

An Improved Model to Increase Retrieval Time and Security by Data Fragmentation and Replication Process in Cloud



P. Peter Jose, S.P.Victor

Abstract: Cloud computing provides several features to users as well as to the organizations. Even though, there are some issues faced by the user while using the cloud. Security is a major concern that is always considered. Likewise Data replication is a significant technique to be considered for retrieval time. Replication helps to fetch the data from remote which is a high-time consuming process. To overcome the security issue along with data replication a novel approach is proposed in this paper. Dynamic fragmentation is utilized for the division of a file into fragments. Each cloud nodes has a different fragment to enhance the data security of the system. Blowfish technique is used for encrypt the files before storing in cloud that divides messages into 64 bits blocks then encrypts them separately. The result of experimental evaluation shows that this schemes increase the overall performance.

Keywords: Fragmentation, Data replication, retrieval time, data security.

I. INTRODUCTION

Nowadays it is becoming common in storing data in the Cloud for both organizations and individuals alike being its lack of high reliability for data backup Praveen Challagidada., (2017). To overcome these problems, Cloud providers stores several copies (replicas) in various data centers. Usually, the data centers separated geographically across the world. Any individual data center can experience down time for many reasons such as natural disasters, software error and attacks. The idea behind the data replication presents a numerous challenges in cloud. One of them is to maintain the cost of data to be copied to all data centers in several locations.

Data retrieval is a major issue in cloud computing Delishiya Moral and Kumar., (2016). Due to this, multiple servers is used for data storage across the cloud. Data is divided before sharing to several servers. Prior techniques are focused at increasing the data operation, reducing the processing time, aid data operation, storage, processing costs and assist the progress of data distribution and transportation Hudic et al., (2012). Sun et al., (2012) study shows that the reducing process of waiting time and fast data access is performed through replication. Nowadays, many data chunk and replication method has been proposed by researchers.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

P. Peter Jose* Research Scholar Department of Computer Science, Bharathiar University, Coimbatore.

Dr.S.P.Victor, Associate Professor, Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The remaining section is arranged as follows. Section II discusses the literature review regarding the replication technology along with security issue. Sections III presents the proposed architecture in a detailed manner. The implementation and the result is discussed in section IV. Eventually the conclusions of the research and its future directions are discussed in section V.

II. RELATED WORKS

Xie et al., (2017) analyzes the research related to data reliability in a distributed data centers. Then the authors proposed a novel data replication strategy that represents the network topology with fixed rules. Azari et al., (2018) proposed a novel replication procedure based on connectivity graph. This Popular Groups of Files Replication (PGFR) procedure enhances the local availability by identifying an association of files in every data grid and replicates the files based on the popularity.

Z. Ou et al., (2013) introduces a mechanism for Cloud Service Provider (CSP) to enhance the services of customers. The presented research found required variety in disk performance, CPU utilization, network speed, memory utilization of same VM size. By considering these factors the author proposed a mechanism that found the best performing VMs, and cancels the remaining ones. This allows the customer to use less number of VMs by achieving the desired performance.

Fei Xie et al., (2017) presents a novel replication approach to decrease the maintenance expense of data storage. Through this method the author splits the storage space into two types, and dataset into three types. This mechanism designed based on the build-time stage with several determinant levels.

S. Manjula et al., (2016) presents a new approach based on division of data that improves the data privacy. RSA is utilized for data encryption and performs the fragmentation and stored on various data centers. The Cloud Manager only has the authority of storing the file in cloud server, not able to process which increases the confidentiality of data. The data fragment in the cloud obtains the security and privacy to user data and improves the efficiency through replication techniques.

Zeng Zeng and Bharadwaj Veeravalli., (2014) presents a data replication mechanism in cloud to decrease the response time of the data access. This research proposed a metadata request balancing procedure for detecting the optimal solution and considers the number of Metadata Servers for serve the data request.

BakhtaMeroufelGhalemBelalem.,(2013)focus on improving the data availability and introduces a replication strategy. This scheme ensures the data availability in an efficient manner with few replicas and significantly reduces system overloading. The researcher also develops the placement and replacement mechanism for replicas.

III. PROPOSED METHODOLOGY

The effectiveness of data access in cloud is depending upon the performance of the retrieval process by

decreasing the response time of the data. The proposed architecture is shown in figure 1. The objective is to enhance the data access performance and data security through dynamic fragmentation process. The uploaded file is encrypted using blow fish algorithm and fragmented. The nodes contains the chunk are divided with particular distance with the help of Meta heuristic algorithm that ban an attacker of guessing the chunk location. Centrality measures are utilized in this system for node selection that enhances data retrieval time.

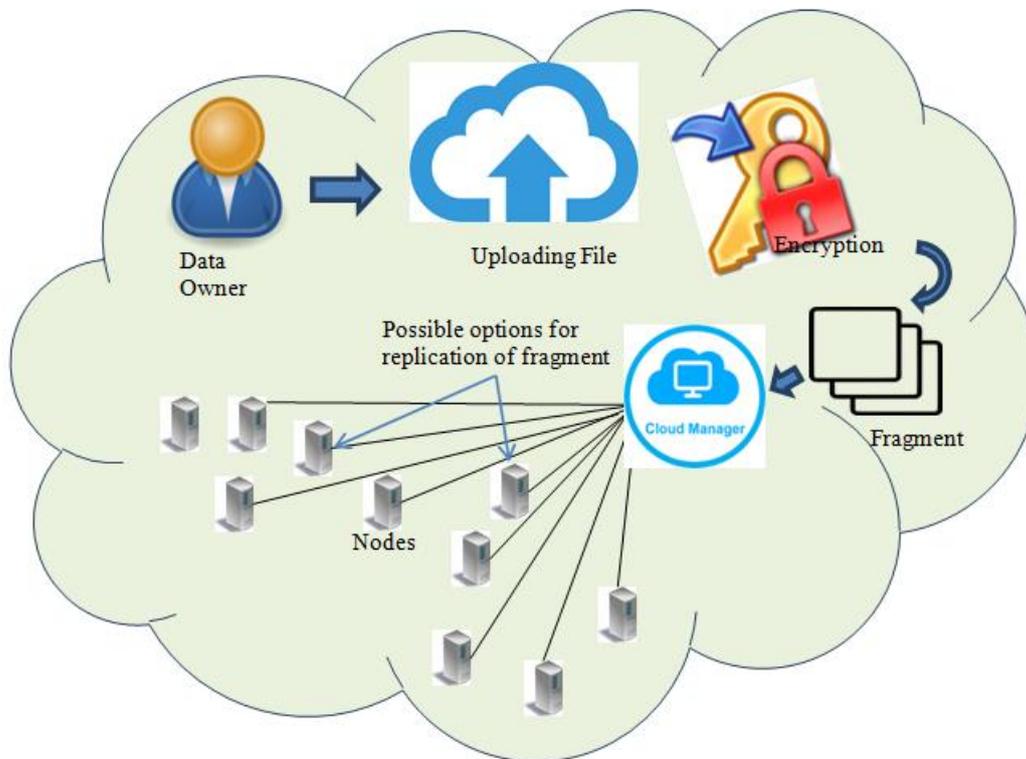


Figure 1. System Architecture

IV. DATA FRAGMENTATION

Cloud security not only relies upon the whole system but also the individual nodes. In order to enhance the security of system individual's node security is most important. When an intruder doesn't have any idea about the fragment location, the chance of identifying the data chunk on entire nodes is very low. The proposed data fragmentation is discussed in this section. Consider a cloud having N nodes and p number of a file chunk. a is consider as the number of interruption on different nodes, therefore $a > p$. The chance that s number of victim nodes have all of the p sites keeping the file chunks $F(a, p)$ is computed by:

$$F(a, p) = \frac{\binom{a}{p} \binom{N-a}{s-p}}{\binom{N}{s}}$$

Eqn-1

Distributing each data chunk in the VMs significantly enhance the data retrieval time. Hence data chunk can be copied based on some procedure which

decreases retrieval time to arrange that does not improves the previous chance. Node selection is done in two phases. The initial phase computes the centrality measures to choose the nodes for the first placement of the chunk. The second phase chooses the nodes for replication.

1. Centrality

Centrality measures are a significant process for data retrieval time enhancement in replication. The required centrality measure computation is given below.

- i. Betweenness Centrality
- ii. Closeness Centrality
- iii. Eccentricity

The initial centrality is betweenness which is measured by the count of the shortened way among other nodes. In a network consider the number of nodes is k then the formula for computing this centrality $C_{bt}(k)$ is:

$$C_{bt}(k) = \sum_{g \neq k \neq h} \frac{\delta_{gh}(k)}{\delta_{gh}}$$

Eqn-2

Where δ_{gh} denotes the sum of shortest wayamongg and h, and $\delta_{gh}(v)$ denotes the count of shortest wayamongg and h crossing through v.

Closeness Centrality is the total interval from all of the other nodes is lower than the total interval of other candidate nodes from all of the other nodes. The computation formula is

$$C_{cls(k)} = \frac{D - 1}{\sum_{g \neq k} dis(k, g)}$$

Eqn-3

The eccentricity is computed by the maximum interval to any node from a node n [24]. When a node has less eccentric then it is central in the network. The eccentricity $Ecn(k_a)$ is computed by:

$$Ecn(k_a) = \max_h dis(k_g, k_h)$$

Eqn-4

Where $dis(k_g, k_h)$ denotes the interval among the node k_g and k_h .

Meta heuristic algorithm

The proposed system utilizes the Cuckoo Search algorithm (CSA) to separate the fragment with certain distance. The pseudo code of CSA is described below:

Algorithm 1.

- Initialize the file fragment separation issue
- Randomly Initialize cloud nodes
- Compute each chunk of node
- begin the search loop
 - declare the best node
 - obtain a Cuckoo Randomly by Levy Flights
 - select a node to store the fragment
 - deserted the poor node and built novel one
- If Termination condition is arrived stop the process
- display the best node

In algorithm 1; the node denotes the result candidate, and the cuckoo bird denotes the novel result candidate. The optimum solution is provided by designing the approach of expanding the eggs of the cuckoo birds in the nature. In this approach eggs are laid by the cuckoo through Levy Flights requires the fraction (pa) as a parameter [10].

In the proposed frame work cloud manager system performs the following process after the user uploads the files in cloud:

- (a) Data fragmentation,
- (b) Initial nodes selection and
- (c) Fragments replication of the selected nodes.

Algorithm 2: pseudo code of fragment employment

```

O = {O1; O2, ..., ON}
o = {sizeof(O1); sizeof(O2), ..., sizeof(ON)}
bestnest = {opennest; closenest}
cm = {cm1, cm2, ..., cmn}
bestnest ← opennest ∀ i
cm ← cmi ∀ i
for each Ok ∈ O do
  choose Mi | Mi ← index of (max (cmi))
  if bestnestMi = opennest and mi ≥ ok
    Mi ← Ok
    mi ← mi - ok
    bestnestMi ← closenest
  Mi' ← dist(Mi, T)
  bestnestMi' ← closenest
End if
End for
    
```

In cloud the data availability and data retrieval time is increased by locating the data fragments on the central nodes. In the proposed approach the data chunk in cloud node gives the low access cost by increasing the recovery time for accessing the chunk for reunion of source document.

Fragment's replication pseudo code

```

for each Ok ∈ O do
  Choose Mi that has max (Rki + Wki)
  if bestnestMi = opennest and mi ≥ ok then
    Mi ← Ok
    mi ← mi - ok
    bestnestMi ← closenest
  Mi' ← dist(Mi, T)
  bestnestMi' ← closenest
End if
End for
    
```

Experimental Result:

The experimental setup and their performance are discussed in this section. This implementation is developed using java language in Net beans IDE and the cloud simulation is performed with CloudSim tool. Cloudsim is java jar file that provides a generalized simulation framework for cloud. The efficiency of the performance study is evaluated and their report is discussed in this section.

Response Time:

Response time of the system is computed by the amount of time taken from when a process is scheduled until the first response is produced.

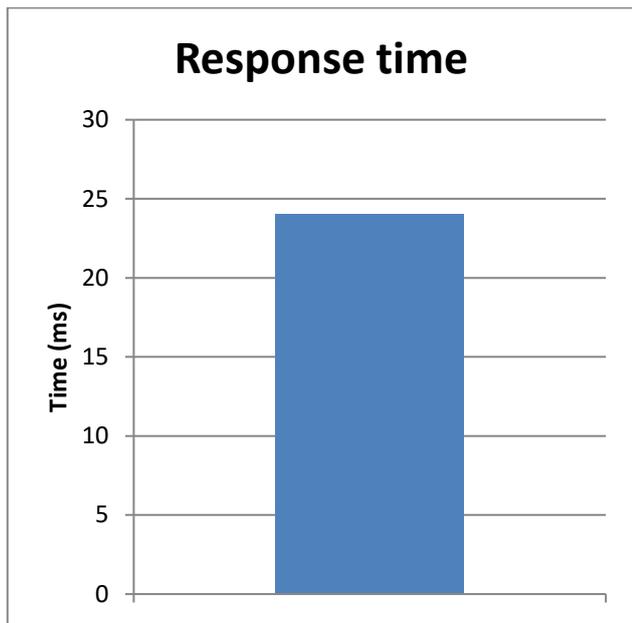


Figure 1. Graph of Response Time

The response time result is shown in figure 2. The Proposed system obtains 24ms response time which means the scheduled task got the response within 24ms.

Throughput

The throughput is a significant metrics computed by total count of job finished per second. Efficiency of this study is shown through this performance metrics. In order to compute the throughput two important parameters is required one is Process Time (PT) and Workload (W).

$$\text{Throughput} = W / PT$$

The throughput is measured and their result is given in Figure 3. The below graph clearly shows that this research obtain the highest throughput that is 4.5 kb per second.

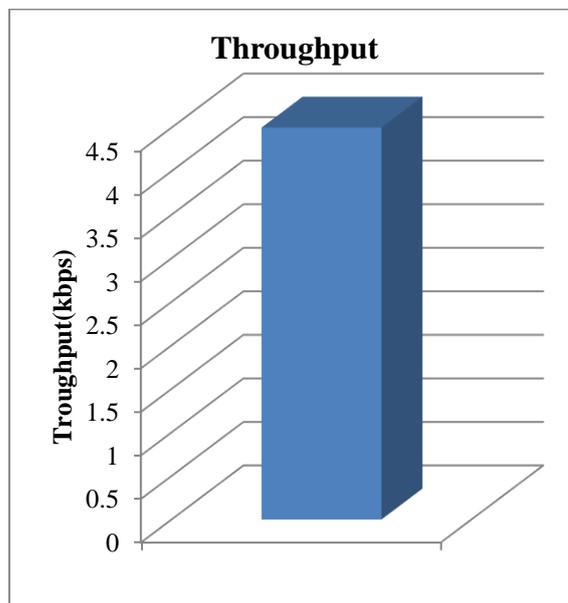


Figure 3. Throughput Report

CPU Memory

The CPU Memory capacity of task submitted in Virtual Machine is shown in figure 4. This result shows that the proposed work provides the improved result with respect to CPU memory

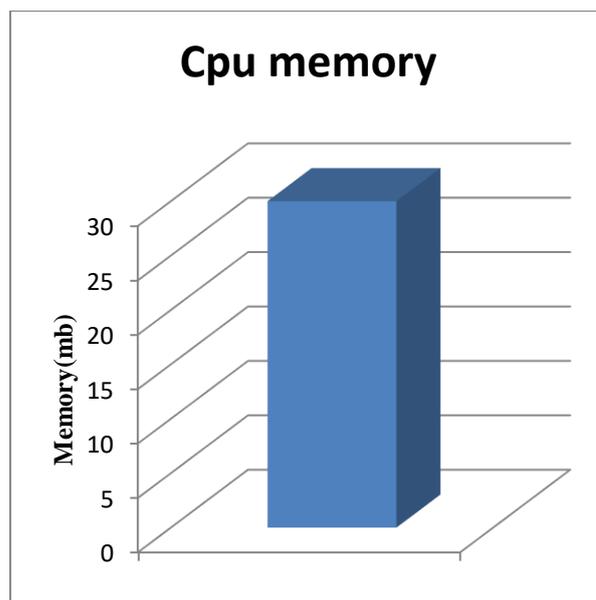


Figure 4. Throughput Report

V. CONCLUSION

Enhancing the security for replication in cloud through file fragmentation approach is the major research concern considered in this study. The framework designed in our work enhances the retrieval time and security of the user data. The prior study regarding this research issue is also discussed.

The approach focused on fragmenting the user file and storing them into the separate node and replicates those files for improving the data access. The implementation is performed with cloudsimsim tool. The performance of this research is evaluated by the performance metrics such as CPU memory, throughput, Response Time.

REFERENCES

1. Praveen Challagidada, Ambika S. Dalawai and Mahantesh N. Birje, "Efficient and Reliable Data Recovery Technique in Cloud Computing", *Internet of Things and Cloud Computing* 2017; 5(5-1): 13-18.
2. Montoya, Gabriela, HalaSkaf-Molli, Pascal Molli, and Maria-Esther Vidal. "Decomposing federated queries in presence of replicated fragments." *Journal of Web Semantics* 42 (2017): 1-18.
3. W. Delishiya Moral and B. M. Kumar, "Improve the data retrieval time and security through fragmentation and replication in the cloud," 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, 2016, pp. 539-545.
4. A. Hudic, S. Islam, P. Kieseberg and E. R. Weippl, "Data confidentiality using fragmentation in cloud computing", *Int. J. Communication Networks and Distributed Systems*, vol. 1, No. 3/4, 2012, pp. 1-10.
5. D. W. Sun, S. Gao, L. Z. Jin and X. W. Wang, "Modeling a dynamic data replication strategy to increase system availability in cloud computing environments", *Journal of Computer Science and Technology*, vol. 27, No. 2, 2012, pp. 256-272.
6. F. Xie, J. Yan, and J. Shen, "Towards cost reduction in cloud-based workflow management through data replication," pp. 94–99, 2017.
7. Azari, L., Rahmani, A. M., Daniel, H. A., Qader, N. N. (2017). A data replication algorithm for groups of files in data grids. *Journal of Parallel and Distributed Computing*, 113 (2018), 115-126.
8. Z. Ou, H. Zhuang, A. Lukyanenko, J. Nurminen, P. Hui, V. Mazalov, and A. Yla-Jaaski, "Is the same instance type created equal? exploiting heterogeneity of public clouds," *IEEE Trans. Cloud Comput.*, vol. 1, no. 2, pp. 201–214, 2013.
9. Xie, F., Yan, J. and Shen, J., 2017, August. Towards cost reduction in cloud-based workflow management through data replication. In *2017 Fifth International Conference on Advanced Cloud and Big Data (CBD)* (pp. 94-99). IEEE.
10. Manjula, S., Indra, M. and Swathiya, R., 2016, January. Division of data in cloud environment for secure data storage. In *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)* (pp. 1-5). IEEE.
11. Zeng, Z. and Veeravalli, B., 2014. Optimal metadata replications and request balancing strategy on cloud data centers. *Journal of Parallel and Distributed Computing*, 74(10), pp.2934-2940.
12. Yang, Xin-She, and Suash Deb. "Cuckoo search via Levy flights." *Nature & Biologically Inspired Computing*, 2009. NaBIC 2009. World Congress on. IEEE, 2009.

of the Department of Computer Science and Director of the Research Centre for six years. Functioned as the Dean of Science of the College for three years. Having Strong educational background with a Master degree in Computer Applications, M.C.A, Master degree in Engineering, M.E (CSE) and a Ph.D degree. Recognized as a research guide by the Manonmaniam Sundaranar University, Tirunelveli. He has published around 116 research articles in refereed International journals with good impact factor. Guiding students in the areas like Cloud computing, Data mining, Web mining, Graph mining, Networks, and Image processing.

AUTHORS PROFILE



Peter Jose P., has completed M.C.A Degree in Computer Applications from Sacred Heart College, Tirupatturin the year 2009. He is doing his Part Time Ph.D (Cloud Computing) in Bharathiyar University, Coimbatore, Tamil Nadu 641046, India. His research area is Cloud Computing. He has published research articles in refereed International journals

with good impact factor.



Dr. S.P. VICTOR, has completed his Ph.D. in Computer science in M.S. University Tirunelveli in the year of 2005. Currently working as Associate Professor of Computer Science, St. Xavier's College Palayamkottai, with over 30 years of teaching experience. Functioned as the Head

