Replication and Automatic Updates of File Fragments in Cloud

Siri Mallugari, Kriti Ohri, N. Sai Poojitha

Abstract: Generally, in cloud computing the data is outsourced to controls of third party, which may lead to issue of securitys. The data in cloud may lost because of attacks by unknown users and nodes inside the cloud .Therefore, the data within in the cloud must be kept in secured state by using different techniques .In this paper, we propose division and replication of data within in the cloud to improve optimal performance and security .In the present work we separate a file into multiple fragment sand replicate fragments on different nodes in the cloud. Each node, should store fragment only once. Therefore, no information can be get by the attacker. The node storing the fragments within the cloud are separated using a method called T-colouring, which avoids the attacker to find the location of nodes that contain related fragments. The traditional cryptographic techniques for data security are not used in this work, so that computationally expensive methodologies are reduced. The more advanced feature in this work is an automatic update of file fragments on nodes whenever any changes are done by the third person on that particular node.

Keywords: Cloud Computing, File Fragments, Replication, nodes, Performance, Security.

I. INTRODUCTION

Security is one of the most urgent angles among those precluding the wide-spread appropriation of cloud computing The neighbouring substances may give an chance to an assailant to sidestep the clients protections. The information redistributed to an open cloud must be verified. Unapproved access by third persons must be reduced. Any kind of information can be placed in cloud. There may be chances for the data to be get corrupted. Apart from these things there are many advantages with using cloud, it provides elasticity, cost effectiveness and many other services.

The information which is to be placed in cloud storage must be in secured state .Unauthorized access must be prevented to other users and process. There must be reduction in data loss. The cloud provides many advantages to the users like reliability, confidentiality and security. Therefore, many people are migrating to cloud in now-a-days. In this, we considered the problems in security and execution of safe information replication issue. The fig 1, demonstrates the overall structure of cloud computing

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Siri Mallugari*, Post-graduate student, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India.

Mrs Kriti Ohri, Assistant Professor, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India.

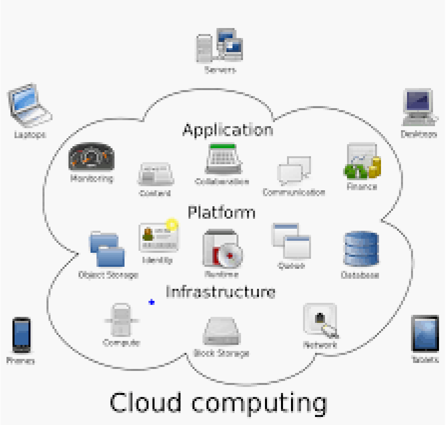


Fig. 1 cloud computing.

II. BACKGROUND

A. Cloud Computing

The cloud computing worldview has changed the use also, the board of the framework of data innovation. Cloud computing can be explained as on demand administrations, pooling of assets, flexibility, and estimated administrations. The formerly mentioned attributes of cloud computing make it more suitable for organizations, and particular clients. Be that as it may, provide features like low cost, immaterial administration (from clients point of view), and prominent adaptability considers improved security concerns.

B. File Fragmentation

File fragmentation is division of file into multiple parts based on the user request. The main purpose of file fragmentation is to provide security and increase performance. The fragments are located such that no node holds more than a fragment in cloud, so the attacker cannot find the appropriate location of a fragment even after an attack also. This methodology includes controlled replication where replication is done only for a fragment in the cloud to improve the security. The fragment size is determined by owner of the file. The owner can determine the discontinuity limit as far as either rate or the number and size of various pieces. The rate discontinuity limit, for example, can manage that each fragment will be of 5 percent size of the absolute size of the file.

III. METHODOLOGY

A cloud must guarantee throughput, unwavering quality, and security. The choice of the nodes is performed in two stages.

In the principal stage, the nodes are chosen for the introductory arrangement of the sections dependent on the centrality measures. In the subsequent stage, the hubs are chosen for replication.

The major publications in this paper are

1. Initially, the file is split into fragments and placed on different nodes with in the cloud.
2. As the nodes are separated by T-colouring the attacker cannot easily find the fragment location.
3. Each node stores only one fragment and the fragment is replicated only once to improve the security.
4. The main contribution is the fragments which are modified by the third persons will be updated automatically.

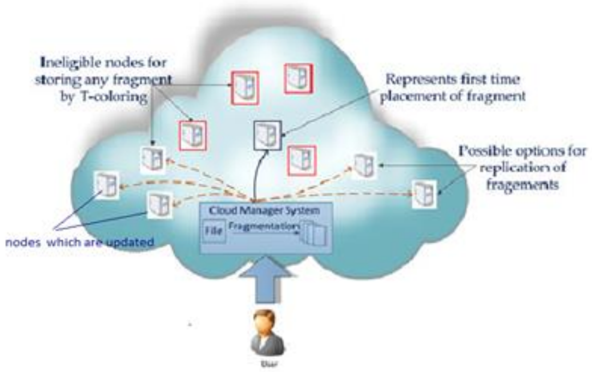

Ineligible nodes for storing any fragment by T-coloring
Represents first time placement of fragment
Possible options for replication of fragements
nodes which are updated
Cloud Manager System
File
Fragmentation
User

Fig. 2. Architecture

Assume a cloud with M number of nodes, each with its own capacity and also N no. of fragments with the end goal that O_k means kth piece of a document while o_k speaks to the size of kth part.

Table-I : Assumptions

Symbols	Meanings
M	Total number of nodes in the cloud
N	Total number of file fragments to be placed
O_k	kth fragment of file
o_k	Size of O_k
S^i	ith node
s_i	Size of S^i
cen_i	Centrality measure for S^i
col_{S^i}	Color assigned to S^i
T	A set containing distances by which assignment of fragments must be separated
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 neighbor of S^i holding O_k
$c(i,j)$	Communication cost between S^i and S^j
P_k	Primary node for O_k
R_k	Replication schema of O_k
RT	Replication time

IV. LITERATURE REVIEW

In this area other scientist's work which is important to the present investigation are displayed.

Manghui Tu [1], In this paper, an encryption key is isolated into n shares also, circulated on various locales inside the system. The system is separated into bunches and essential site is chosen in every one of the groups that distributes the imitations inside the bunch. Concentrates just on the security of encryption key.

Alessandro Mei [2], this paper proposed a model to gauge the confirmation level of static and dynamic record assignment plans in a huge scale dispersed framework and thinks about high assurance, availability, performance and adaptability. Concentrates just on powerful record portion and does not identify vindictive hub.

Nirmal kumar gupta [3], In this paper,the open key framework is utilized to improve the degree of trust in authentication,integrity and privacy of information and correspondence between included gatherings. Believed outsider is liable for age and management of open/private keys. There might be loss of information because of issues emerging from virtualization. Here the files are not divided.

Yang tang [4], In this paper, the believed outsider produces open/private keys. To accomplish security goals ,it is based upon a lot of cryptographic key tasks that are self-kept up by a majority of key directors that are free of outsider mists. This plan don't secure information records against treating.

Ari juels [5] This paper, centers around integrity ,freshness and accessibility of information in cloud. Here iris document framework is utilized for information movement into cloud and uses a passage application in the association to guarantee trustworthiness and freshness of information. It vigorously depends on client's utilized.

Qiu [6] In this paper, the online issue of setting web server copies in substance conveyance systems (CDNS) is examined, to limit the expense for customers to get to information recreated on servers. Here the expense speak to latency, hop count, etc. A voracious calculation is utilized for illuminating the web server copy situation problem .Using the given calculation there perhaps chance for an assailant to figure the areas of reproductions.

Dejene Boru [7] In this paper information replication in distributed computing server farms is considered. The replication depends on Developed Mathematical model and re-enactments utilizing green cloud ,the test system concentrating on vitality efficiency and correspondence process. Concentrates just on vitality productivity and transfer speed. It does not think about security of information inside the cloud.

Giorgos Kappe [8] In this paper, virtualized and multitenancy related issues are considered. In this embankment approval is utilized for local access control and occupant namespace isolation. The spillage of basic data incase of noxious vm isn't dealt with. Information documents are most certainly not divided and took care of as a solitary record.

Lo’ai Tawalbeh , Nour S. Darwazeh [9] This paper, proposed a proficient privacy based distributed storage structure that upgrades the handling time and guarantees secrecy and trustworthiness through information order and applying TLS, AES and SHA dependent on the kind of grouped information. This paper does not incorporate programmed information grouping.

Ali [10] This procedure gives information confidentiality, integrity, access control, data sharing.

In this cryptographic systems are utilized and encryption/unscrambling is finished by confided in outsider. This approach totally relies upon cryptographic methods.

V. METHODOLOGY FOR PROPOSED MODEL

A. Problem Definition

High safety efforts are required to ensure information inside the cloud. In the current framework, the information records are not divided and taken care of as single document. Some current frameworks just spotlight on security of the encryption keys however not on the security of information. In the present frameworks, client needs to download the record, update substance and transfer it once more ,which is tedious.

B. Proposed Model Methodology

The proposed model uses the following methodology to improve security and performance within the cloud by fragmenting file into multiple parts.

Step-1: The file (any type of file) which is to be placed in cloud is to be divided into multiple fragments.

Step-2: Nodes on which fragments should be placed are identified .

Step-3: Each fragment is placed on identified nodes within the cloud.

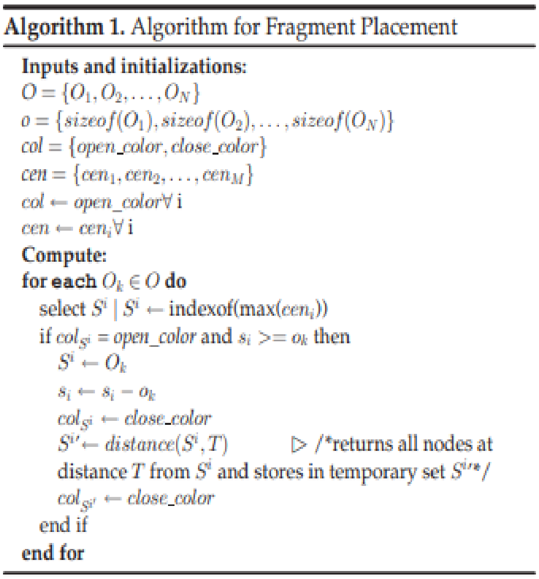
Step-4: Particular fragment is replicated only once with in the cloud.

Step-5: The fragment gets updated automatically whenever any changes are done by the attacker to that particular fragment.

1) file fragmentation

The fragment limit of the information record is determined to be produced by the document proprietor. The record proprietor can determine the discontinuity limit as far as either rate or the number or size of various sections. The rate discontinuity limit, for example, can manage that each section will be of 5 percent size of the all out size of the document. The proprietor can best part the document with the end goal that each section does not contain critical measure of data as the proprietor is conscious of the considerable number of realities relating to the information. The default rate fracture edge can be made a some portion of the administration level understanding (SLA), if the client does not determine the fracture edge while transferring the information document.

The following fig 3 represents the algorithm to perform fragmentation of files. In this O represents inputs, o represents size of inputs, cen represents the centrality of the hubs.



```
Algorithm 1. Algorithm for Fragment Placement
Inputs and initializations:
$O = \{O_1, O_2, \ldots, O_N\}$
$o = \{sizeof(O_1), sizeof(O_2), \ldots, sizeof(O_N)\}$
$col = \{open\_color, close\_color\}$
$cen = \{cen_1, cen_2, \ldots, cen_M\}$
$col \leftarrow open\_color \forall$ i
$cen \leftarrow cen_i \forall$ i
Compute:
for each $O_k \in O$ do
  select $S^i \mid S^i \leftarrow$ indexof(max($cen_i$))
  if $col_{S^i} = open\_color$ and $s_i >= o_k$ then
    $S^i \leftarrow O_k$
    $s_i \leftarrow s_i - o_k$
    $col_{S^i} \leftarrow close\_color$
    $S^{i'} \leftarrow distance(S^i, T)$        ▷ /*returns all nodes at
    distance $T$ from $S^i$ and stores in temporary set $S^{i'}$*/
    $col_{S^{i'}} \leftarrow close\_color$
  end if
end for
```

Fig. 3. Algorithm for file fragmentation

2) Node Selection

The nodes within the cloud can be selected by using two methods :centrality and T-colouring.

1. Centrality

The centrality of hub in diagram gives the proportion of overall significance of a hub in network. Different centrality measures are available; for example, closeness centrality, degree centrality, betweenness centrality, erraticism centrality, and eigenvector centrality

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

here,

δ_{ab} → Represents total no. of shortest paths between a, b.

$\delta_{ab}(v)$ → Represents the no. of shortest paths between a, b passing through v.

$C_b(v)$ → Represents the betweenness centrality for node v.

2. T-colouring

T-coloring is a mapping function which assigns colors to vertices. i.e.,the disance between the colors of adjacent vertices must not belong to T,where T is a set of non-negative integers.

3) file Replication

Put the piece on hub that gives diminished cost for access with a target to modify recovery time for getting to the sections for reproduction of unique record. While duplicating the section, the division of parts as clarified in the position procedure through T-colouring. On the off chance that of countless pieces or modest number of hubs, it is additionally conceivable that a portion of the pieces are left without being duplicated as a result of the T-shading.

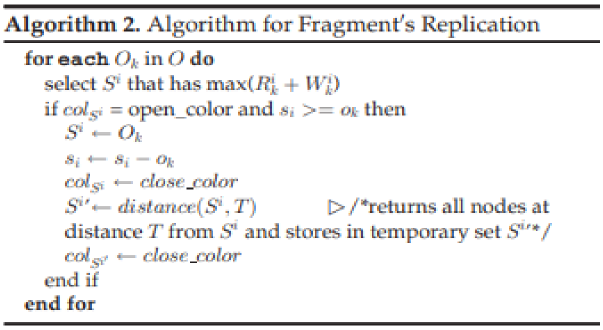
A fragment is replicated on a node which contains maximum read and write costs.

$$R_k^i = r_k^i o_k c(i, NN_k^i)$$

Retrieval Number: A1789109119/2019©BEIESP
DOI: 10.35940/ijeat.A1789.109119
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$w_k^i = w_k^i o_k(c(i,p_k)+\sum_{(j\varepsilon R_k),j=i} c(p_k,j))$



Algorithm 2. Algorithm for Fragment's Replication

```
for each $O_k$ in $O$ do
    select $S^i$ that has max$(R_k^i + W_k^i)$
    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
```

Fig. 4 Algorithm for file replication

C. Results

In this study other models were implemented and a comparison of no. of nodes, no.of fragments and read/write ratio is done.

1) Improvement in number of nodes

The quantities of hubs chose for the recreations are 100,1024,30000,2400 and 500. The quantity of hubs in the Dcell design increments exponentially. For a Dcell design, with 2 hubs in the Dcell_0, the design comprises 2,400 hubs.

Be that as it may, expanding a solitary hub in the Dcell_0, the aggregate hubs increments to 30,000 . The quantity of document pieces was set to 50. For the principal explore we utilized C = 0.2. Figs. 5, 6 and 7 demonstrate the outcomes for 3 level, Dcell ,Fat tree, designs, separately.

Fascinating perception is, albeit the majority of the calculations indicated comparative pattern in execution inside a particular design, the presentation of the calculations was better in the Dcell engineering when contrasted with three level and fat tree models. This is on the grounds that the Dcell engineering displays better bury hub availability and heartiness.

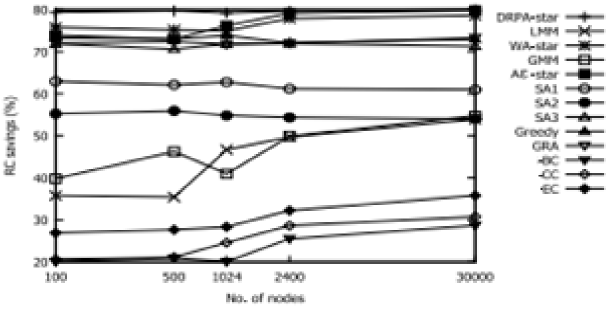


Fig. 5 comparison between RC and no. of nodes with Dcell

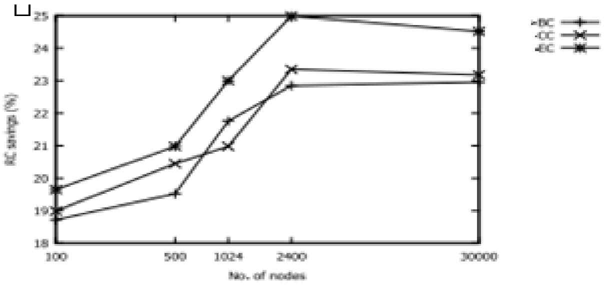


Fig. 6 comparison between RC and no. of nodes with three tier

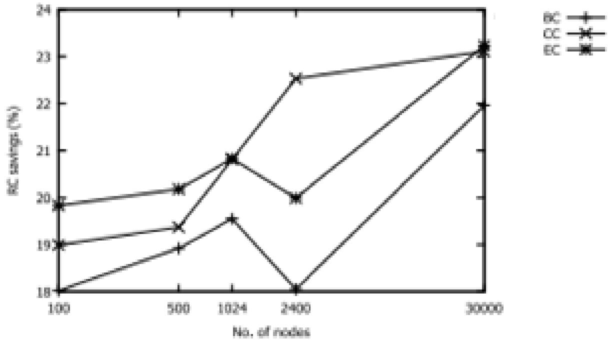


Fig. 7 comparison between RC and no. of nodes with fat free.

2) Improvement in number of nodes

The number of nodes can say the capacity limit of cloud that, thusly may influence the choice of the nodes. The outstanding task at hand was produced with C = 45% to watch the impact of increment number of record parts with genuinely sensible measure of memory and to perceive the exhibition of the considerable number of calculations. The outcomes are appeared in Figs. 8, 9 and 10. However, from Dcell design, plainly this strategy with whimsy centrality keeps up the matchless quality on the centralities of other two.

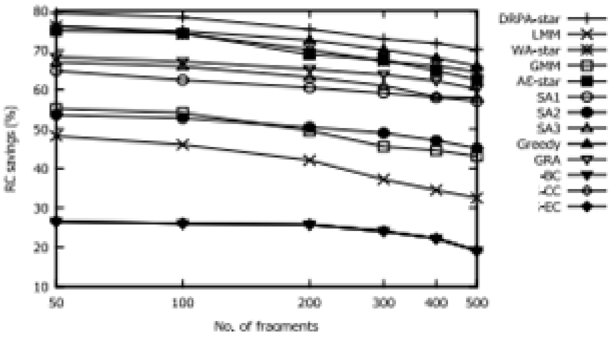


Fig. 8 comparison between RC and no. Of file fragments with three tier.

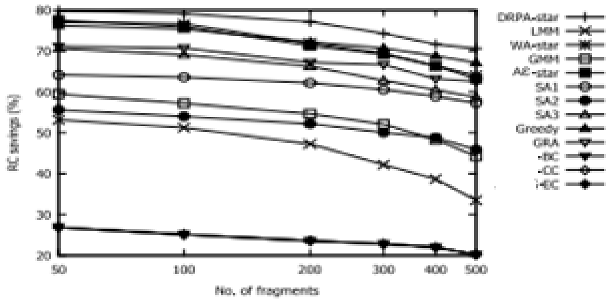


Fig. 9 comparison between RC and no. of file fragments with fat free.

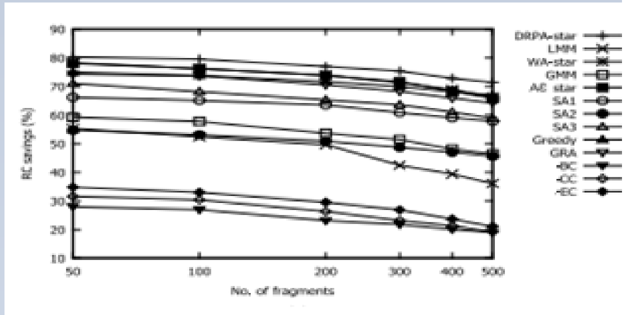


Fig. 10 comparison between RC and no. of file fragments with Dcell.

3) Improvement in Read/write ratio

The adjustment in R/W proportion influences the exhibition of the talked about similar procedures. Be that as it may, the expanded number of composes requests that the imitations be put nearer to the essential hub. The nearness of copies nearer to the essential hub brings about diminished RC related with refreshing reproductions. The higher compose proportions may expand the traffic on the system for refreshing the reproductions. Figs. 11, 12 and 13 demonstrate the presentation of the relative strategies Reduction in quantity of composes caused decrease of expense related by refreshing the imitations of pieces. Nonetheless, the majority of the similar systems demonstrated a type of diminishing in RC putting something aside for R/W proportions above 0.50.Therefore, the worldwide view of these calculations brought about superior.

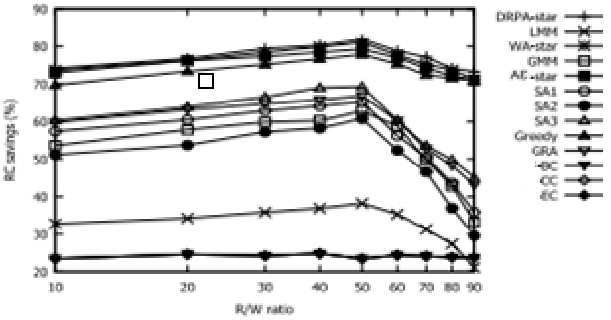


Fig. 11 comparison between RC and R/W ratio for three tier

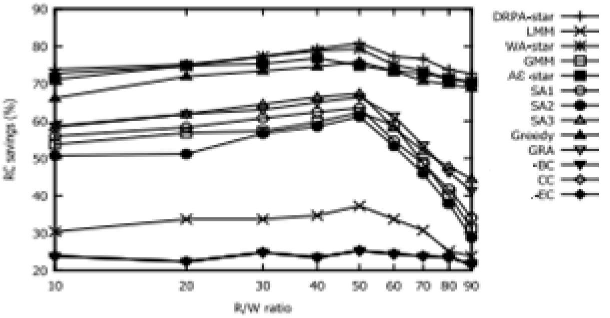


Fig. 12 comparison between RC and R/W ratio for Fat free

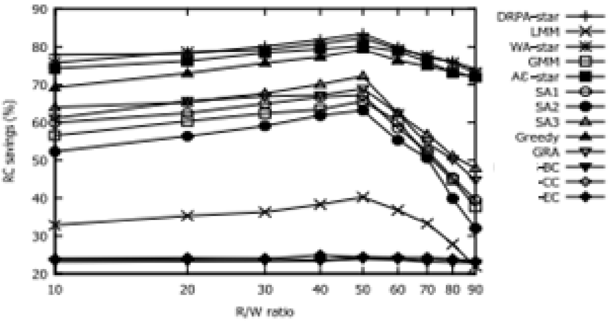


Fig. 13comparison between RC and R/W ratio for Dcell

VI. CONCLUSIONS

The proposed strategy is a cloud storage scheme of security that largely manages security and issues of performance. In this approach, the data files (the files can be in any format) were divided and fragments are transferred on multiple nodes and furthermore altered parts are refreshed automatically. The nodes were isolated by methods for T-colouring. The discontinuity and dispersal guaranteed that no critical data was realistic by a foe if there should arise an occurrence of a fruitful assault. No node in the cloud, put away more than a solitary piece of a similar document. The effects of reproductions uncovered the security of synchronous spotlight and execution, increased security level.

REFERENCES

1. M. Tu, P. Li, Q. Ma, I.-L. Yen, and F. B. Bastani, “On the optimal placement of secure data objects over Internet,” in Proc. 19th IEEE Int. Parallel Distrib. Process. Symp., 2005, p. 14.
2. A. Mei, L. V. Mancini, and S. Jajodia, “Secure dynamic fragment and replica allocation in large-scale distributed file systems,” IEEE Trans. Parallel Distrib. Syst., vol. 14, no. 9, pp. 885–896, Sep. 2003.
3. D. Zissis and D. Lekkas, “Addressing cloud computing security issues,” Future Gener. Comput. Syst., vol. 28, no. 3, pp. 583–592, 2012.
4. Y. Tang, P. P. Lee, J. C. S. Lui, and R. Perlman, “Secure overlay cloud storage with access control and assured deletion,” IEEE Trans. Dependable Secure Comput., vol. 9, no. 6, pp. 903–916, Nov. 2012.
5. A. Juels and A. Opera, “New approaches to security and availability for cloud data,” Commun. ACM, vol. 56, no. 2, pp. 64–73, 2013.
6. L. Qiu, V. N. Padmanabhan, and G. M. Voelker, “On the placement of web server replicas,” in Proc. IEEE Comput. Commun. Soc. 20th Annu. Joint Conf., 2001, vol. 3, pp. 1587–1596.
7. D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, and A. Y. Zomaya, “Energy-efficient data replication in cloud computing datacenters,” in Proc. IEEE Globecom Workshops, 2013, pp. 446– 451.
8. G. Kappes, A. Hatzieleftheriou, and S. V. Anastasiadis, “Virtualization-aware access control for multitenant filesystems,” in 30th IEEE Symposium on Mass Storage Systems and Technologies (MSST), pp. 1–6, 2014
9. Lo’ai Tawalbeh , Nour S. Darwazeh “A Secure Cloud Computing Model based on Data Classification,” 2015.
10. M. Ali, R. Dhamotharan, E. Khan, S. U. Khan, A. V. Vasilakos, K. Li, and A. Y. Zomaya, “SeDaSC: Secure data sharing in clouds,” IEEE Syst. J., DOI: 10.1109/JSYST.2014.2379646, 2015.

AUTHORS PROFILE

Siri Mallugari, currently pursuing Master of Technology in Software Engineering at VNR Vignana Jyothi Institute of Engineering and Technology (VNRVJIET) affiliated to JNTU Hyderabad.. Her interests include Cloud Computing data analytics, data mining and machine learning.

Krithi OHRI, is currently working as Assistant Professor in Department of Computer Science and Engineering at VNRVJIET, Hyderabad, India..Her Interests include Cloud Computing, Object Oriented Analysis and Design, C programming, Software Quality Assurance and Testing, Database Management, Networks area of research includes data mining, data analytics and machine learning.

Sai Poojitha Nimmagadda, currently pursuing Master of Technology in Software Engineering at VNR Vignana Jyothi Institute of Engineering and Technology (VNRVJIET) affiliated to JNTU Hyderabad. She worked as quality analyst in an MNC at Hyderabad. Her interests include data analytics, data mining and machine learning.