International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 10, Issue 5 (May 2014), PP.01-06

Privacy Preserving Of Intermediate Data Sets In Cloud

[1]Mr C.Radhakrishnan PG scholar, [2] Ms. C.P.Darani
[1]Department of CSE/IT University College of Engineering BIT Campus, Tiruchirappalli
[2]Teaching Fellow, Department of CSE/IT University College of Engineering BIT Campus, Tiruchirappalli

Abstract:- Cloud Computing is a long-term dream of computing as a service has the portable to convert a huge part of the Information Technology Industry and organization without any investment of infrastructure. In cloud uses the more number of intermediate data sets generated, but still now preserving the privacy of intermediate datasets becomes a difficult achievement problem because adversaries may improve privacy secret information by analyzing multiple intermediary datasets. Encrypting all datasets in cloud is usually adopted in existing approaches pointed to this challenge. Encrypting all intermediate datasets are either well organized or gainful because it will take more time and costly for data demanding applications to encrypt and decrypt datasets regularly while performing any operation on the data set. In my project, propose a novel upper-bound privacy leakage constraint based approach to identify which intermediate datasets need to be encrypted or not, then the privacy-preserving cost to be saved while the privacy requirements of data holders to be Satisfied. That data can be transferred to the cloud only in an encrypted form, available only to users with the correct keys, that protecting its confidentiality against un-intentional errors and attacks**.**

Keywords:- Encryption and Decryption, Privacy Preserving, Intermediate Dataset, Privacy Upper Bound, Economics of scale.

I. INTRODUCTION

A. Cloud Computing:

Cloud computing is a new technology for to provide the hardware, software services to the clients without investment by using the internet. The cloud computing model enable the reliable and on demand network access to sharing the computing resources that can be quickly provisioned and released with minimum management effort or service provider interaction. The cloud environment gives option for to reduce the cost of the communication. The cloud computing services are mainly referred to three ways. First one is refers to the applications delivered as a services through the internet and the hardware and systems in the data centers that provide those services. The service referred to software as s Service (SaaS). Public cloud is referred as that cloud will be used by various clients in pay as you go manner. In this service is being to Utility Computing. The example of public Utility computing include Amazon Web Services, Google Application Engine and etc. The term private cloud refers to internal data center of a company or other business that are not made available to the public cloud. The cloud computing is a grouping of SaaS and Utility Computing, but does not include Private controls. If the clarity demands in the cloud computing then it is replaced by one of other terms. The advantage of the Software as a Service is well understood of the both user and service providers.

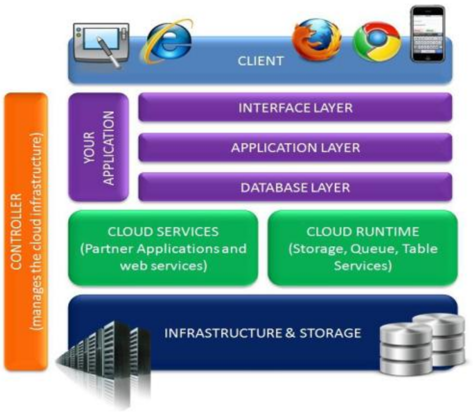
CLIENT
INTERFACE LAYER
APPLICATION LAYER
DATABASE LAYER
YOUR APPLICATION
CONTROLLER (manages the cloud infrastructure)
CLOUD SERVICES (Partner Applications and web services)
CLOUD RUNTIME (Storage, Queue, Table Services)
INFRASTRUCTURE & STORAGE

Fig 1: Cloud Computing Architecture

Service providers like significantly easy software installation, maintenance and centralized control over versioning. The end users allowed to the service anytime, anywhere, sharing the datas and manage more easily. The data are stored safely in the cloud environment infrastructure. Cloud computing provide more application to develop the product as SaaS but this does not change arguments. It is permit to setup SaaS and the level of order without structuring a data center.

Cloud computing is developing of a attractive grouping of a different technology, establish a novel business model by offer information technology services and using market of scale. This business chain is very useful for who are all involved in the cloud computing environment model. Cloud environment clients can save huge capital savings of information technology infrastructure, and focus on their own company. So that many industries or organizations are interested in their business or company into cloud environment. Moreover many customers are still struggle to take advantage of cloud because of security and privacy concerns. The privacy concerns reason by hold the intermediate data sets in cloud are essential but these data's are need attention. Storage and working out services in cloud environment are equivalent from an inexpensive point of view because they are paid the amount in proportion to their usage.

The cloud users can store selectively necessary intermediate data sets when using original data sets in data secret applications like medical analysis, in order to reduce cost by stopping regular recomputation to getting those data sets. This type of situation are common because data users frequently results, creating a analysis on intermediate data sets, or share their intermediary results with others for relationship. The idea of an intermediate data set is identified without loss of generality of intermediate data set. Still, the storage of intermediate data increase attack surfaces then privacy requirements of data owner are at risk of violated. Regularly intermediate data sets in cloud are processed by more no of users, but time to time controlled by own data set holders. This enables to gather intermediate data sets together and necessary privacy secret information present them and bring considerable gainful loss or severe public status damage to data owner. But, little awareness has been paid to a cloud for particular privacy problem.

Earlier approach is describes that the preserving the privacy of data sets in cloud generally includes encryption and identification. Encrypting all data set is a simple and effective approach and it is usually implemented in current research. But processing of encrypted data sets efficiently is difficult task, because most existing approaches only run on unencrypted data sets. While recent progress has been made in similar encryption which permits to performing computation on encrypted data sets, applying recent algorithms are somewhat costly due to ineffectiveness. Limited information of data users need to describe the data users in cloud like data mining and analytics. This type of cases, data sets are identified relatively encrypted to provide both data efficiency and privacy maintained. Current privacy-preserving techniques can withstand most privacy attacks on its own data set the preserving privacy for multiple data sets is till now a difficult problem. Thus, for preserving privacy of multiple data sets, it is promising to first identify data sets and then encryption done on those data's before storing it into cloud. Usually, the quantity of intermediate data sets is high. Hence, we go down out that encrypting all intermediate data sets leads to high operating cost. The intermediate data sets provide the low efficiency when they are accessed. So that planned to encryption is applied only an intermediate data set compare to all for reducing privacy-preserving cost.

II. PRIVACY LEAKAGE CONSTRAINTS

Privacy-preserving cost of intermediate datasets from regular encryption and decryption with charged by cloud services. Cloud service providers can have different amount of models to maintain pay and use model, e.g., Amazon Web Services pricing model.

Specifically encryption and decryption required working out power, data storage space and other cloud services. To avoid amount details and points to the discussion of our core ideas. In this approach combine the prices of different services required by encryption and decryption into one. The status of location-based, dynamic interaction a key distinction is also a possible supply of privacy leakage. The combination of location information, unique identifiers of devices, and traditional leakage of other PII all combine against protection of a user's privacy.

Protecting privacy on the Web is becoming more difficult because of the significant quantity of personal and sensitive information left by users in many locations during their Web browsing. The actions of third party sites that collect data, grouping information and create personal profiles of Internet users in order to offer free and personalized services. The most of people are not learned their information may be collected online.

III. CAUSES OF INFORMATION LEAKAGE

The problem of privacy protection for a user is that the user must know that which users to access the sensitive information during the privacy setting. Whenever the information's are outsourced to online social networks, with in this situation which users to may see the data and the data available to third parties. Once the opponent model is defined the correct privacy framework is created means to countable measure the privacy and the private information leakage. The assessment of the leakage that a given protocol gives to finds its fitness for a set of privacy requirements and that fix a privacy level for a cloud application dealing with sensitive signals.

Privacy preservation of the current situations look many challenges associated to the development of secure protocol. The development of the secure protocol that efficiently give the exact functionalities without delay of the provider capabilities to actual develop of its activities and arrange its services. In order to effectively promise privacy and assess the contact of a given main preserving protocol on the utility condition. In this paper to implement the privacy leakage upper bound constraint in the intermediate dataset. This is to achieve the data privacy preserving and to reduce the privacy cost of the data over the existing approach.

IV. SYSTEM ARCHITECTURE

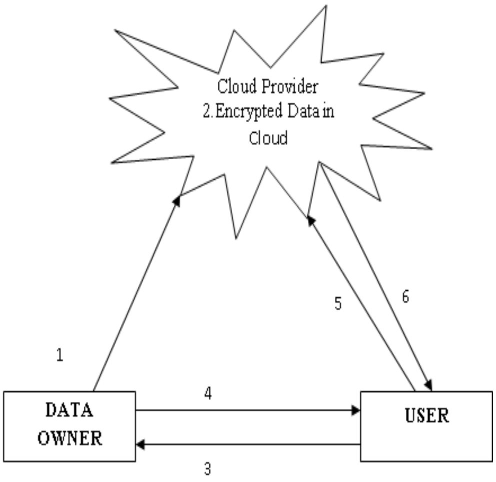

Fig. 2. System Architecture

Encryption is a main technique for data privacy in cloud. The data's are need necessary to encrypt and decrypt data sets regularly in many applications. Encryption is generally included with other technique to attain cost reduction, high data usability and privacy protection the data privacy problem. This type of problem reason by Map Reduce and presented which include compulsory access control with differential privacy that identifies all data to get module privacy preserving and high usefulness of origin information using carefully hiding a subset of intermediate data. The sensitivity of data is required to be making in advance to create the above approaches available that combines encryption and data fragmentation to attain privacy protection for distributed data storage with only the data sets are in encrypted format.

The following are the transition steps referred to above fig 2 System Architecture,

1. Data encrypted before in to cloud storage.
2. Encrypted data in cloud.
3. Request for an authentication.
4. Response for credentials.
5. Request to access data.
6. Responded with the encrypted data request.

Transition Steps Description:

- Initially the data's are present in the data owner. Once the data owner decide to share the data to who required the data from various users.
- The data is stored in the cloud environment in encrypted format.
- The data to be encrypted before the data outsourced in cloud. In cloud environment all the data's are present in encrypted format only.
- If the user need to access the data means the user need authentication.
- For authentication purpose the user send the request to the data owner.
- From the authentication request of the user is analyzed the data owner and then send the authentication permission to the user when the data owner satisfied the user.
- From the authentication permission itself the user receives the decrypted key of the cloud data.
- After getting the authentication permission of the data, the user sends the request to access the data in cloud with decrypted key.

- At last the cloud send the necessary encrypted data to the authorized user. After receiving data, the user decrypt the data.
- The decryption the process is done only the correct key known by user.

V. IMPLEMENTING APPROACH

A. Privacy-Preserving Cost Reducing Heuristic

The heuristic privacy preserving cost reduction algorithm idea is that the algorithm procedure iteratively choose a state node with the highest heuristic value and then expand its child state nodes until to reaches a goal state node.

Description: To identify which intermediate dataset is need to be encrypted with low privacy cost.

Input: A Sensitive Intermediate data set Tree with root node d0. All attribute values of each intermediate dataset is given.

Output: The global privacy cost.

```
Steps:
1. Initialize the following Attributes.
        a. Define Priority Queue.
        b. Construct the Initial Search node will the root of the SIT.
        c. Add the node to Priority Queue.
2. Retrieve the search node from Priority Queue.
        a. Retrieve the search node with the highest heuristic from Priority Queue.
        b. Check the Encrypted data set. If the encrypted dataset is null the it has a solution and goto step 3
        c. Label the datasets in CDE as encrypted if the privacy leakage is larger than $\varepsilon$.
        d. Generate all possible local solutions.
        e. Select a solution: $\pi \leftarrow$ SELECT (A)
             i. Calculate the privacy leakage upper bound of the above solution and encryption cost.
            ii. Calculate remaining privacy leakage.
        f. Calculate the heuristic value.
        g. Design the new search node from the obtained values and add to Priority Queue. Proceed the step 2.a
3. Obtain the global encryption cost.
```

VI. ENCRYPTION TECHNIQUES

Encryption is the process of convert the information into unreadable format. Here the source information referred as plaintext and the unreadable format information is referred as cipher text. The conversion techniques are implements with help of key. Encryption is mainly used to protect data while transmitting over the networks.

A. AES OFB Mode Techinques

AES OFB technique is a strange declaration model. A block cipher in OFB mode is generally a stream cipher which creates a key-dependent stream of pseudorandom bytes. Eencryption is done with XOR of a stream with the data to be encryption. OFB which is not requires no padding and not be used with less than full block feedback. In earlier days, OFB used a partial feedback and in input of the block chipper is the only part of output from the previous block. Partial OFB feedback is slower than ordinary OFB and was planned to support communication standard which lose synchronization. If the partial OFB feedback is weaker than full OFB then the partial OFB feedback is turned out. And the standard deprecated it.

This mode is a somewhat less common mode, quite similar to CFB. CFB does not use a flow of cipher due to natural fault when the data size doesn't match the block size of the causal encryption algorithm. The latency from side to side the OFB function can be reduce, as the only dealing out applied to the data is an XOR function.

Advantages and Disadvantages of OFB:

- Bit errors do not propagate in OFB.
- OFB to provide more vulnerable to message stream modification.
- It is shall not to be use or reuse the same sequence (key+IV).
- Both the sender & receiver are must to be remain in synchronous.
- Easily to altered the original plain text.

The output feedback can accept cipher text bit errors, but is unable of self-synchronization after trailing cipher text bits, as it disturb the synchronization of the support key stream. A problem of output feedback is that the plaintext can be straightforwardly altered. The plan of an upper bound constraint-based approach to choose the essential subset of intermediate data sets that needs to be encrypted for minimizing privacy -preserving cost. Then specify related basic notations and detailed two useful properties on an SIT. The privacy leakage upper bound control is decomposed layer by layer. A controlled optimization problem with the Privacy leakage control is then changed into a recursive form. A heuristic algorithm is planned for my approach. Extend my approach to an responsive intermediate data set tree which is defined as a Sensitive Intermediate data set Tree and it is a tree structure.. The computing services of this system are placed along with number of labs at UTS. On top of hardware and operating system, virtualization software which virtualizes the infrastructure and provides joined computing and expand the approach to an Sensitive intermediate data set tree.

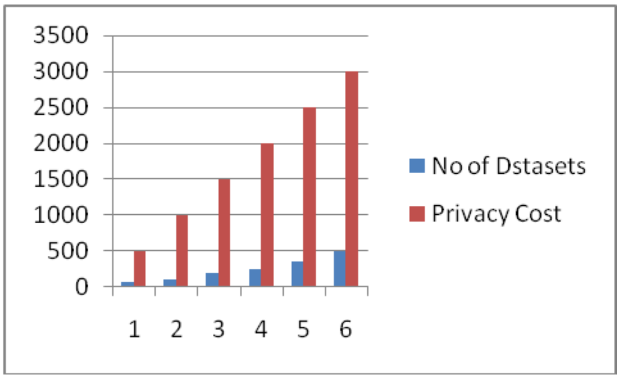

Fig 3 No of datasets Vs Cost

Data source is working to handle intermediate data sets in our research which can be consider as the information of data generation. Reproducibility of data origin can helps to redevelop a data set from its nearby existing predecessor data sets rather than from separate. In this experiments on the intermediate data set which generally used data set in the privacy research society. Intermediate data sets are created from the original data set, and identified by the algorithm proposed. The privacy leakage of each data set is calculated and as formulated. The information of these data sets is described that are available in the online supplemental material. Further, expand trial to intermediate data sets of large amounts. SITs are generated. The values of data size and practice frequencies are randomly created in the period according to the standardized allocation.

In this technique can save privacy-preserving cost considerably over All-Encryption approach. Further, we can see that the difference Csav between Call and Cheu increases when the privacy leakage degree raise, because looser privacy leakage chains involve more data sets can remain unencrypted. The reason about the difference between heuristic and encryption with different privacy leakage degree. The difference changes with various numbers of wide data sets while is convinced. The data owners like the data privacy leakage to be much low and the leakage is bigger when the number of intermediate data sets increased. The more cost can be minimized when the number of data sets becomes larger. This development is the effect of the impressive rise in call and relatively slower increase in heuristic while the number of data sets is getting larger.

VII. CONCLUSIONS AND FUTURE WORK

In cloud computing saving privacy-preserving cost is the major constrained problem. Thus the proposed work decompose the problem into simple small problem and to determine which part of intermediate data sets needs to be encrypted and it save the privacy preserving cost. The heuristic algorithm is designed accordingly to obtain a global privacy preserving solution Finally practical heuristic algorithm is constructed and based on the result of the algorithm which the data sets are to be encrypt is determined. From the execution the mode results on real-world and general data sets illustrate that privacy-preserving cost of intermediate data sets can be much reduced with this approach over existing ones of all data sets are encrypted.

In agreement with various data and computation concentrated applications on cloud, intermediate data set management is becoming an significant research area. Privacy preserving for intermediate data sets is one of important yet demanding research issues, and needs concentrated study. With the contributions of my work, I am planning to further investigate privacy aware efficient scheduling of intermediate data sets in cloud by taking privacy preserving as a metric together with other metrics such as storage and computation. Optimized balanced scheduling strategies are expected to be enveloped toward overall highly efficient privacy aware data set scheduling.

REFERENCES

[1]. C.M. Fung, K. Wang, and P.S. Yu, “Anonymizing Classification Data forPrivacy Preservation,” IEEE Trans. Knowledge and Data Eng., vol. 19, no. 5, pp. 711-725, May 2007.

[2]. C. Gentry, “Fully Homomorphic Encryption Using Ideal Lattices,” Proc. 41st Ann. ACM Symp. Theory of Computing (STOC ’09), pp. 169-178, 2009.

[3]. H. Lin and W. Tzeng, “A Secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding,” IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 6, pp. 995-1003, 2012. N. Cao, C. Wang, M. Li,

[4]. B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, “Privacy-Preserving Data Publishing: A Survey of Recent Developments,” ACM Computing Survey, vol. 42, no. 4, pp. 1-53, 2010.

[5]. H. Takabi, J.B.D. Joshi, and G. Ahn, “Security and Privacy Challenges in Cloud Computing Environments,” IEEE Security & Privacy, vol. 8, no. 6, pp. 24-31, Nov./Dec. 2010.

[6]. 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.H.

[7]. K. Zhang, X. Zhou, Y. Chen, X. Wang and Y. Ruan, “Sedic:Privacy-Aware Data Intensive Computing on Hybrid Clouds,”Proc. 18th ACM Conf. Computer and Communications Security (CCS'11), pp. 515-526, 2011.

[8]. S.B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen and Y. Chen, “On Provenance and Privacy,” Proc. 14th Int'l Conf.Database Theory, pp. 3-10, 2011.

[9]. G. Wang, Z. Zutao, D. Wenliang, and T. Zhouxuan, “Inference Analysis in Privacy-Preserving Data Re-Publishing,” Proc. Eighth IEEE Int’l Conf. Data Mining (ICDM ’08), pp. 1079-1084, 2008.

[10]. K.P.N. Puttaswamy, C. Kruegel, and B.Y. Zhao, “Silverline: Toward Data Confidentiality in Storage-Intensive Cloud Applications,” Proc. Second ACM Symp. Cloud Computing (SoCC ’11), 2011.

[11]. X. Zhang, C. Liu, J. Chen, and W. Dou, “An Upper-Bound Control Approach for Cost-Effective Privacy Protection of Intermediate Data Set Storage in Cloud,” Proc. Ninth IEEE Int’l Conf. Dependable Autonomic and Secure Computing (DASC ’11), pp. 518-525, 2011.

[12]. V. Ciriani, S.D.C.D. Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati, “Combining Fragmentation and Encryption to Protect Privacy in Data Storage,” ACM Trans. Information and System Security, vol. 13, no. 3, pp. 1-33, 2010.

[13]. S.B. Davidson, S. Khanna, S. Roy, J. Stoyanovich, V. Tannen, and Y. Chen, “On Provenance and Privacy,” Proc. 14th Int’l Conf. Database Theory, pp. 3-10, 2011.

[14]. C. Gentry, “Fully Homomorphic Encryption Using Ideal Lattices,” Proc.

[15]. 41st Ann. ACM Symp. Theory of Computing (STOC ’09), pp. 169-178, 2009.

[16]. I. Roy, S.T.V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel, “Airavat: Security and Privacy for Mapreduce,” Proc. Seventh USENIX Conf.

[17]. Networked Systems Design and Implementation (NSDI’10), p. 20, 2010.