www.ioirp.com
International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 12. November 2015

HIGH SECURITY FOR DATA IN CLOUD THROUGH FRAGMENTATION

Nivetha F

Department of Computer Science and Engineering
K. Ramakrishnan College of Technology, Samayapuram
Tiruchirappali, India
kannan.nive5@gmail.com

***Abstract* – Now-a-days, the term cloud computing has the feel of a buzzword any more than the term the web is. The main goal of cloud computing is to allow users to benefit from all technologies without the need for deep knowledge about that. It aims to cut costs by "pay-as-you-go" model. In cloud computing, usually we transfer the data or submit the data to a third party administrator. It gives rise to security concerns. Due to attacks by other users and nodes within cloud, data compromise may occur. Even though, there are many security measures applied to protect data, still we are facing some security issues. Therefore, we propose Division and Replication of Data in the cloud for Optimal Performance and Security (DROPS) that provide solution to the existing issues. In this methodology, the files in the cloud storage are encrypted and then divided into number of fragments and replicate the fragmented data over the cloud nodes. Each of the nodes in cloud stores only a single fragment of a particular data file fragment. No meaningful information is revealed to the attacker even in case of successful attack. To enhance the security, cryptographic technique is used for encryption of data. Thus it results with higher level of security with slight performance overhead.**

***Keywords* -- Cloud data, data fragmentation, node allocation, security, centrality**

I. INTRODUCTION

The cloud computing is a flexible, cost-effective and proven delivery platform for providing business or consumer IT services over the Internet. It combines a number of computing concepts and technologies such as Service oriented Architecture (SOA) [8]. It contains potential privacy risks which lead to privacy and security concerns because data is travelling over the internet and is stored in remote locations. Security is one of the most crucial aspects among those prohibiting the wide-spread adoption of cloud computing because essential services are often outsourced to a third-party, which makes it harder to maintain data security and privacy, support data and service availability [10]. The fundamental elements of the cloud require security which depends and varies with respect to the deployment models that is used, the way by which it is delivered and the character it exhibits [17].

The offsite data storage cloud utility requires users to move data in cloud's virtualized and shared environment that may result in various security concerns. Currently the security has lot of loose ends which scares away a lot of potential users. The security modules should cater to the entire issues element in the cloud should be analyzed at the macro and micro level and an integrated solution must be designed and deployed in the cloud to attract and enthrall the potential customers. Though, there are extreme advantages in using a cloud-based system, there are yet many practical problems which have to be solved [5].

The cloud manager will manage all the files that are all stored by multiple data owners. Once the file is stored in cloud, cloud manager will encrypt the file and then start fragmentation with the help of fragment engine. The fragmented files will be stored in cloud nodes using allocation techniques. The main advantage is that the data security gets increased because of storing each fragment in distinct location. Even a successful attack on a single node must not reveal the locations of other fragments within the cloud [10]. The amount of data compromise can also be reduced by making fragments of data file and storing them on separate nodes. The focus of this concept is on the security of the data in the cloud and we doo no take into account the security of the authentication system.

II. T-COLORING ALGORITHM

In the existing cloud system, the data are outsourced to a third-party administrative control, gives rise to security concerns. Due to attacks by other users and nodes within the cloud, the data compromise may occur [10]. In

traditional cloud storage, the data owner sends copies of files over internet to the data serves, which the records the information to an off-site storage system that is maintained by a third-party. Whenever the data owner wants to retrieve the information, they access the data server through web-based interfaces. The concerns that are facing in existing system are reliability and security. To secure data, most systems use a combination of techniques, including: (i) Encryption is a complex algorithm is used to encode information, (ii) Authentication, which requires creating a user name and password, (iii) Authorization, the data owner lists the people who are authorized to access information stored on the cloud storage system. Even with these protective measures in cloud system, there's always the possibility that a hacker will find an electronic back door and access data. In order to overcome the problem with the existing system, the method called DROPS is used to the cloud storage system that contains node allocation algorithm that is designed to secure the data.

DROPS, a methodology that divides the owner's file into number of fragments and stores them in different locations, so that the attacker cannot tract the stored data. Even, the data was tracked means no meaningful information can be found. This methodology uses T-coloring algorithm, which aims at achieving high security for the data in the cloud storage. This will provide absolute value of the difference between two colors of adjacent vertices must not belong to fixed set of values.

Based on this algorithm, DROPS methodology can handle the attacks in which attacker gets hold of user data by avoiding or disrupting security defenses. The attacks that are handle by this methodology are data recovery, cross VM attack, improper media sanitization, E-discovery, VM escape, VM rollback. It is noteworthy that in case of successful attacks, no useful information can be revealed to the attacker because the attacker cannot trace the location of the fragment or the attacker cannot do anything with the information what is found [3]. The attacker can only keep on guessing the location of other fragments.

III. T-COLORING ALGORITHM WORKING

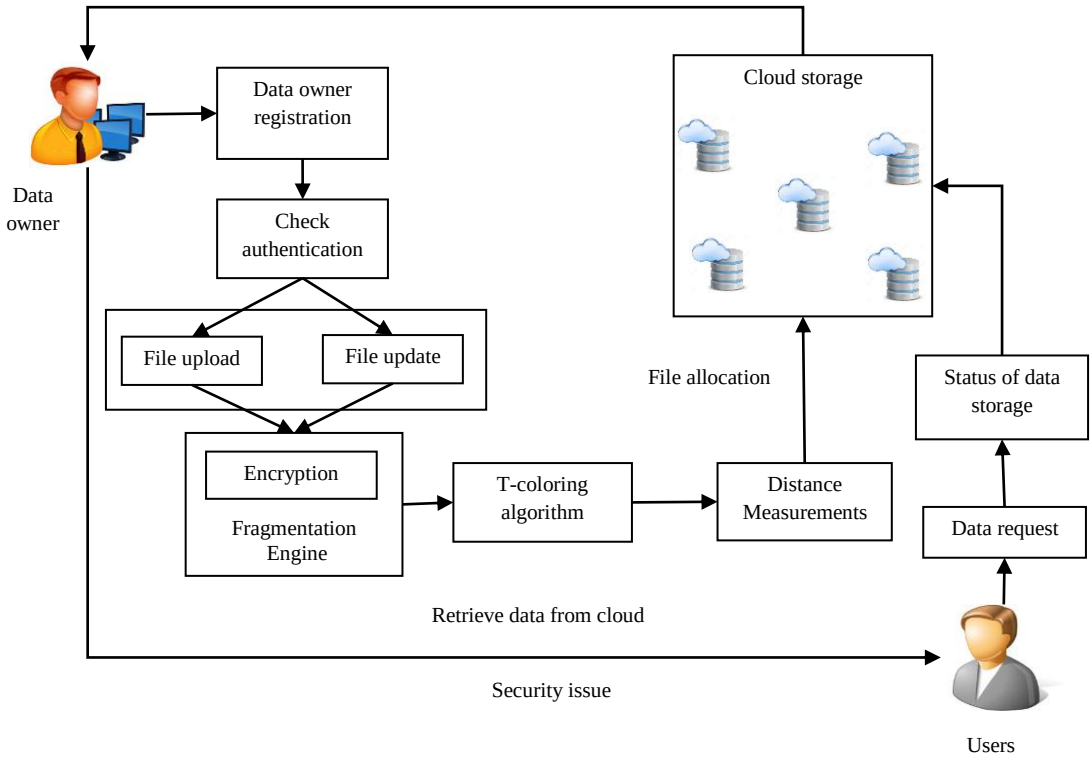

Figure 1.1 System Architecture

Figure 1.1 describes the system architecture. The client node is considered data owner which contains number of files. The data owner have to create an account in cloud system, so that the user can store their files in the cloud when there is internet connection exist.

The cloud manager will manage all the files that are all stored by multiple data owners. Once the file is stored in cloud, the file will get encrypted. Then, cloud manager will start fragmentation with the help of fragmentation engine. Based on the fragmentation threshold value, the file will get fragmented into number of pieces. Then it will be stored in cloud nodes using allocation techniques. After fragmentation, the primary node will be determined and it gets stored initially. Then, all the remaining k^{th} fragments will be placed in remaining available nodes. The replication of the fragment will be maintained in same node. During the retrieval process, cloud manager will collect all the files and return to the requested user based on the primary node that is maintained.

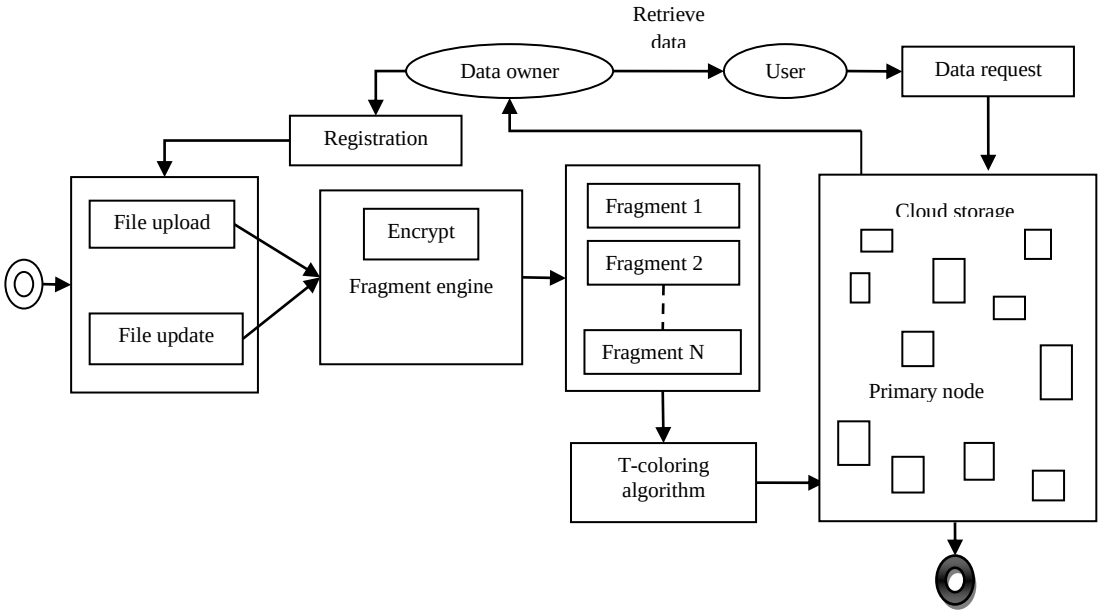

Figure 1.2 Workflow execution

Figure 1.2 defines the overall workflow of T-coloring algorithm. The following steps will demonstrate the working of T-coloring algorithm:

(i) Once the data owner registered in the cloud storage, authentication has to be checked. The process has to be identified that the file has to upload or have to update.
(ii) Then, fragmentation engine is used to fragment the files into number of pieces. Using the centrality measures the primary node has to be placed in the storage area.
(iii) Further using the T-coloring algorithm, the available nodes have to be collected as such the nodes that are not adjacent to the primary node are considered as available nodes.
(iv) The other nodes are considered as unavailable nodes, the remaining fragments should be placed in the available nodes based on the routine of T-coloring algorithm.
(v) Once all the fragments are placed in the cloud storage means its status have to be maintained by the cloud manager and the data owner also aware of the security status of the storage.
(vi) If any user request for data to the cloud storage means the data owner will get intimation about the request.

(vii) The data owner will provide a key to the user, so that the user can retrieve the file from cloud.

IV. COMPONENTS

The various components used here are: (A) Fragmentation Engine, (B) T-coloring algorithm, (C) Nearest Neighbor Node Identification, (D) Cloud management system. The working of each of the components is elaborated in the following section:

A. Fragmentation engine

The fragmentation engine is an initial component which is located in the cloud storage. This component is mainly used for fragmenting the user files. Initially, data owner will upload a file to cloud storage system. The cloud manager will collect a file and encrypt the file using cryptographic techniques. Convert that encrypted file into 'n' number of fragments using fragments engine. All the fragments can upload to distinct node in different region. Once the file is fragmented, the primary node among the fragments should be determined. Based on the primary node, the retrieval process can be done. The cloud manager should maintain the primary node in secured manner, so that no intruder can affect the file without knowing the primary node. The amount of compromised data can be reduced by making fragments of data file and storing them on separate nodes.

A successful intrusion on a single or few nodes will only provide access to a portion of data that might not be of any significance. If an attacker is uncertain about the locations of the fragments, the probability of finding fragments on all of the nodes is very low. In cloud systems with thousands of nodes, the probability for an attacker to obtain a considerable amount of data reduces significantly. To improve the data retrieval time, fragments can be replicated in a manner that reduces retrieval time to an extent that does not increase the aforesaid probability.

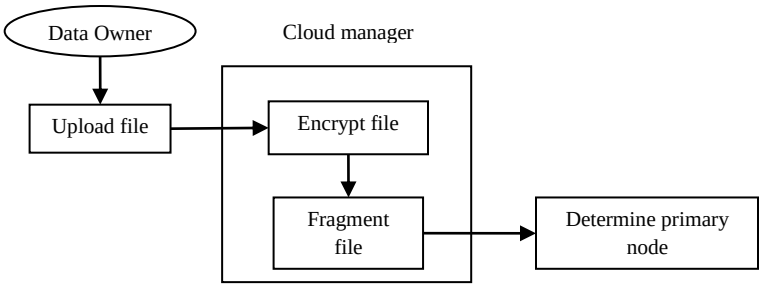

Figure 1.3 Fragmentation Engine

B. T-coloring algorithm

The node identifier component uses the T-coloring concept for determining the available and unavailable nodes for allocation of fragments. Once the file is fragmented into number of pieces, the primary node is determined. Then, a non-negative random number will get generate and build the set T starting from zero to the generated random number. After that, the primary fragment should be placed in the cloud storage, based on centrality measures. Among the cloud nodes that are all available in the storage area, the central node should be identified, based on the betweenness centrality measures. The set T is used to restrict the node selection to the nodes within the neighborhood at a distance belonging to T are assigned close_color and remaining nodes are assigned open_color.

By these notifications, the close_color nodes are considered as unavailable nodes and the open_color nodes are considered as available nodes. Initially, all of the nodes are given the open_color. Once a fragment is placed on the node, all of the nodes within the neighborhood at a distance belonging to T are assigned close_color. In the aforesaid process, we lose some of the central nodes that may increase the retrieval time but a higher security level can be achieved. If somehow the intruder compromises a node and obtains a fragment, then location of the other fragments cannot be determined.

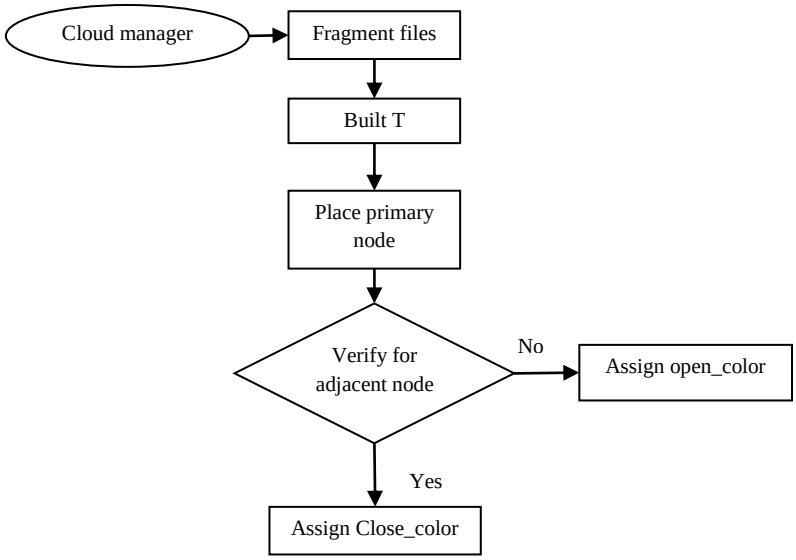

Figure 1.4 T-coloring algorithm

C. Nearest-Neighborhood Node Identification

After determining the available nodes and unavailable nodes, the 'n' number of file fragments should be placed in a distinct node. Each and every fragment should have a size. Every single storage node calculate read and write a fragment, same time primary node stores primary copy of fragment. Here, every storage area has two field records, first field is to store primary node for primary fragment using centrality measures, and second field is to identify the nearest neighbor node for storing k^{th} fragment data using T-coloring and centrality measures.

The most central node to the cloud network should be choosing to provide better access time. The concept of centrality is used to reduce access time. Based on the available nodes that are determined and using the centrality values, remaining nodes can be placed. Once the nearest nodes are identified, the nodes should be verified that whether it is open_color node or close_color node. If it is open_color nodes means the k^{th} fragment can be placed, otherwise the nearest node to the close_color node should be identified and aforesaid process have to be followed, until all the fragments get placed in storage area.

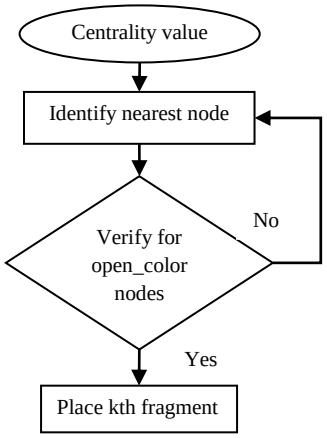

Figure 1.5 Nearest-Neighborhood Node Identification

D. Cloud management system

Once all the fragments are placed on its appropriate locations, the cloud manager should maintain all the nodes in the cloud. Cloud storage has a different unique storage in different region. This all node should be followed by a single primary node that represent first placement of fragment. Then, T-coloring algorithm is used to plan the remaining nodes and also it uses centrality measures. T-coloring prohibits storing the fragment in neighborhood of a node storing a fragment, resulting in the elimination of a number of nodes to be used for storage.

In such a case, only for the remaining fragments, the nodes that are not holding any fragment are selected for storage randomly. Based on replication algorithm, a controlled replication is performed to increase the data availability, reliability and improve data retrieval time. Cloud node store a fragment file in single time and replicate the file for single time because to reduce the storage cost for end user.

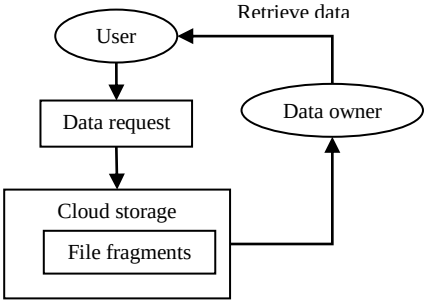

Figure 1.6 Cloud management system

V. CONCLUSION

The cloud storage scheme collectively deals with the security and performance in terms of retrieval time. The user file was fragmented and the fragments are dispersed over multiple nodes. Once the fragment is placed in primary node, remaining nodes are placed over multiple nodes. The cloud manager will store and maintain that primary node for retrieval process. The fragmentation and dispersal ensured that no significant information was obtainable by an adversary in case of a successful attack. Each node in cloud should contain only one fragment. The nodes were separated by means of T-coloring. The performance of this system is in increasing manner to provide less access time. Mainly, this system will resulted in increased security level of data in cloud. Currently, with this security system, the user has to download the file, update the contents and again upload to the cloud. To overcome the aforesaid issues, as a future work, an automatic update mechanism that can identify and update the required fragment only. By this, the time and resource utilization will be saved.

References

[1] Hale W.K. (1995), 'Frequency assignment: Theory and applications', *Proceedings of the IEEE,* Vol. 68, No. 12, pp. 1497-1514.

[2] Hashizume K., Rosado D.G., Fernndez-Medina E., and Fernandez E.B. (2013), 'An analysis of security issues for cloud computing', *Journal of Internet Services and Applications*, Vol. 4, No. 1, pp. 1-13.

[3] Juels A. and Opera A. (2013), 'New approaches to security and availability for cloud data', *Communications of the ACM*, Vol.56, No. 2, pp. 64-73.

[4] Kamara S. and Lauter K. (2010), 'Cryptographic cloud storage', *In Workshops on Real-Life Cryptographic Protocol and Standardization.*

[5] Khan S.U., and Ahmad I. (2008), 'Comparison and analysis of ten static heuristics-based Internet data replication techniques', *Journal of Parallel and Distributed Computing*, Vol. 68, No. 2, pp. 113-136.

[6] Mazhar Ali, Kashif Bilal, Samee U.Khan, Bharadwaj Veeravalli, Keqin Li, and Albert Y. Zomaya (2015), 'DROPS: Division and Replication of Data in Cloud for Optimal Performance and Security', *IEEE Transactions on Cloud Computing*, No. 99, pp. 1.

[7] Subashini S., Kavitha V. (2011), 'A survey on Security issues in service delivery models of Cloud Computing', *J Netw Comput Appl* 34(1):1–11.