Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 6, June 2017, pg.57 – 62

A Novel Approach Towards Data Replication

Mahesh Paragannavar[1], Prof. S. L. Deshpande[2]
[1]Department of Studies in Computer Network Engineering, VTU Belagavi, India
[2]Professor Department of Studies in Computer Network Engineering, VTU Belagavi, India
[1] maheshparagannavar@yahoo.com; [2] sld@vtu.ac.in

Abstract*— An infrastructure needs to be built, that allows users sharing and managing data expedients effectively over the internet. Data sharing can be enhanced through dynamic data replication strategies to degrade access latencies and bandwidth consumption. Replica placement is a strategy to generate and place copies of the most wanted file in specified locations in the data grid environment. To minimize the total execution time of work, storage consumption and effective usage of network usage, a novel method for replica placement is introduced here. A novel method is proposed to find the popular file present in the environment using dynamic threshold mechanism and to find the best node to replicate the popular file using replication replacement strategy. The proposed method illustrate how many replicas are created, access latency which is elaborated in result section.

Keywords*— Data Grid, Dynamic replication, Dynamic Threshold, Replication replacement strategy etc.

I. INTRODUCTION

Information plays a devoted role in today's era of internet. Everyday pool of configured computing expedients are delivered and/or received on demand to the outside customers over the net. In order to make data always available whenever request is made by user or any deterioration occurs for this to happen we need to form a replica of that data. Replication is habituated to enhance the availability by permitting access to the information even when the few replicas are not found. User access nearby replicas with this it enhances the throughput.

In presently developed cloud systems static replication is employed to create copies of data to datacenter. In static location cannot be changed and there is no automatic replica generation and replacement processes are carried out. All the users will access to the data on demand enabling flexibility across different locations and also optimization with minimal resource consumption.

Dynamic replication is a technique where decisions are made depending upon the present access motifs and resource availability. There are two phases the first one deals with replication location and creation using catalogue and index. Second one depicts optimization among copies of replications. Data sharing, gathering and management and also selection plays tectonic role in grid technology also. This is considered as an important method for effectively utilizing the distributed resources. These resources are effectively distributed among the different locations geographically. If centralized location is considered to manage data it takes large access time and more execution time. To minimize these two problems distributed environment is considered here. The perfect solution to this is to create replicas of data to be accessed. It is considered as an optimistic approach that reduces data access time and storage expedients effectively.

Data replication is a method of creating and placing the copies of data entries. The maximum data entries made in the grid environment are considered and the popular data entry among them is chosen as the data to replicate. Once the popular data is favoured that needs to be placed or located somewhere. For that the best place among the network environment needs to be selected that is done by exploiting replica placement strategy.

In Data grid environment, if one site is failed, then with the replicas of data the sharing accessing of data can be effectively managed. Hence we can say this proposed scheme can able to reproduce the advantages as fault tolerance, increased availability, scalability with lower bandwidth and lesser access latencies.

In our work we focus on dynamic replication strategy which takes the above mentioned parameters into considerations. There are many research work and interpretation of replication algorithms are carried out by many authors and are illustrated in the following sections.

II. LITERATURE SURVEY

The storage capacity of an individual grid site is limited hence replication utilization has to be carried out wisely. For this it is very important to design a strategy that handles replica replacement. As Data Grid is considered as an important and/or useful technology for huge number of data processes produced by data computing applications, a solution to enhance the availability and access time in a Data Grid is replicating the data. A novel replication method proposed is based on Latest Access Largest Weight dynamic data replication strategy [1].

A Data Grid assembles a collection of distributed computational storage expedients that allows users to exchange data and other resources. Data replication is much discussed by Data Grid researchers in recent years creates more number of copies of file and places them in various positions to limit file access periods. A dynamic data replication strategy called Modified Dynamic Hierarchical Replication (MDHR) strategy is developed that replaces replicas on the basis of last time the replicas requested, number of accesses and replica size. The best replica place among many replicas considering response time as parameter is determined with considering data transfer period, the storage access delay, the replica requests waiting in the queue, distance between nodes and CPU capability of processing [2].

The eminent replication tactic is necessary because of limited storage. A novel replacement tactic is also proposed that deletes files when the space is not enough for the new replica storage and is named as New Dynamic Hierarchical Replication (NDHR). NDHR enhances replica selection by selecting best location for replica and also in efficient way data replication method is proposed to improve overall performance [3]. Amount of data transferred among nodes is minimized by passing jobs to the nodes that has maximum files requested by exploiting scheduling technique and minimized access delay by using data replication strategies. All the techniques of scheduling and replication issues are addressed in [4] by considering attributes like network bandwidth, dynamic behaviour of user, latency and cost of replication.

Irrespective of storage capacity, it is necessary to prepare a strategy for replication replacement. Prediction-Base Dynamic Replication (PBDR) method deletes files with parameters like number of requests, replica size, availability and last requested replica. It also reduces access delay by electing the best replica. PBDR enhances data access time. This algorithm is simulated[5].

III. METHODOLOGY

The proposed methodology includes two phases namely Selection of Replica and Replacement of inconsistent Replica.

A. Selection of Replica:

The request for a file is updated in the history table. If the frequency of requested file is greater than the average access frequency of all the files, then it is decided to replicate the required file and it is called as popular file.

B. Replacement of inconsistent Replica:

When a requested replica is not available in the local storage, replication should take place. According to the temporal and geographical locality the replica is placed in the best sites or nodes. To select the best site or node characterizes the number of appropriate sites for replication by a generating two lists. List L1 will be based on the communication cost of each node present in the network and ranking them in descending order helps us to select an appropriate node for replication. List L2 will be based on the storage space available in all the nodes to replicate the popular file. The decision for replication is based on the two lists cascading. An appropriate decision is taken to recover the issue of storage. Once the replication of popular file is done then both the lists have to be updated simultaneously.

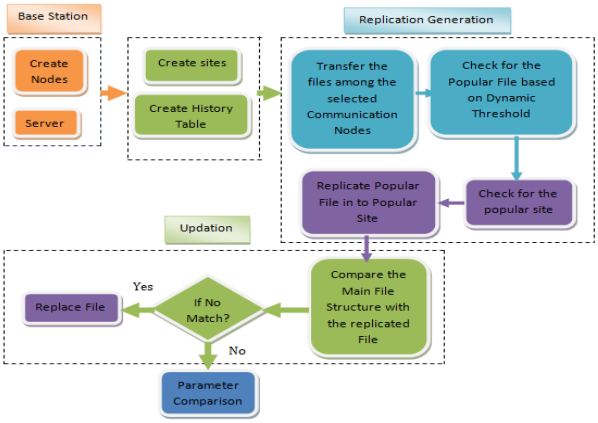

Figure 1: Architecture Block Diagram

IV. IMPLEMENTATION

As discussed in the methodology part with the interpretation is carried out for the proposed architecture with creating network and considering server with different sites and deploying few nodes in the respective sites. Then executing few communications history of the executions is created, according to history the most popular file is computed through the dynamic threshold equation.

Once the popular file is obtained the most popular place where the replica of popular file needs to be copied is a question that is computed by considering the maximum number of users from their respective sites requested for the popular file. That site is called as best or popular site or node and replica is placed in that site or node. Similarly different cases are considered and optimal replicas are found and replaced in the presented paper. The flow chart of the best node elected is given stepwise is Figure 2 below. The corresponding results of the interpretation are outsourced in result section.

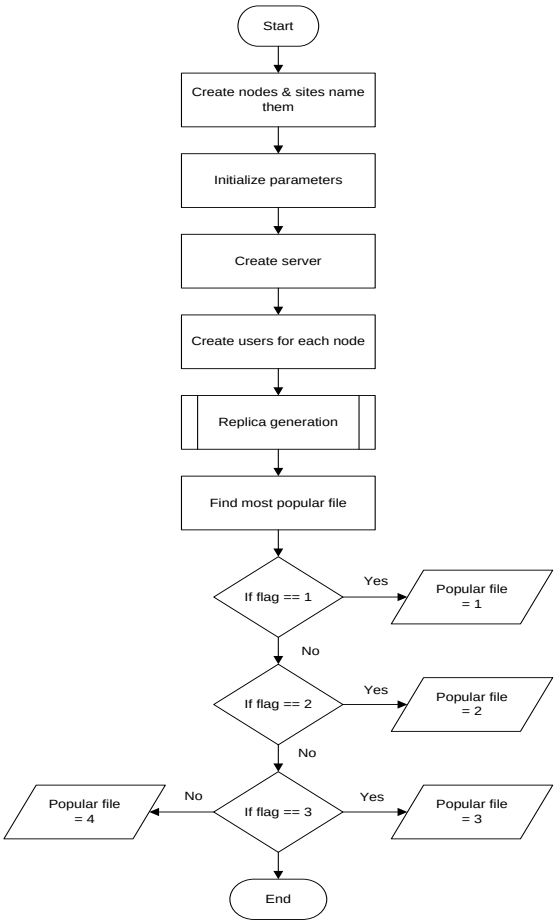

Figure 2: Flow Chart of Network Creation

V. EXPERIMENTAL RESULTS

A network is created with server, nodes as users, and four sites. Based on how the communication with the server is happening by different sites and their associated nodes the user history is created for finding a popular file.

The figure 3 depicts traffic at server side before and after replication. The time taken for accessing the files from the server by different sites at different communications before replication is obviously more compared to after replication. Hence time taken for access is minimized and overhead at server is minimized to around 67%.

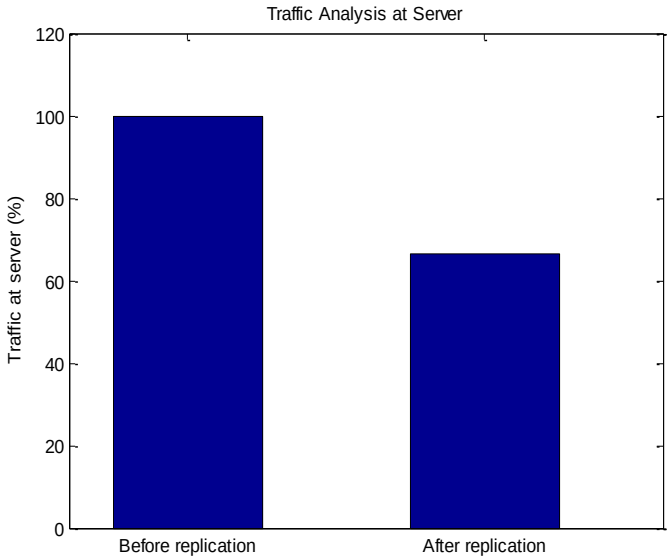

Figure 3 : Traffic at Server side before and after replication

The Figure 4 and 5 depicts the comparison of the proposed method with the already existing techniques of replication referenced as [11]. Figure 4 depicts the number of replications formed compared to Replica Placement with Load Balancing (RPLB) algorithm. It is clear from the figure that our proposed method has formed less replications as compared to RPLB. The comparison of access latency of proposed method and existing methods are given in Figure 5. The proposed method has access latencies of 344.87sec and 613.68sec for 100 and 192 file accesses respectively. RPLB holds 781sec for 100 files accesses and 841 for 192 file accesses.

Table 1: List of No. of Replicas with respect to No. of jobs of Proposed system

No. of Jobs	No. of Replicas
50	**4**
100	**9**
150	**13**
200	**17**

The Table 1 shown above indicates that as the number of jobs increases the number of replications increases for our proposed technique. In case of RPLB method at 192 or 200 jobs the number of replicas will be increased by only one.

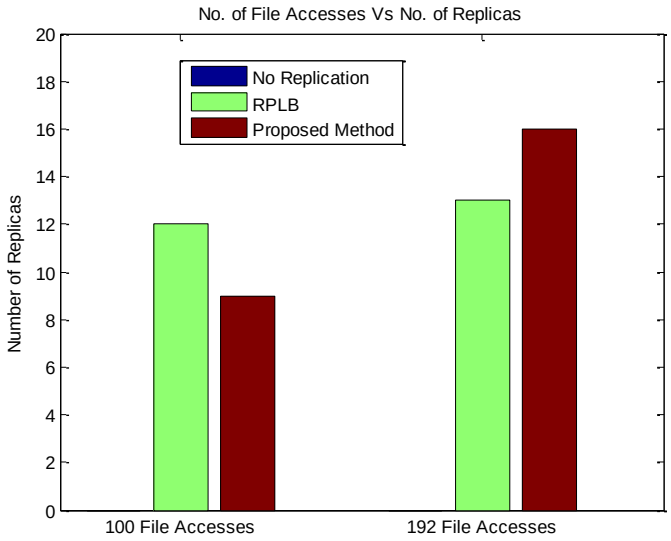

Figure 4: Plot of No. of File Accesses Vs No. of Replicas

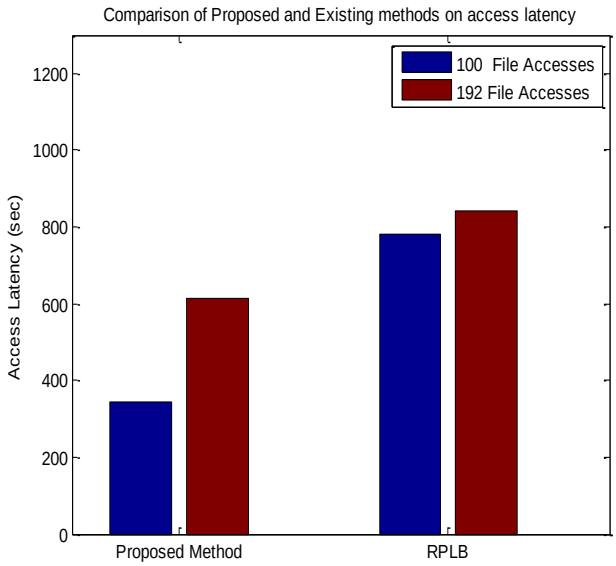

Figure 5: Comparison of Access latency with the proposed and existing methods

VI. CONCLUSION

As represented in methodology part the dynamic optimal replication strategy is one of importance factor to save memory space, avoid traffic at server where all the users request for their respective files directly, and most importantly access latency, it has to be minimized by finding an optimal replica using the dynamic threshold and is capable of saving memory space, access latencies. From Figures 4 and 5 it is cleared that our nominated method is better as compared to the already existing tactics. The access latencies of proposed is 345 and 613 seconds for 100 and 192 file accesses which are less as compared to RPLB and other algorithms.

REFERENCES

[1] N. Mansouri, A. Asadi, "Weighted Data Replication Strategy for Data Grid Considering Economic Approach," International Journal of Computer, Electrical, Automation, Control and Information Engineering, Volume 8, Issue 8, 2014.

[2] N. Mansouri and Gh. Dastghaibyfard, "Improving Data Grids Performance by using Modified Dynamic Hierarchical Replication Strategy," International Journal of Electrical & Electronics Engineering, 2014.

[3] Najme Mansouri, Ehsan Mansouri, Amir Yazdinezhad, Sahar Abedpour, "Proposing and Evaluating Dynamic Data Replication Strategy in Data Grid Environment," Majlesi Journal of Multimedia Processing, Volume2, Issue 3, 2013.

[4] Arul Xavier V. M, Lakshmi C. S, "Analysis on Performance with Combination of Scheduling and Replication in Data Grids," Indian Journal of Computer Science and Engineering (IJCSE), Volume 4, Issue 3, 2014.

[5] Najme Mansouri, "A Prediction-Based Replication Algorithm for Improving Data Availability in Grid Environment," Volume 6, Issue 1, 2014.

[6] Nagamani H Shahapure, P Jayarekha, "Replication: A Technique for Scalability in Cloud Computing," International Journal of Computer Applications, Volume 122, Issue 5, 2015.

[7] K.Sabarigirivason, R. Giridharan, "A Novel On Fast Parallel File Transfer Using Replication," International Journal of Computer Science and Mobile Computing, IJCSMC, Volume 3, Issue 2, 2014.

[8] Sanjay Kumar Yadav, Gurmit Singh, Divakar Singh Yadav, "Mathematical Framework for A Novel Database Replication Algorithm," International Journal of Modern Education and Computer Science, Volume 5, Issue 9, 2013.

[9] Rajshri S Patil, L K Gautam, "Full & Partial Replication in Data Grid Using Gridsim," International Journal of Advance Research in Computer Science & Management Studies, Volume 3, Issue 4, 2015.

[10] Da-Wei Sun, Gui-Ran Chang, Shang Gao, Li-Zhong Jin, Xing-Wei Wang, "Modeling a Dynamic Data Replication Strategy to Increase System Availability in Cloud Computing Environments," Journal Of Computer Science And Technology, Volume 27, Issue 2, 2012.

[11] Kingsy Rajaretnam, Manimegalai Rajkumar, Ranjith Venkatesan, " RPLB: A Replica Placement Algorithm in Data Grid with Load Balancing",The International Arab Journal of Information Technology,2014.

