A

Seminar Report

Cloud DROPS in Cloud Computing

Submitted in partial fulfillment of the requirements for the Award of the Degree

of

Master of Computer Applications

of

APJ Abdul Kalam Technological University

Submitted by

ABHILASH THANKACHAN

RegNo: TVE16MCA1

Department of Computer Applications

COLLEGE OF ENGINEERING TRIVANDRUM

AUGUST 2018

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING TRIVANDRUM

CERTIFICATE

Certified that this Seminar report entitled, ***"DROPS in Cloud Computing "*** *is the paper presented by* ***" Abhilash Thankachan "(Reg No: TVE16MCA1)*** *in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications of APJ Abdul Kalam Technological University during the year 2018.*

Prof. Baby Syla L.

Co-ordinator

Prof. Jose T Joseph.

Head of the Department

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this seminar. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my seminar.

I would like to thank **Dr.Hari V.S**, Principal, College of Engineering Trivandrum, who helped me during the entire process of work.

I am extremely grateful to **Prof.Jose T Joseph**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I would like to thank my coordinator, **Prof. Baby Syla**, Dept of Computer Applications, who motivated me throughout the work of my seminar.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this seminar. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Abhilash Thankachan

Abstract

Cloud Computing (CC) is an emerging trend that offers number of important advantages. One of the fundamental advantages of CC is pay-as-per-use, where customers will pay only according to their usage of the services. Outsourcing data to a third-party administrative control, it is done in cloud computing, gives rise to security concerns. The data compromise may occur due to attacks by other users and nodes within the cloud. Therefore, high security measures are required to protect data within the cloud. However, the employed security strategy must also take into account the optimization of the data retrieval time. The DROPS (Division and Replication of Data in the Cloud for Optimal Performance and Security) technique collectively approaches the security and performance issues. In the DROPS technique, divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a single fragment of a particular data file that ensures that even in case of a successful attack, no meaningful information is revealed to the attacker. Moreover, the nodes storing the fragments are separated with certain distance by means of graph T-coloring to prohibit an attacker of guessing the locations of the fragments. The DROPS technique does not depend on the traditional cryptographic techniques for the data security; thereby relieving the system of computationally expensive methodologies. So the probability to locate and compromise all of the nodes storing the fragments of a single file is extremely low.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The utilized security procedure should likewise consider the improvement of the information recovery time. The DROPS (Division and Replication of Data in the Cloud for Optimal Performance and Security) technique that aggregately approaches the security and performance issues. In the DROPS technique, separate a document into sections, and duplicate the divided information over the cloud nodes. Security is one of the most crucial aspects among those prohibiting the wide- spread adoption of cloud computing. Cloud se- curity issues may stem due to the core technologys implementation (virtual machine (VM) escape, session riding, etc.), cloud service offerings (structured query language injection, weak authentication schemes, etc.), and arising from cloud characteristics (data recovery vulnerability, Internet protocol vulnerability, etc.). For a cloud to be secure, all of the participating entities must be secure. In any given system with multiple units, the highest level of the systems security is equal to the security level of the weakest entity.

Therefore, in a cloud, the security benefit does not solely depend on an individuals security measures. The neighbouring entities may provide an opportunity to an attacker to bypass the users defences. The off-site data storage cloud utility requires users to move data in clouds virtualized and shared environment that may result in various security concerns. Pooling and elasticity of a cloud, allows the physical resources to be shared among many users. Furthermore, the shared resources may be reassigned to other users at some instance of time that may result in data compromise through data recovery methodologies. Further- more, a multi-tenant virtualized environment may result in a VM to escape the bounds of virtual machine monitor (VMM). The escaped VM can interfere with other VMs may access to unauthorized data. Similarly, crosstenant virtualized network access may also compromise data privacy and integrity. Due to improper media sanitization can also leak customers private data.

Chapter 2

Related Work

Juels et the author of the paper " New approaches to security and availability for cloud data,Communications of the ACM ", Vol. 56, No. 2, 2013, pp. 64-73 presented a technique to ensure the integrity, freshness, and availability of data in a cloud. The data migration to the cloud is performed by the Iris file system. A gateway application is designed and employed in the organization that ensures the integrity and freshness of the data using a Merkle tree. The file blocks, MAC codes, and version numbers are stored at various levels of the tree. The proposed technique in the paper heavily depends on the user s em- ployed scheme for data confidentiality. Moreover, the probable amount of loss in case of data tempering as a result of intrusion or access by other VMs cannot be decreased. Our proposed strategy does not depend on the traditional cryptographic techniques for data security. Moreover, the DROPS methodology does not store the whole file on a single node to avoid compromise of all of the data in case of successful attack on the node.

The authors G. Kappes, A. Hatzieleftheriou, and S. V. Anastasiadis, Dike of the paper "Virtualization-aware Access Control for Multitenant Filesystems" University of Ioannina, Greece, approached the virtualized and multi-tenancy related issues in the cloud storage by utilizing the consolidated storage and native access control. The Dike authorization architecture is pro- posed that combines the native access control and the tenant name space isolation. The proposed system is designed and works for object based file systems. However, the leakage of critical information in case of improper sanitization and malicious VM is not han- dled. The DROPS methodology handles the leakage of critical information by fragmenting data file and using multiple nodes to store a single file.

Chapter 3

Existing System Approach

In existing system data reliability, data availability, and response time are dealt with data replication strategies. However, storing replicas data over a number of nodes increases the attack surface for that particular data.For example, storing m replicas of a file in a cloud instead of one replica increases the probability of a node holding file to be chosen as attack sufferer, from 1/n to m/n where n is the total number of nodes. Existing system was not achieving proper security.

3.1 Disadvantages

- A key factor determining the throughput of a cloud that stores data is the data retrieval time.
- In large-scale systems, the problems of data reliability, data availability, and response time are dealt with data replication strategies.
- However, placing replicas data over a number of nodes increases the attack surface for that particular data.
- Affected on security and performance.

3.2 Data Fragmentation

The security of a large-scale system, such as cloud depends on the security of the system as a whole and the security of individual nodes. A successful intrusion into a single node may have severe consequences, not only for data and applications on the victim node, but also for the other nodes. The data on the victim node may be revealed fully because of the presence of the whole le. A successful intrusion may be a result of some software or administrative vulnerability. In case of homogenous systems, the same aw can be utilized to target other nodes within the system. The success of an attack on the subsequent nodes will require less effort as compared to the effort on the rst node. Comparatively, more effort is required for heterogeneous systems. However, compromising a single le will require the effort to penetrate only a single node. The amount of compromised data can be reduced by making fragments of a data le 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 signicance. Moreover, if an attacker is uncertain about the locations of the fragments, the probability of nding fragments on all of the nodes is very low. Consider a cloud with M nodes and a le with z number of fragments. Let s be the number of successful intrusions on distinct nodes, such that s ¿ z

The probability that s number of victim nodes contain all of the z sites storing the le fragments (represented by $P(s,z)$)is given as:

$$P(s,z) = \frac{\begin{pmatrix} s \\ z \end{pmatrix} \begin{pmatrix} M-s \\ s-z \end{pmatrix}}{\begin{pmatrix} M \\ s \end{pmatrix}}$$

3.3 Centrality

The centrality of a node in a graph provides the measure of the relative importance of a node in the network. The node is important in the network if it:
(a) interconnects more nodes than others,
(b) can be reached easily by other nodes, or
(c) can reach other nodes easily.

The objective of improved retrieval time in replication makes the centrality measures more important. There are various centrality measures; for instance, closeness centrality, degree centrality, betweenness centrality, eccentricity centrality, and eigen- vector centrality. Here only elaborate on the closeness, betweenness, and eccentricity centralities because here using the aforesaid three centralities in this work.

3.3.1 Betweenness Centrality

The betweenness centrality of a node n is the number of the shortest paths, between other nodes, passing through n. Formally, the betweenness centrality of any node v in a network is given as:

$$C_b(v) = \sum_{a \neq v \neq b} \frac{\delta_{ab}(v)}{\delta_{ab}}$$

3.3.2 Closeness Centrality

A node is said to be closer with respect to all of the other nodes within a network, if the sum of the distances from all of the other nodes is lower than the sum of the distances of other candidate nodes from all of the other nodes. The lower the sum of distances from the other nodes, the more central is the node. Formally, the closeness centrality of a node v in a network is dened as:

$$C_c(v) = \frac{N-1}{\sum_{a \neq v} d(v,a)}$$

where N is total number of nodes in a network and d(v,a)represents the distance between node v and node a.

3.3.3 Eccentricity

The eccentricity of a node n is the maximum distance to any node from a node n. A node is more central in the network, if it is less eccentric. Formally, the eccentricity can be given as:

$$E(v_a) = max_b \; d(v_a, v_b)$$

3.4 T-Coloring

Consider a graph G=(V,E)and a set T containing non-negative integers including 0. The T-coloring is a mapping function f from the vertices of V to the set of non-negative integers, such that f(x)- f(y) is not an element of T, where(x,y) is an element of E. The mapping function f assigns a color to a vertex. In simple words, the distance between the colors of the adjacent vertices must not belong to T. Formulated by Hale, the T-coloring problem for channel assignment assigns channels to the nodes, such that the channels are separated by a distance to avoid interference.

Chapter 4

Proposed System

The proposed system called DROPS(Division and Replication of Data in Cloud for Optimal Performance and Security) jointly approaches the security and performance issues. The proposed DROPS scheme ensures that even in the case of a successful attack, no meaningful information is disclosed to the attacker. The DROPS technique doses not depend on traditional cryptographic techniques for data security. The non-cryptographic nature of the proposed scheme makes it faster to perform the required operations (placement and retrieval) on the data. It make sure a controlled replication of the file fragments, where each of the fragments is replicated only once for the purpose of improved security. A cloud storage security scheme jointly deals with the security and performance in terms of retrieval time.

4.1 Advantages

- Improve security.
- Improve performance.
- No information is revealed to the attacker. (If an attacker is uncertain about the locations of the fragments, the probability of finding fragments on all of the nodes is very low.)
- No load on single node of cloud.
- Numbers of fragments are decided according to owners choice.

Chapter 5

DROPS System Model

A cloud that consists of M nodes, each with its own storage capacity. Let Si represents the name of ith node and si denotes total storage capacity of Si Communication time between Si and Sj is the total time of all of the links within a selected path from Si to Sj represented by c(i, j). Consider N number of file fragments such that Ok denotes k -th fragment of a file while ok represents the size of k-th fragment. Pk denote the primary node that stores the primary copy of Ok, replication scheme for Ok denoted by Rk is also stored at Pk and Whenever there is an update in not as an independent document. Please do not revise any of the current designations Ok , the updated version is sent to Pk that broadcasts the updated version to all of the nodes in Rk. Let colSi store the value of assigned color to Si. The colSi can have one out of two values, namely: open color and close color. The value open color represents that the node is available for storing the file fragment. The value close color shows that the node cannot store the file fragment The set T is used to restrict the node selection to those nodes that are at hop-distances not belonging to T.

In the DROPS methodology, not to store the entire file at a single node. The DROPS methodology fragments the file and makes use of the cloud for replication. The fragments are distributed such that no node in a cloud holds more than single fragment, so that even a successful attack on the node leaks no significant information.

Symbols	Meanings
M	Total number of nodes in the cloud
N	Total number of fragments to be placed
O_k	k-th fragment of the file
o_k	Size of O_k
S^i	i-th node
s_i	Size of S^i
cen_i	Centrality measure of S^i
col_{S^i}	Color assigned to S^i
r_k^i	Number of reads for O_k from S^i
R_k^i	Aggregate read cost of r_k^i
w_k^i	Number of writes for O_k from S^i
W_k^i	Aggregate write cost of w_k^i
NN_k^i	Nearest neighbour of S^i holdingO_k
c(i,j)	Communication cost between S^i and S^j
P_k	Primary node for O_k
R_k	Replication schema for O_k
RT	Replication Time

Table 5.1: Notations and their meanings.

In the DROPS methodology, user sends the data file to cloud. The cloud manager system(a user facing server in the cloud that entertains users requests) upon receiving the file performs:

1. fragmentation
2. first cycle of nodes selection and stores one fragment over each of the selected node and
3. second cycle of nodes selection for fragments replication. The cloud manager keeps record of the fragment placement and is assumed to be a secure entity.

Chapter 6

DROPS Architecture

This system mainly consist of 3 modules:

6.1 Cloud Client

Cloud client should be Data owner or Data user.

6.1.1 Data Owner

Data owner is responsible for uploading file on cloud as well as view files uploaded by him or others. Data owner has information about the placed fragment and its replicas with their node numbers in cloud.

6.1.2 Data User

Data user is the one who is responsible for downloading files or view files uploaded by others. To download file from cloud he has to be authenticated user otherwise he will be considered as attacker.

6.2 Cloud Server

6.2.1 Fragmentation

This approach is used for fragmenting the file for security purpose at sever side. This approach runs the Fragmentation algorithm. It has file as input and produces the file fragments as output.

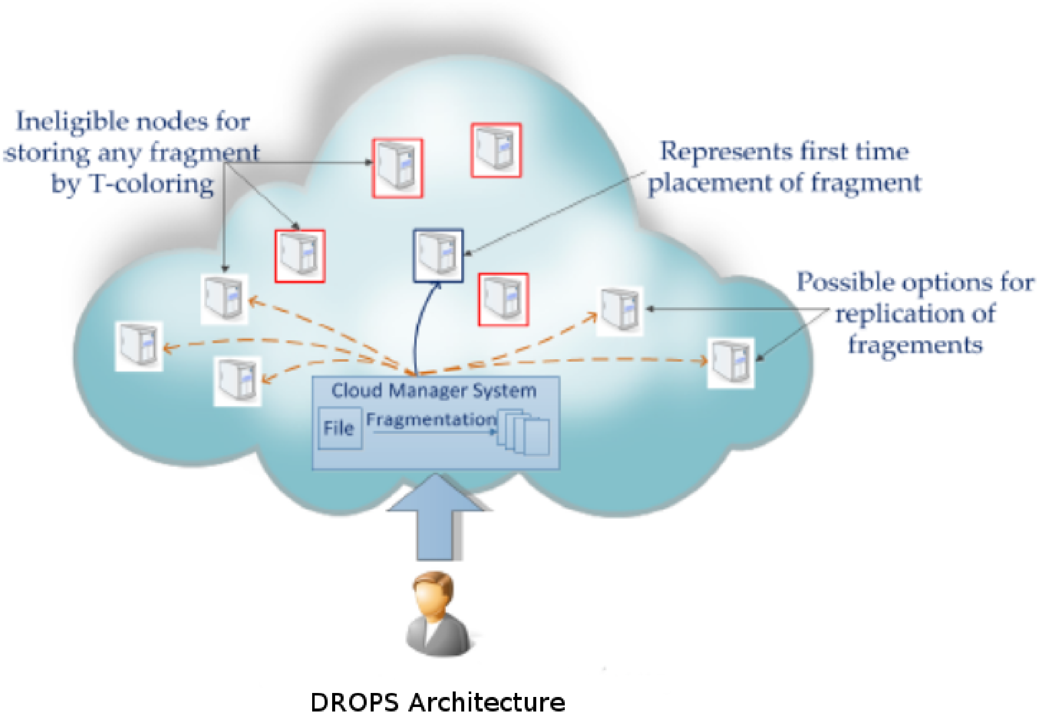

Ineligible nodes for storing any fragment by T-coloring
Represents first time placement of fragment
Possible options for replication of fragements
Cloud Manager System
File
Fragmentation
DROPS Architecture

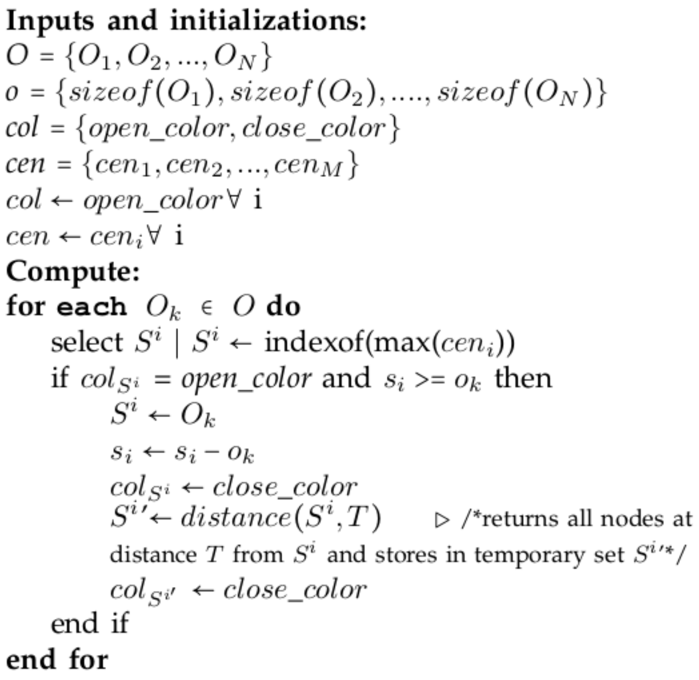
Figure 6.1: DROPS Architecture

6.2.2 Replication

This approach creates replicas (duplicate copy) of fragments. These replicas are useful when one of fragment is corrupted by attacker then to provide file for user admin replaces its replica at that place and combine all fragments and send file to authenticated user or data owner. To make replicas of file fragments this approach runs replication algorithm which takes input as fragments and produces its replicas as output.

6.2.3 Allocation

After the file is spitted and replicas are generated then allocate that fragments at cloud server for storing data. While storing or allocating that fragments, consider security issues. So the proposed method using T- Coloring Graph concept for placing fragments at different nodes on cloud server. This approach runs Fragment allocation algorithm which takes input as fragments and produces the output as fragments allocated with node numbers.



```
Inputs and initializations:
$O = \{O_1, O_2, ..., O_N\}$
$o = \{sizeof(O_1), sizeof(O_2), ...., sizeof(O_N)\}$
$col = \{open\_color, close\_color\}$
$cen = \{cen_1, cen_2, ..., cen_M\}$
$col \leftarrow open\_color \forall$ i
$cen \leftarrow cen_i \forall$ i
Compute:
for each $O_k \in O$ do
    select $S^i$ | $S^i \leftarrow$ indexof(max($cen_i$))
    if $col_{S^i}$ = $open\_color$ and $s_i >= o_k$ then
        $S^i \leftarrow O_k$
        $s_i \leftarrow s_i - o_k$
        $col_{S^i} \leftarrow close\_color$
        $S^{i\prime} \leftarrow distance(S^i, T)$    ▷ /*returns all nodes at
        distance $T$ from $S^i$ and stores in temporary set $S^{i\prime}$*/
        $col_{S^{i\prime}} \leftarrow close\_color$
    end if
end for
```

Figure 6.2: Algorithm for Fragment Placement

6.3 Admin

Admin is an authorized person who has rights to validate authorized data owner and user. He is also responsible for allocation of block and maintains information and authentication.

6.4 Fragment Placement

In the DROPS methodology, not to store the entire file at a single node. The DROPS methodology fragments the file and makes use of the cloud for replication. The fragments are distributed such that no node in a cloud holds more than a single fragment, so that even a successful attack on the node leaks no significant information. The DROPS methodology uses controlled replication where each of the fragments is replicated only once in the cloud to improve the security. Although, the controlled replication does not improve the retrieval time to the level of full-scale replication, it significantly improves the security.The fragment placement strategy is presented in the algorithm given below.

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, lose some of the central nodes that may increase the retrieval time but achieve a higher security level. If somehow the intruder compromises a node and obtains a fragment, then the location of the other fragments cannot be de- termined. The attacker can only keep on guessing the location of the other fragments. However,the probability of a successful coordinated attack is extremely minute. The pro- cess is repeated until all of the fragments are placed at the nodes. The above algorithm represents the fragment placement methodology.

6.5 Fragment Replication

In addition to placing the fragments on the central nodes, this approach also per- form a controlled replication to increase the data availability, reliability, and improve data retrieval time. Place the fragment on the node that provides the decreased access cost with an objective to improve retrieval time for accessing the fragments for reconstruction of original le. While replicating the fragment, the separation of fragments as explained in the placement technique through T- coloring, is also taken care off.

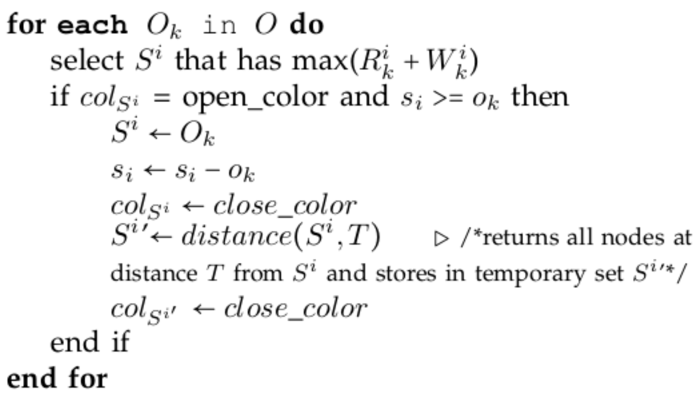
In case of a large number of fragments or small number of nodes, it is also possible that some of the fragments are left without being replicated because of the T-coloring. As discussed previously, T-coloring prohibits to store the fragment in neighborhood of a node storing a fragment, resulting in the 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. The replication strategy is presented in the algorithm given below.

6.6 Replication Time

Aim is to minimize the overall total network transfer time or replication time (RT) or also termed as replication cost (RC). The RT is composed of two factors: (a) time due to read requests and (b) time due to write requests.

- The total read time of Ok by Si from NNi k is denoted by Ri k and is given by:

$$R^i_k = r^i_k o_k c(i, NN^i_k)$$



```
for each $O_k$ in $O$ do
    select $S^i$ that has max($R^i_k + W^i_k$)
    if $col_{S^i}$ = open_color and $s_i$ >= $o_k$ then
        $S^i \leftarrow O_k$
        $s_i \leftarrow s_i - o_k$
        $col_{S^i} \leftarrow close\_color$
        $S^{i\prime} \leftarrow distance(S^i, T)$    ▷ /*returns all nodes at
        distance $T$ from $S^i$ and stores in temporary set $S^{i\prime}$*/
        $col_{S^{i\prime}} \leftarrow close\_color$
    end if
end for
```

Figure 6.3: Algorithm for Fragment Replication

- The total time due to the writing of Ok by Si ad- dressed to the Pk is represented as Wi k and is given:

$$W_k^i = w_k^i o_k (c(i, P_k) + \sum_{(j \in \neq R_k), j \neq i} c(P_k, j))$$

- The overall RT is represented by:

$$RT = \sum_{i=1}^{M} \sum_{k=1}^{N} (R_k^i + W_k^i)$$

- The storage capacity constraint states that a file fragment can only be assigned to a node, if storage capacity of the node is greater or equal to the size of fragment.

- To handle the download request from user, the cloud manager collects all the fragments from the nodes and reassemble them into a single file. Afterwards, the file is sent to the user.

6.7 Features of DROPS

- It ensures that even in case of a successful attack, no meaningful information is revealed to the attacker.

- A successful attack on a single node must not reveal the locations of other fragments within the cloud.

- The nodes storing the fragments, are separated with certain distance by means of graph T-coloring to prohibit an attacker of guessing the locations of the fragments.

- Does not rely on the traditional cryptographic techniques for the data security.
- Faster
- Higher level of security with slight performance overhead.

Chapter 7

Conclusion and future scope

The DROPS methodology is a cloud storage security scheme that jointly deals with the security and performance in terms of retrieval time. The data file was fragmented and the fragments are scattered over multiple nodes. The nodes were separated by means of T-coloring. The fragmentation and dispersal make sure that no significant information was obtainable by an antagonist in case of a successful attack. No node in the cloud, stored more than a single fragment of the same file.

The scope of this work is divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a single fragment of a particular data file that ensures that even in case of a successful attack, no meaningful information is revealed to the attacker.

Bibliography

[1] Lade Mayuri S.Lolage Pranali S. Nagare Mayuri S.Sadaphal Priyanka V. Prof.Jorwekar Y.S, *"An Study on-Cloud Security Using DROPS Technique"*, Vol. 2 Issue. 6, 2016.

[2] Priya Patni, Dr. S.N Kakarwal , *"Optimal performance and security by division and replication of data in cloud"*, Global Journal of Engineering Science and Research Management, Nov. 2015.

[3] Nidhi Jain, Archana jadhav, *"A Survey Paper on Drops: Division and Replication of Data in Cloud for Optimal Performance and Security"* , International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, Oct. 2016

[4] Mazhar Ali, K. Bilal, S. U. Khan, B. Veeravalli, K. Li, Albert Y. Zomaya and A. J. R. Neves, *"dropd: Division and Replication of Data in the Cloud for Optimal Performance and Security"*, IEEE transactions of cloud computing., vol. 13, no. 11, pp. 14111418, Nov. 2015.