9th International Conference on Recent Development in Engineering Science, Humanities and Management
Mahratta Chamber of Commerce, Industries and Agriculture , Pune (India) ESHM-17
23th December 2017, www.conferenceworld.in ISBN: 978-93-86171-88-7

Increasing Security and Performance of Cloud Storage Using Data Division and Replication Strategies

Suyog Ghodey[1], Niranjan Dandekar[2], Sagar Kate[3], Vikrant Bhalerao[4], Shubham Bhange[5], Alka Londhe[6]

[1,2,3,4,5,6]*Department of Computer Engineering,*
Pimpri Chinchwad College of Engineering, Pune, (India)

ABSTRACT

The security of large scale systems, such as the Cloud depends upon the security of the whole system and security of individual nodes. A successful intrusion into a single node may have severe consequences, not only for data of host node but also for other nodes. As storage is the most precious factor in the Clouds, data fragments having same values can be considered as redundant. In this paper we will propose a system which will implement a new storage strategy that uses the Division and Replication methodology for storing the data.

In this system, the file will be divided into fragments and then those fragments will be replicated several times according to the replication factor before storing on the cloud. The fragments are then 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.

Furthermore, this methodology does not rely on the traditional Cryptographic techniques for the data security; thereby relieving the system efficiency in clouds. The system will analyze the performance parameters by simulating it in a controlled environment of computationally expensive methodologies.

This system will enhance cloud security using Division and Replication of data in Cloud for Optimal Performance and Security (DROPS) methodology. The system will increase space.

Index Terms— Cloud, Data Fragments, Redundancy, Security, Space Efficiency, T-coloring.

I.INTRODUCTION

Cloud computing is an information technology (IT) paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility.

Cloud computing poses privacy concerns because the service provider can access the data that is in the cloud at any time. It could accidentally or deliberately alter or even delete information. Many cloud providers can share information with third parties if necessary for purposes of law and order even without a warrant. That is permitted in their privacy policies, which users must agree to before they start using cloud services. Solutions to

privacy include policy and legislation as well as end users' choices for how data is stored. Users can encrypt data that is processed or stored within the cloud to prevent unauthorized access. [1]

Cloud computing and storage provides users with capabilities to store and process their data in third-party data centres. Organizations use the cloud in a variety of different service models (with acronyms such as SaaS, PaaS, and IaaS) and deployment models (private, public, hybrid, and community). Security concerns associated with cloud computing fall into two broad categories: security issues faced by cloud providers (organizations providing software-, platform-, or infrastructure-as-a-service via the cloud) and security issues faced by their customers (companies or organizations who host applications or store data on the cloud). The responsibility is shared, however. The provider must ensure that their infrastructure is secure and that their clients' data and applications are protected, while the user must take measures to fortify their application and use strong passwords and authentication measures. [2]

Cloud storage is based on highly virtualized infrastructure and is like broader cloud computing in terms of accessible interfaces, near-instant elasticity and scalability, multi-tenancy, and metered resources. Cloud storage services can be utilized from an off-premises service (Amazon S3) or deployed on-premises (ViON Capacity Services).Cloud storage typically refers to a hosted object storage service, but the term has broadened to include other types of data storage that are now available as a service, like block storage.

Cloud storage is made up of many distributed resources, but still acts as one, either in a federated or a cooperative storage cloud architecture. It is highly fault tolerant through redundancy and distribution of data. It is highly durable through the creation of versioned copies. It typically eventually consistent with regard to data replicas.

This paper will aim to improve upon these current storage and security policies. Therefore, in this paper, we collectively approach the issue of security and performance as a secure data replication problem. We present Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that judicially fragments user files into pieces and replicates them at strategic locations within the cloud. The division of a file into fragments is performed based on a given user criteria such that the individual fragments do not contain any meaningful information. Each of the cloud nodes (we use the term node to represent computing, storage, physical, and virtual machines) contains a distinct fragment to increase the data security. A successful attack on a single node must not reveal the locations of other fragments within the cloud. To keep an attacker uncertain about the locations of the file fragments and to further improve the security, we select the nodes in a manner that they are not adjacent and are at certain distance from each other. The node separation is ensured by the means of the T-coloring. [3]

To improve data retrieval time, the nodes are selected based on the centrality measures that ensure an improved access time. To further improve the retrieval time, we judicially replicate fragments over the nodes that generate the highest read/write requests. We develop a scheme for outsourced data that takes into account both the security and performance. The proposed scheme fragments and replicates the data file over cloud nodes. The proposed DROPS scheme ensures that even in the case of a successful attack, no meaningful information is

revealed to the attacker. We do not rely 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. We ensure a controlled replication of the file fragments, where each of the fragments is replicated only once for the purpose of improved security. [4].

II.ALGORITHMS USED

1) Fragment Replication

Replication technique called sole replication used, that every fragment is replicated only once. Sole replication increase the security without having multiple copies of fragments thereby reducing the unnecessary storage space. Node selection and fragment placement for replicated fragments is same as the original file fragments. Data owner will select the number of nodes greater than the number replicated fragments as done while fragmenting. By repeating the same process will avoid the storage of any replicated fragments in any node holding the original fragment. Algorithm represents the process of sole replication.

Algorithm for fragment′s replication

for each Ok in O do

select Si that has max(Rik+Wik)

ifcolSi = open_color and si>=ok then

Si←Ok

si←si−ok

colSi←close_color

Si'←distance(Si,T)

colSi'←close_color

end if

end for

2) File Fragmentation

Being specified the fragmentation threshold value by the user or considered default percentage through SLA the fragmentation process will be preceded. With those value CMS will spilt the file into Number of fragments. Once the file is fragmented, this methodology selects the cloud nodes for fragment placement. The selection is made by keeping an equal focus on both security and performance in terms of the access time. Using

fragmentation technique and choosing distinct node for every single fragment enhance the security without using any cryptographic techniques. Data owner will provide details to the user about the file size, number pages and number of fragments. This information is provided in order to user choose to provide data owner access or to decline. [5]

3) Graph coloring concept

Node selection places an important part inplacing the fragments. Placements of fragments aresignificant because for improving data retrieval timethese fragments should be highly reliable. Hereproposed the concept of hotbed measure. In this acentral node is selected primarily which called ashotbed node of the network. To find central node thenode should be less eccentric. Eccentricity measure ofthe node is the maximum distance of other node fromthe selected node n. [6]T-coloring concept is used, which helpsin improving data retrieval time and security. Fragmentplacement is done securely by selecting nodes whichare separated by a distance to avoid interference. In thisconcept being selected the hotbed node, selection ofnext node for the placement of next fragment is done insuch way that all adjacent node to the hotbed node willbecome ineligible. In figure 1, nodes boxed in red colourare adjacent to the hotbed node, which becomeineligible to place next fragment. Nodes boxed in greencolour are eligible for next fragment placement becauseit is not adjacent to the hotbed node. This process iscarried out for upcoming fragments of that file. Byplacing fragments using these techniques avoidsleakage of information to attackers. Even in case ofsuccessful intrusion of single node does not provideany clue to next fragment location. [6]

Algorithm for fragment placement

O = {O1;O2; :::;ON}

o = {sizeof(O1); sizeof(O2); :::::; sizeof(ON)}

col = {open color; close color}

cen = {cen1; cen2; :::; cenM}

col _ open color¦ *i*

cen _ ceni¦ *i*

Compute:

for each Ok > O do

select Si S Si _ indexof(max(ceni))

ifcolSi = open color and si>= ok then

Si _ Ok

si _ si − ok

colSi _ close color

*Si'_ distance(Si; T) P /*returns all nodes at*

distance T from Si and stores in temporary set Si'/*

colSi. _ closecolor

end if

end for

III.RELATED WORK

Juels 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 [10] heavily depends on the users employed 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. [7] The authors in [11] 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 proposed 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 handled. The DROPS methodology handles the leakage of critical information by fragmenting data file and using multiple nodes to store a single file. The use of a trusted third party for providing security services in the cloud is advocated. The authors used the public key infrastructure (PKI) to enhance the level of trust in the authentication, integrity, and confidentiality of data and the communication between the involved parties. The keys are generated and managed by the certification authorities. At the user level, the use of temper proof devices, such as smart cards was proposed for the storage of the keys. Similarly, Tang et. al. have utilized the public key cryptography and trusted third party for providing data security in cloud environments. However, the authors have not used the PKI infrastructure to reduce the overheads. The trusted third party is responsible for the generation and management of public/private keys. The trusted third party may be a single server or multiple servers. The symmetric keys are protected by combining the public key cryptography and the (k, n) threshold secret sharing schemes. Nevertheless, such schemes do not protect the data files against tempering and loss due to issues

arising from virtualization and multi-tenancy. [8] A secure and optimal placement of data objects in a distributed system is presented here. An encryption key is divided into n shares and distributed on different sites within the network. The division of a key into n shares is carried out through the (k, n) threshold secret sharing scheme. The network is divided into clusters. The number of replicas and their placement is determined through heuristics. A primary site is selected in each of the clusters that allocates the replicas within the cluster. The scheme presented combines the replication problem with security and access time improvement. Nevertheless, the scheme focuses only on the security of the encryption key. The data files are not fragmented and are handled as a single file. The DROPS methodology, on the other hand, fragments the file and store the fragments on multiple nodes. Moreover, the DROPS methodology focuses on the security of the data within the cloud computing domain that is not considered in [9]

IV.SYSTEM MODEL

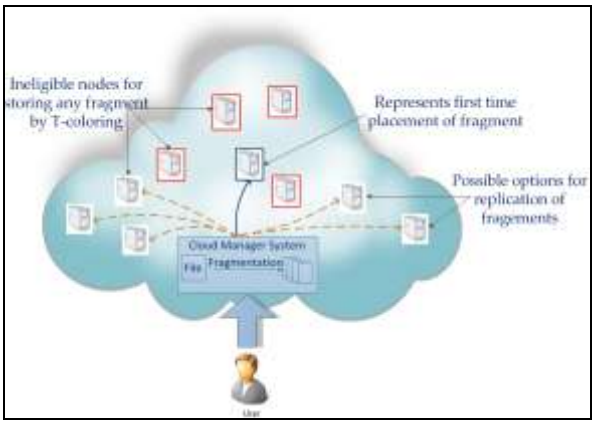

Fig 1: Cloud management system

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 neighbourhood 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.

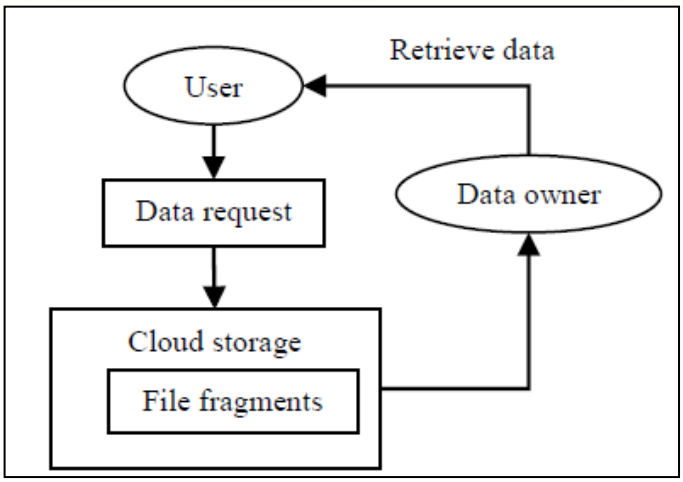

Fig 2: System Flow

V.CONCLUSION

We proposed the DROPS methodology, a cloud storage security scheme that collectively deals with the security and performance in terms of retrieval time. The data file was fragmented and the fragments are dispersed over multiple nodes. The nodes were separated by means of T-coloring. The fragmentation and dispersal ensured that no significant information was obtainable by an adversary in case of a successful attack. No node in the cloud, stored more than a single fragment of the same file. The performance of the DROPS methodology was compared with full-scale replication techniques. The results of the simulations revealed that the simultaneous focus on the security and performance, resulted in increased security level of data accompanied by a slight performance drop.

REFERENCES

[1] Ali, Mazhar, Kashif Bilal, Samee Khan, BharadwajVeeravalli, Keqin Li, and Albert Zomaya. "*DROPS: Division and Replication of Data in the Cloud for Optimal Performance and Security*." IEEE Transactions on Cloud computing (2015).

[2] Wang, Shulan, Kaitai Liang, Joseph K. Liu, Jianyong Chen, Jianping Yu, and WeixinXie. "*Attribute-based data sharing scheme revisited in cloud computing.*" IEEE Transactions on Information Forensics and Security 11, *no. 8 (2016): 1661-1673*.

[3]Salunkhe, Sujata D., and Dhanshri Patil. *"Division and replication for data with public auditing scheme for cloud storage."* In Computing Communication Control and automation (ICCUBEA), 2016 International Conference on, *pp. 1-5. IEEE, 2016*.

[4]Singla, Sanjoli, and Jasmeet Singh. "*Cloud data security using authentication and encryption technique*." Global Journal of Computer Science and Technology 13, no. 3 (2013).

[5] Moral, W. Delishiya, and B. Muthu Kumar. "*Improve the data retrieval time and security through fragmentation and replication in the cloud*." In Advanced Communication Control and Computing Technologies (ICACCCT), 2016 International Conference on, *pp. 539-545. IEEE, 2016*.

[6] Kang, Seungmin, BharadwajVeeravalli, and KhinMiMiAung. "*A Security-aware Data Placement Mechanism for Big Data Cloud Storage Systems*." In Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on, *pp. 327-332. IEEE, 2016*.

[7]Boru, Dejene, DzmitryKliazovich, FabrizioGranelli, Pascal Bouvry, and Albert Y. Zomaya. "*Models for efficient data replication in cloud computing datacenters*." In Communications (ICC), 2015 IEEE International Conference on, *pp. 6056-6061. IEEE, 2015*.

[8]Singla, Sanjoli, and Jasmeet Singh. "*Cloud data security using authentication and encryption technique*." Global Journal of Computer Science and Technology 13.3 (2013).

[9]*https://en.wikipedia.org/wiki/T-coloring* accessed on 10th December, 2017.