Journal of Computer Science 10 (12): 2548-2552, 2014
ISSN: 1549-3636

doi:10.3844/jcssp.2014.2548.2552 Published Online 10 (12) 2014 (http://www.thescipub.com/jcs.toc)

CIRCULAR NEIGHBOR REPLICATION

Ahmad Shukri Mohd Noor, Nur Farhah Mat Zian and Yazid Md Saman

School of Informatics and Applied Mathematics, University Malaysia Terengganu, Kuala Terengganu, Malaysia,

Received 2014-08-26; Revised 2014-10-29; Accepted 2014-11-10

ABSTRACT

Distributed system rely on replication techniques to tolerate data failure and site disconnection thus ensur-ing flexibility of the system so as to preserve its dependability. The idea of replication is robust however practical implementation of the replication techniques is often rigid that would bring down system's de-pendability and performance. This study intended to evaluate existing techniques and then develop a new technique which later will be compared with the existing with the goal of to achieve better fault tolerance, dependability and performance in distributed systems. The new technique is constructed based on circular neighbor relationship and quorum-based protocol. The consistency and integrity of the replicated data that involved write and read operations on the replicas is ensured using Replica maintenance protocol. This techniques focused on synchronous solution as for its quorum execution or commitment protocol showed higher reliability and convenience to avoid conflicts compared to asynchronous solution.

Keywords: Availability, Replication, Neighbor Replication, Quorum Voting, Replica Management Protocol

1. INTRODUCTION

Replication method is a powerful tool to ensure availability and reliability in fault tolerant computing. Site replication gives very high availability as it masks environmental failures, hardware failures, operator error and even some software faults (Helal *et al*., 1996). Although the idea of replication is robust, it is often a difficult challenge to determine practical implementation of a replication technique since each replication technique involves different levels of complexity (Mohd Noor *et al*., 2014). While a simpler replication technique would be preferred, it comes with weaknesses that would bring down the system's availability, reliability and performance. Thus, providing a replication mechanism and control protocol that can promise data integrity, easy access, reliability and availability is highly desired to make a good distributed environment especially if it is large in size or heavy in transaction number. An excellent replication technique is also first step to develop a reliable fault tolerance mechanism.

The rest of the paper is organized as follows: Section 2 discusses several replication techniques and presents the research background of the new circular neighbor replication. The management protocols are also discussed. Section 3 describes the circular neighbor replication technique with the possible maintenance protocol. Section 4 presents a rough evaluation on the new technique. Conclusion and future works are given in sections 5.

2. REPLICATION TECHNIQUES

There are synchronous and asynchronous solutions. An example for the latter is Lotus Notes which works reasonably well for single object update but fails when it comes to multiple objects in a single update (Helal *et al*., 1996). Additional approach is needed to overcome this problem for asynchronous replication such as vector timestamps and log records. In this study the synchronous solution is preferred for its comparable convenience when it comes to avoiding and resolving conflicts in replica access and updating and the higher availability that results. The

Corresponding Author: Ahmad Shukri Mohd Noor, School of Informatics and Applied Mathematics, University Malaysia Terengganu, Kuala Terengganu, Malaysia

synchronous solution can be based on quorum execution and/or coupled with commitment protocol to reach one copy serializability and higher degree of consistency.

Replication has two schemes; full replication (all-data-to-all-sites) or partial replication (all-data-to-some-sites). Full replication causes high update propagation and larger storage capacity which causes higher overhead (Mamat *et al*., 2006). The partial replication minimizes storage capacity while potentially compromising availability due to less number of backup copies. However, this issue is not significant if individual sites and the network are consistent, as well as having a superior failure detection and recovery action plan.

A straightforward protocol in full replication is Read-Once-Write-All (ROWA). In ROWA, a logical read operation on a replicated data item is converted to one physical read operation on any one of its copies, but a logical write operation is translated to physical writes on all of the copies. The ROWA protocol is good for environments where the data is mostly read-only because it provides read operation with a high degree of availability at low communication overhead. However, the write operation has very high overhead as all replicas must be updated simultaneously. Also, it has very low availability as an update cannot be performed in the presence of a single replica failure or network partitions (Helal *et al*., 1996). If one site is not accessible, the processing of an object is noted in the partial commit state and resolved after some time delay. The problem with this approach is that it increases the response time, which is one of the major performance parameters in replicated systems and therefore decreases the performance of the system.

The concept of neighbor replication is all-data-to-some-sites (partial replication) where only neighbors are considered to have the replicated data. This assignment provides a higher availability of executing write operations in replicated database due to the minimum number of quorum size required in Neighbor Replica Triangular Grid (NRTG) (Mamat *et al*., 2006). In NRTG, the organization of sites is as in **Fig. 1.** The most number of replicas a site can have is five which is site 5.

Another neighbor replication is Neighbor Replication Distributed Technique (NRDT) proposed in (Mamat *et al*., 2004). In this technique nodes are logically arranged in a n x n grid for N number of sites so N. Each site holds a primary data and copies of its adjacent neighbors' data. At most, one site holds 5 data copies (site d5 in **Fig. 2**). This technique incur smaller overhead cost as smaller number of nodes can hold more replicas.

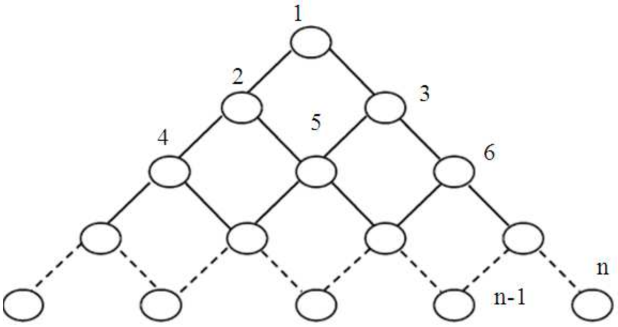

Fig. 1. The Organization of Sites in triangular-grid Replication

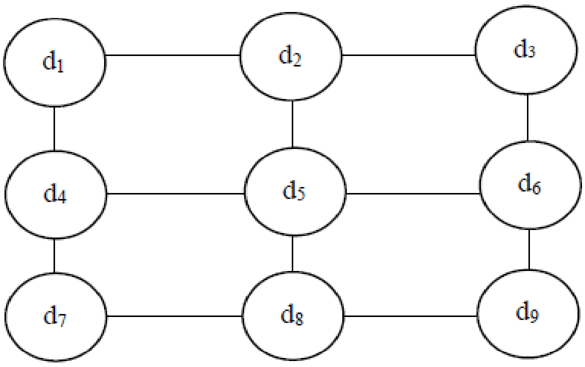

Fig. 2. The organization of sites in NRDT

3. MODEL AND METHOD

This section will describe the newly proposed replication technique, namely Circular Neighbor Replication Technique (CNT). The neighboring relationship is similar to that of a circle where a number of sites are related to each other back to back. This way each site has two neighbors, the one preceding it and the one following. This relationship property is true for sites of three or more, whereas two sites will simply be neighbors to each other. The proof of the property is as following:

The set of sites is defined as where N is the total number of sites if there are more than three sites. As easily illustrated in a circle, in the base case of three sites, each member site has two neighbors i.e., has and as neighbors, has and as neighbors and has and as neighbors. For instance, we want to add one more site to the site. This can be done by inserting the additional site, to follow. This breaks up the neighbor relationship between and consequently, this neighbor relationship is replaced with. Thus now, has and as neighbors and has and as neighbors and immediately has and as neighbors. By iteration, adding one more site will have the same process and consequence. Thus, it is proven for any number of sites of three or more, each site has two neighbors preceding and following it, in the circular neighborhood relationship.

By the previous definition, at most each site has a primary data and data copies of its two neighbors. Thus, each site holds three copies of unique data (one being its own primary data) and each individual data has three copies anywhere on the system. Compared to full replication, this new technique provides minimization on storage capacity without compromising availability. In terms of storage capacity, this technique minimizes it further compared to other techniques such as Triangular-Grid and NRDT. This is because all sites have 2 neighbors whereas in Triangular-Grid and NRDT, sites either have 2, 3, 4 or 5 neighbors.

Quorum is defined as the number of replicas to be in agreement for an operation to take place on them. It is important to mind the two operations read and write so conflict will not happen if any operations happen concurrently except read-read operations. Synchronicity is very important. These are two rules to ensure this one-copy serializability through quorum voting; 1) R + W > v and 2) W > v/2 being the weight of all votes. This avoids 1) concurrent writes and 2) read and write happening on the same data at the same time. This will ensure that read results always reflect the result of the most recent write (because the read quorum will include at least one replica that was involved in the most recent write) (Attiya *et al*., 1995). In optimistic voting, updates can be done concurrently in different partitions. It allows potentially conflicting updates in separate replicas. Majority is defined as the replicas whose votes will count in an operation involving a particular partition. In Dynamic Linear Voting, the total number of votes is allowed to evolve over time to correspond to configuration changes.

The management protocol of the CNT can be similar to the one proposed for Triangular-Grid where a commitment protocol is used. Except in CNT it will require different definition of majority since there can be different partitions in the circular neighborhood **Fig. 3.** The quorum size for the number of replicas involved in each transaction is smaller because it is partial replication i.e., not all sites contain same data. It is described as the following.

The quorum for an operation is defined as a set of sites/replicas whose number is sufficient to execute that operation. In other words, it is modeled as a set of replicas: C = {C1, C2,…, Cn}, where i = 1,2,…,n, are called the sequence numbers of these replicas. Each replica, Ci manage a set of data.

In CNT, there are only 2 neighbors to each site and only 3 sites contain the same copy of a data item. For one-copy serializability in CNT, for read operation, the quorum will be defined as R = 2 and for write operation, W = 3.

To prove the correctness of the quorum rules, we need to show that the quorum have non-empty intersection. Read quorums are any two adjacent sites, while write quorums are the sites of the read quorums plus one site that either precedes or follows, depending on which site has the primary data. It is obvious that, the corresponding set of write quorums intersect with read quorum set from this description. It is also proven that write-write operations and read-write operations have non-empty intersection. Thus the CNT protocol is one-copy serializable.

In each site, there are Coordinating Algorithm for the primary replica and Cooperating Algorithm for the non-primary replica. The transaction manager manages request from clients and locate the primary replica. After receiving a procedure call from the transaction manager, the primary replica perform the coordinating algorithm. 2PC protocol is used to ensure consistency. In the first phase, the primary replica asks to form a quorum for the operation (asking all related replicas to give vote). If quorum is formed, it will return 1 to the transaction manager else 0 and abort operation, or -1 in the case of partial commitment from some of the replicas. In the second phase, the transaction manager will ask to commit. Primary replica will ask all related replicas to commit to the operation i.e., lock the data. If there was partial commitment (PC), the primary replica records who gives it and will use this record in conflict resolution. The co-operating algorithm on the side of non-primary replicas will act accordingly through this whole process. During operation execution, the non-primary replicas perform the operation and unlock the data when it is done.

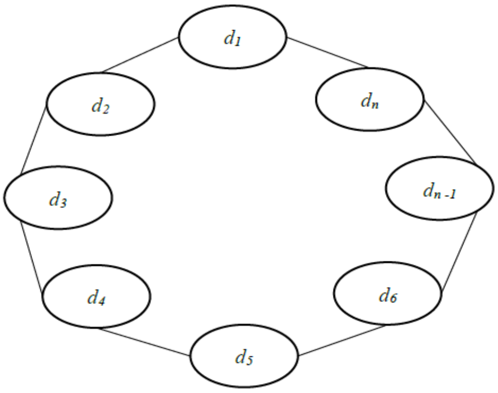

Fig. 3. Site organization in CNT of size n (any integer)

4. EVALUATION

In this section, an analysis of the availability of the Replication techniques will be presented. Availability refers to the probability a system is completely working over a period of operating time. In other words, availability is the measure of how often or how long a service or a system component is available for use (Parziale *et al*., 2009; EPSMA, 2005) Equation 1:

$$Availability = \frac{Operational}{Operational + Non - Operational} \quad (1)$$

Let be the number of nodes that are operating correctly at time t, be the number of nodes that have failed at time t and N be the number of nodes that are in operation at time t as in (EPSMA, 2005) Equation 2:

$$A(t) = \frac{N_o(t)}{N} = \frac{N_o(t)}{N_o(t) + N_j(t)} \quad (2)$$

The availability in series can be expressed as in (Mohd Noor *et al*., 2012) Equation 3:

$$A = A_V \times A_Z \quad (3)$$

And the availability in parallel can be expressed as Equation 4:

$$A = 1(1 - A_V)(1 - A_Z) \quad (4)$$

If however there is a mixed environment between parallel and serial the availability a can be defined as Equation 5:

$$A = (1 - (1 - A_W)(1 - A_X)) \times (1 - (1 - A_V)(1 - A_Z)) \quad (5)$$

When a system is comprised of two redundant components, then the availability of the system can be calculated by using parallel formula as expressed (5).

In terms of system availability score, NRDT is the most excellent followed by CNT but only outperforms CNT by 0.026%. This result shows that CNT is almost as good as NRDT in achieving high availability. Looking at individual components, 3 out of 9 components have better availability in CNT model than in NRDT model **Table 1**.

Table 1. The nine components of interdependent servers and its availabilities

Component	Availability
Web	0.9500
Application	0.9550
Database	0.9500
DNS	0.9700
Firewall	0.9600
Switch	0.9700
Data Center	0.9500
Applications2	0.9500
Manager	0.9900
Total availability	0.6956

Although this analysis shows that NRDT is slightly more superior, it is to note that CNT can achieve high availability as good as NRDT model at a much lower update propagation. This is because, CNT uses 3 copies of data throughout the entire system, while NRDT uses 2, 4 and 5 copies of data depending on the grid location of the component.

5. CONCLUSION AND FUTURE WORKS

Replication technique primarily concentrates on the two fault tolerance manners ensuring flexibility of the distributed system so as to preserve its dependability. There are many existing replication approach with different level of complexity and performance but gives better availability at the expense of bigger storage or higher overhead cost. Replica management protocol is important in order to sustain its data consistency and integrity. This new technique utilize synchronous replication scheme that will involve two protocols; quorum consensus and commitment protocol. A better design in quorum will lead to better fault tolerance and better availability in read and write operation plus reducing the cost while the use of commitment protocol can provide higher degree of consistency. The new technique will exploit the concept of neighbor replica that helps to minimize storage capacity in order to reduce overhead cost and smaller quorums will expedite in reducing both cost, access time and availability of the read and write operations. Therefore it is possible for the new technique to utilize circular neighbor relationship amongst sites. It is essential to test the score of system availability that this technique can provide so that this technique can be evaluated and compared with other techniques.

6. ADDITIONAL INFORMATION

6.1. Funding Information

This project is supported by Ministry of Science, Technology and Innovation (MOSTI) ScienceFund Grant, titled Efficient Autonomous Recovery Algorithm for Parallel and Distributed. Systems Project code 01-01-12-SF0154

6.2. Author's Contributions

All authors equally contributed in this work.

6.3. Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

7. REFERENCES

Attiya, H., A. Bar-Noy and D. Dolev, 1995. Sharing memory robustly in message-passing systems. J. ACM (JACM), 42: 124-142. DOI: 10.1145/200836.200869

EPSMA, 2005. Guidelines to Understanding Reliability Prediction. 24th Edn., Wellingborough, Northants, UK, pp: 29.

Helal, A., A. Abdelsalam, B. Heddaya and B. Bhargava, 1996. Replication Techniques in Distributed Systems (Advances in Database Systems). 1st Edn., Springer, ISBN-10: 0792398009, pp: 156.

Mamat, A., M. Deris, J. Abawajy and S. Ismail, 2006. Managing data using neighbor Replication on triangular-grid structure. Comput. Sci. ICCS, 3994: 1071-1077. DOI: 10.1007/11758549_142

Mamat, R., M. Deris. and M. Jalil, 2004. Neighbor replica distribution technique for cluster server systems. Malaysian J. Comput. Sci., 17: 11-20.

Mohd Noor, A.S., M. Deris and M.Y. Saman, 2014. Co-existance neighbourhood model for optimizing cloud Infrastructure as a Service (IaaS) within interdependent environment. Int. J. Machine Learning Comput., 4: 85.

Mohd Noor, A.S., T. Herawan and M. Deris, 2012. Neighbor-replica distribution technique model for availability prediction in distributed interdependent environment. Int. J. Cloud Applic. Comput., 2: 98-109. DOI: 10.4018/ijcac.2012070105

Parziale, L., A. Dias, L.T. Filho, D. Smith and J.V. Stee *et al*., 2009. Achieving High Availability on Linux for System z with Linux-HA. 1st Edn., IBM Redbooks, ISBN-10: 0738432598, pp: 310.