. Shi, and Y. Liang,"In Cloud, Can Scientific Communities Benefit from the economies of Scale?," IEEE Trans. Parallel and Distributed System, Vol . 23, no. 2, pp. 296-303, Feb.2012.

[4] H. Lin and W. Tzeng ,"A Secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding", IEEE Trans. Parallel and Distributed Systems, Vol . 23, no. 6,pp. 995-1003, June 2012.

[5] M. Li, S. Yu, N. Cao, and W. Lou ,"Authorized Private Keyword Search Over Encrypted Data in Cloud Computing ", Proc. 31st IntIconf . Distributed Computing Systems,pp.383-392,2011.

[6] C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices ",Proc.41st Ann. ACM symp. Theory of Computing pp.169-178,2008.