RESEARCH ARTICLE OPEN ACCESS

Data Leakage Detection and Security Using Cloud Computing

S.Geetha[1] , M.Nishanthini[2], G.Shanthi[3], K.Sivabharathi[4], M.Suganya[5]
Computer Science and Engineering, Anna University, Chennai, India / Saranathan College of Engineering

ABSTRACT
The data owner will store the data in the cloud. Every user must registered in the cloud. Cloud provider must verify the authorized user. If someone try to access the account, data will get leaked. This leaked data will present in an unauthorized place (e.g., on the internet or someone's laptop). In this paper, we propose Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that collectively approaches the security and performance issues. In DROPS methodology, we have to select the file and then store the particular file in the cloud account. In order to provide security we are going to implement DROPS concepts. Now we divide the file into various fragments based on the threshold value. Each and every fragments are stored in the node using T-Coloring. After the placement of fragments in node, it is necessary to replicate each fragments for one time in cloud.

Keywords—Data leakage, cloud security, fragmentation, replication, performance.

I. INTRODUCTION

Cloud computing is one of the most important emerging and promising field in Information Technology. It provides services to various organization over a internet with the ability to scale up or scale down their service requirements.

There are six key properties of cloud computing:

1. Cloud Computing is Flexible.
2. Cloud computing is Innovation.
3. Cloud computing is Opportunities.
4. Cloud computing is Accessibility.
5. Cloud computing is Savings.
6. Cloud computing is Efficiency.

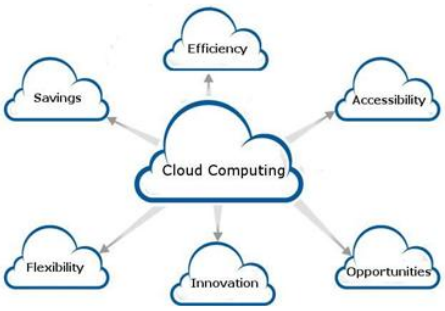

Fig.1 Cloud computing features.

In current situation, the original data will be stored in cloud. Sometimes the unauthorized person try to access the original data, So the data get leaked. For example, a hospital may give patient records to research person who will find new treatments for the disease. Similarly, a company may have partnerships with other companies so that sharing customer data is mandatory. Another enterprise may be outsource its data processing, so that data must be given to the different companies.

We can call the owner of the data as an distributor. Our goal is to detect when the owner's original data have been leaked by intruder, and if possible to identify the intruder that leaked the data. However, in some cases, it is important not to alter the original data. For example, if an outsourcer is doing our payroll, he must know the exact salary and customer bank account numbers. If medical research person will be analyzing patients records, they may need accurate data for the patients.

Traditionally, data leakage detection is handled by watermarking technique. For example, a unique code is embedded in each distributed copy using various types of watermarking algorithm. If that copy is later discovered in the hands of an unauthorized person, the leaker cannot be identified. This is the major drawback in watermarking technique. Watermarks can be very useful in real-time environment, but again, it involves some modification of the original data. In sometimes watermarks can be destroyed if the data recipient is malicious.

Security is one of the most crucial aspects in cloud computing. Hence this prohibit the adoption of cloud computing. Therefore, in this paper, we collectively approaches the issue of security and performance. We present Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) this will fragments the owners files into various parts and replicate them in the cloud space. The division of a file into fragments is performed based on the threshold value decided by the data owner. Such that the individual fragments doesn't hold any meaningful information. Each and every cloud

nodes we use the term node to represent computing and storage contains a distinct fragment to increase the data security.

A successful attack on each and every single node must not reveal the locationsof other fragments within the cloud space. The intruder cannot predict the locations of file fragments so that it improves the security in the cloud. Now we select the nodes in such a manner that they are not adjacent and are at certain distance from each other. The node separation is done by the means of the T-coloring concept. Moreover, the nodes storing the fragments are placed with a certain distance by means of graph T-Coloring inorder to restrict an intruder of guessing the locations of the fragments.

II. EXISTING SYSTEM

A Watermark is a signal that is securely, indiscernibly embedded into original content such as an image, video, text, or audio signal, producing a watermarked signal and it describes information that can be used for verification of original copy. It provides an effective watermarking technique along with the relational data. This technique ensures that some bit positions of some of the attributes of some of the tuples contain required specific values.

The tuples, bit positions in an attribute, and specific bit values are all algorithmically determined under the control of a private key that key is only known to the data owner. This bit pattern constitutes the watermark code. If only one person access to the private key then it is possible to detect the watermark with higher probability.

The watermark can be detected even in a small subset of a relational data as long as the sample contains some of the watermarks. Protection is based upon the insertion of digital watermarks into the original data. The watermarking technique introduces small errors into the data being watermarked. These intentional errors are called marks and all these marks collectively constitute the watermark. The marks must not have a impact on the usefulness of the data and that data should be placed in such a way that a malicious user cannot destroy them.

III. PROPOSED SYSTEM

3.1 Data Leakage :

The data owner first registered into the cloud account. Each and every user has to registered into the cloud. Now the data owner and user will become the authorized person. The data owner will upload the file into the cloud. Now the data owner login into the account, at that time the cloud provider verify the already registered owner or not. If they are registered owner and then they will transfer the file to the registered user. Now the registered user login into their account the cloud provider again verify the registered user. The registered user will download the file sent by the data owner.

If someone try to copy the URL, the data get leaked in someone's laptop. Now the details about the unauthorized person will be tracked. This tracked information sent as a mobile intimation to the data owner. The mobile intimation will hold informations like IP address, MAC address and GPS location.

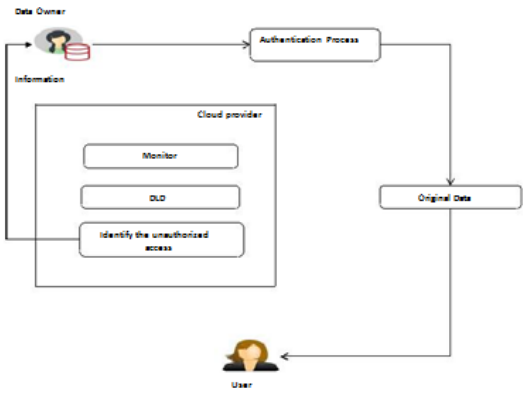

Fig.2 Data leakage Architecture.

3.2 Drops Methodology:

In DROPS methodology, we are not storing the entire file in cloud space. Now we are splitting the entire file into various fragments. These fragments have to distribute in the cloud space. Each and every fragment has to place in a particular node. So that each node contain only a single fragments. In each successful attack the node will not reveal the significant information. After the fragmentation process replication will takes place. In replication process, each fragment has to replicate its content once in the cloud space. In this way we can achieve the security in the cloud computing. In DROPS methodology, user sends the data file to cloud space. Upon receiving the file the cloud manager performs:

(1) File Fragmentation
(2) Nodes selection
(3) Stores fragment
(4) Nodes selection for fragments replication.

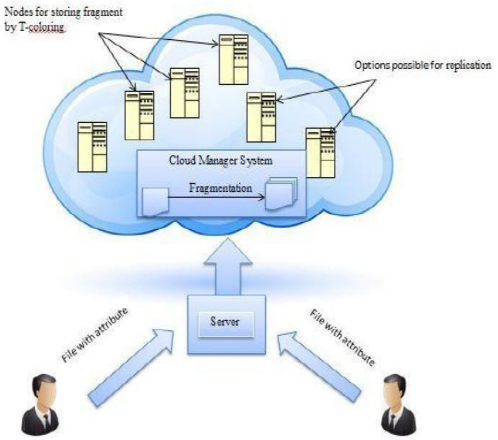

Fig.3 DROPS Methodology.

3.2.1 Drops Implementation:

3.2.1.1 File Fragmentation

In cloud, security is the major aspects for a large-scale system. This provides security of the system as whole and individual nodes. On every successful intrusion into a single node it provides more consequences for data and other nodes. A successful intrusion may lead to software failure. It may also lead to administration defenseless. File fragmentation is a term that describes a group of files that are scattered throughout the cloud. The data owner splits the file into various pieces called fragments. The size of fragmentation is decided by data owner using threshold value. In various aspects the threshold value can be fixed, they are,

1. Decide percentage based on file size.
2. Numbering of each fragments.

For example, first fragment the file by 10% of total size or split the file by using various size like fragment 404MB, fragment 1000MB. It is also possible to number the fragments like fragment1, fragment2 etc., The numbering of each fragments only known to the data owner. There are three types of fragmentation:

1. Horizontal.
2. Vertical.
3. Mixed (Hybrid).

3.2.1.2 Fragment Placement

In order to provide the security while placing the fragments in cloud, the concept of T-coloring is required. T-coloring is mainly used for the channel assignment problem. This will generates a non-negative (i.e. positive number) random number and builds the set T starting from zero to generated random number. It assigns colors to the each and every node, such that, initially, all of the nodes will be in open color. When a fragment is placed on the particular node, all of the neighborhood nodes are at a certain distance belonging to T and there are assigned to close color. In this process, this will lose some of the central nodes in cloud so that may increase retrieval time. If anyhow the intruders try to track the node position and take the fragment, he cannot determine the location of the other fragments. The intruder can only keep on guessing the location of the other fragments in cloud. Because the nodes are separated by using T-coloring concepts.

3.2.1.3 Fragment Replication:

In replication process, unique copy exists and the same copy will exists once again. The replication process is used to increase the data availability and their by improve the retrieval time. This performs a controlled replication. In replication process, copies of the same data item have the same value. There are three types of replication,

1. No replication.
2. Fully replicated.
3. Partial replication.

It places the fragment on the node that provides the limited access cost. Hence improve retrieval time for accessing the fragments and reconstruction of the original file. While replicating the fragment, the separation of fragments is achieved by T-coloring, is also taken care of node placement. In case of a large number of fragments or small number of nodes, it is also possible that some of the fragments can be left without being replicated because of using T-coloring concepts.

3.2.1.4 T-Coloring

As discussed previously, T-coloring prohibits storing the fragment and also avoid the neighborhood of a node storing a fragment, resulting in the eliminating the nodes used for storage. In such a case, only for the remaining fragments, the nodes that are not holding any of the fragments are selected for storage randomly. The nodes were separated by means of T-coloring concepts. The fragmentation process ensured that no successful information was obtained in successful attack. No node in the cloud, stored more than a single fragment of the same file. The DROPS methodology performance is compared with full-scale replication techniques. This results in simultaneous focus on the security and performance. Resulted in increased security level of data.

IV. CONCLUSION

We have shown that it is possible to access the likelihood of an unauthorized person is responsible for a leak, based on the overlap of his data with the leaked data. Our model is relatively simple, but we believe that it provides the essential trade-offs. Hence we have presented a variety of data distribution strategies that can improve the distributor's chances of identifying a leaker. We proposed the DROPS methodology, a cloud storage security scheme that collectively deals with the security and performance that increase the retrieval time. The data file was fragmented and the fragments are dispersed over multiple nodes. The fragmented file will be replicated their by increasing data availability and provide security to the cloud.

REFERENCES

[1]. B. Mungamuru and H. Garcia-Molina, "Privacy, Preservation and Performance: The 3 P's of Distributed Data Management,"technical report, Stanford Univ., 2008.
[2]. V.N. Murty, "Counting the Integer Solutions of a Linear Equation with Unit Coefficients," Math. Magazine, vol. 54, no. 2, pp. 79-81, 1981.
[3]. S.U. Nabar, B. Marthi, K. Kenthapadi, N. Mishra, and R. Motwani, "Towards Robustness in Query Auditing," Proc. 32nd Int'l Conf. Very Large Data Bases (VLDB '06), VLDB Endowment, pp. 151-162,2006.
[4]. P. Papadimitriou and H. Garcia-Molina, "Data Leakage Detection," technical report, Stanford Univ., 2008.
[5]. P.M. Pardalos and S.A. Vavasis, "Quadratic Programming with One Negative Eigenvalue Is NP-
[6]. Hard," J. Global Optimization, vol. 1, no. 1, pp. 15-22, 1991.
[7]. J.J.K.O. Ruanaidh, W.J. Dowling, and F.M. Boland, "Watermarking Digital Images for Copyright
[8]. Protection," IEE Proc. Vision, Signal and Image Processing, vol. 143, no. 4, pp. 250-256, 1996.
[9]. R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," Proc. ACM SIGMOD, pp. 98-109, 2003.
[10]. L. Sweeney, "Achieving K-Anonymity Privacy Protection Using Generalization and Suppression," http://en.scientificcommons.org/ 43196131, 2002.
[11]. K. Bilal, S. U. Khan, L. Zhang, H. Li, K. Hayat, S. A. Madani, N. Min-Allah, L. Wang, D. Chen, M.
[12]. Iqbal, C. Z. Xu, and A. Y. Zomaya, "Quantitative comparisons of the state of the art data center architectures," Concurrency and Computation: Practice and Experience, Vol. 25, No. 12, 2013, pp. 1771-1783.
[13]. K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Zomaya, "On the characterization of the structural robustness of data center networks," IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013, pp. 64-77.