2011 2nd International Conference on Networking and Information Technology
IPCSIT vol.17(2011) © (2011) IACSIT Press, Singapore

Balanced QoS Replica Selection Strategy to Enhance Data Grid

Ayman Jaradat, Anang Hudaya Muhamad Amin and Mohamed Nordin Zakaria
Faculty of Science & Information Technology, Department of Computer & Information Sciences, Universiti Teknologi Petronas
Bandar Seri Iskandar, Malaysia
ayman418@yahoo.com,ananghudaya@petronas.com.my,nordinzakaria@petronas.com.my

Abstract. Data grids provide groundwork for huge, data-intensive applications that produce enormous data. Data replication increases the availability and reliability of data by providing identical copies of the replicated files distributed at grid sites across the world. Typically these data grid sites vary in their capabilities and their abilities to provide different levels of QoS. Replica selection in data grids remains a significant challenge and requires more attention because the current algorithms do not offer balanced QoS levels and the mechanism of rating QoS parameters. In this paper, we proposed a new replica selection strategy Known as Balanced QoS Replica Selection Strategy (BQSS), which based on a sound mathematical model and supporting metric to measure balanced QoS time, availability and security (BTAS). Computer simulation performance results demonstrate that the BQSS performs best when compared to the competitor's D-system.

Keywords: data availability; grid applications; replica selection algorithm;QoS parameters; security

1. Introduction

Scientific advances have resulted in high-intensity computer applications, which frequently require accessing, storing, transferring, analyzing, and sharing huge amounts of distributed data [1], sometimes combined at other replica sites for high-performance computer processing on the grid. The creation of data grids was prompted by such data-intensive applications requiring management and sharing of distributed data across many organizations in different locations via data replication, wherein identical mirrored replicas of the same data are produced and stored at different locations. An essential function of data replication is efficient replica selection. The selection of an appropriate optimal replica location is the principal determining factor of overall system performance. Each grid site has its own capabilities and characteristics; consequently, selecting one specific site with the best Quality of Service (QoS) offering from among numerous sites that have the required replica is an important, complicated and challenging decision. Replica selection algorithms, such as greedy [2,3,4], random [4,5], partitioned [6,7,8], and weighted algorithms [8,9,10], were mainly used in the replica selection function and have exclusively focused on response time as a criterion for the selection process.

Recent works [11,12,13,14,15,16] addressed the notions of utilizing security to select resources in grid environment. They defined security in different ways namely trust, self-protection, reputation and reliability. A. Farag et al [12] defined the reputation of an entity as the expectation of its behavior based on other entity's observations or information concerning the entity's previous behavior within a specific context at a given time. While S. Thamarai et al [13] introduced two new terms, which are 'affordability' and 'success rate'. They defined affordability as the ratio between the number of times the resource was available to the grid and the number of attempts made to access the resource. Their definition of success rate is the number of successful executions of a job by a computational resource against the total number of jobs submitted to the resource, which indirectly reveals the expertise of the resource. In their system, they integrated the previous knowledge of the resources apart from current performance. Their trust model aggregates affordability, success rate and bandwidth to evaluate the resource. Conversely, trust factor (TF) as proposed by V. Vijayakumar et al [11] consists of self-protection capability (SPC) and reputation weight. They defined the self-protection capability of a site as its ability to detect intrusions, viruses, unauthorized access as well as offering secured file storage and job completing capabilities. Furthermore, they defined reputation as a mechanism that offers a way to build trust through social control by using community based feedback on the previous experiences of entities. T. Naseera [16] has related trust to QoS as being a representation of an estimate concerning how likely a resource fulfills its QoS commitments. They defined reliability as the availability of the resource over a period of time. Their second performance measurement is access-cost. Access-cost of the replica is measured in terms of the time expended for the requesting resource to obtain a unit of data from the replica location. This measurement will explicate the proximity of the resource to the required data. According to G. Kavitha et al [14], most of the existing trust models are designed only to protect the resource provider nodes. Security of the grid users is also equally important but not addressed effectively. Therefore, he proposed a trust model to calculate the quantitative value of execution trust. Execution trust is the belief that a resource provider node will faithfully execute a user code and complete the job request. B. Zhang et al [15] differentiates between trust and reputation concepts for individual descriptions of credibility. In this view, trust is a subjective belief that demonstrates a creditable relationship between two or more individuals. Reputation presents the whole belief from all the members who have the qualifications to evaluate.

A. Jaradat, et al., [17] proposed D-system by which three QoS parameters are considered. The parameters are time, security and availability. The system presents a solution in which all sites holding the required replica are evaluated by consolidating the three aforementioned QoS parameters (i.e. time, security and availability) in one value. Time is the estimated time to transfer the replica from source to destination. Availability is a proportion of the declared time by the service provider of allowed services to the estimated time to transfer the replica from that site. Availability is a part of the site's policies since each site has a declared timeframe to offer services to the outsiders, sometimes during evenings, night time or weekends. Security or trust is a composition of the site's capabilities for self-protection, site reputation and site reliability. In D-system, time, availability and security are evaluated before selecting a replica. Each parameter is rated according to the site capabilities by assigning a value between 0% and 100%. Subsequently, they utilized an ideal model imaginary site with the rate of 100% for each QoS parameter as a benchmark to evaluate each prospective candidate site. Their approach utilizes Euclidean distance to measure the distances between sites and the model site. The QoS metric was the Euclidean distance between the model site and the prospective candidate site. The metric utilized is a composition of time, security and availability, which they entitled (TAS). The shortest distance or the lowest value of TAS represents the best site.

In this paper, an extension to D-System in overcoming related issues of time, security, and availability will be proposed. This extension intends to answer his two important questions:

How to determine the best replica location from among many replicas distributed across the grid sites in a high- quality, timely, and consistent balanced rates of QoS parameters.

How to estimate the values of the QoS parameters.

This paper is organized as follows: Section 2 is the related work, section 3 is the strategy's design to highlight and clarify our approach. Section 4 is a case study. Section 5 results and discussion. Section 6 presents the paper's conclusion.

2. Related Work

This research addresses the replica selection problem as a critical decision by focusing on the efficiency and the satisfaction of grid-users by delivering the required replicas with balanced level of security, time expenditure no extreme values for security over time or availability or the vice versa.

This research extends the past efforts D-system [17], which has some limitations published by its authors that the best explained through the illustrations in Figure 1, which for simplicity uses only two parameters instead of three. In Fig. 1, B represents a site that holds the values 50% for time and 50% for security. D-system gives the same rate for sites A, B, C and D because they have the same distance from the model value. Therefore, it will select any of them randomly. Consequently, the important limitation of the D-system is that any sites' representative that falls in the diameter of the quarter circle will be chosen randomly for the smallest circle near the model value. However, the fact site B in Fig.1, should be selected because the rates QoS parameters are equal which means a balanced selection best from all QoS parameters points of view.

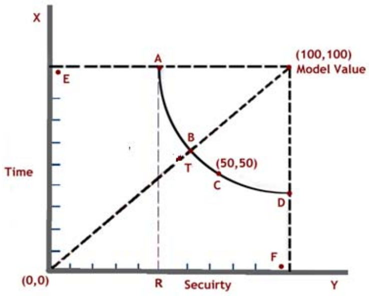


Figure 1. D-system utilizing 2 parameters

The limitations of D-System motivated authors of [18] as they proposed a smart replica selection (RSDG) to overcome them. They argued that the D-System attributes cannot be represented using the exact numerical numbers because they are linguistic variables which better expressed using grey numbers. Grey numbers are used for an accurate attribute expression and to prevent getting replicas with the same attributes values. Their approach uses decision attributes that are derived manually to satisfy the decision table based on the observation of the history or knowledge about the sites. In other work [19] they enhanced their previous strategy by making RSDG strategy work even in the absence of decision attribute (history information of replicas). However their works were absent from the balanced solution that focus in almost same rated QoS parameters and they did not show how to rate sites' QoS parameters. Omar et al [20] enhanced the speed of D-system utilizing genetic algorithm with same aforementioned limitations.

3. Replica selection Strategy Design

Figure 2 illustrates an overview of our proposed Balanced QoS Replica Selection strategy (BQSS). BQSS utilizes many current successful data grid core services, such as Replica Location Service (RLS) and the Network Weather Service (NWS) [21]. The RLS provides the system with the physical file locations, and the NWS provides information about network status. The BQSS selects the best site location which houses the required replica. In this context, the best site is the site closest to the model site with almost equal values of QoS parameters: security, availability and time. For example selecting the site B as shown in Fig. 1, and if B does not exist selecting C and so on as to overcome the limitations of D-system which selects any of A, B, C and D randomly even it is clear that site T is better than C, D and A in terms of balanced values of its parameters. So the proposed solution should deliberately find the best site by selecting the one that is close to the model site and at the same time the value of its parameters are almost the same. This can be achieved by employing both the distance and the scaled standard deviation to find the best site. In the next subsections time rating, security rating, and availability rating are demonstrated.

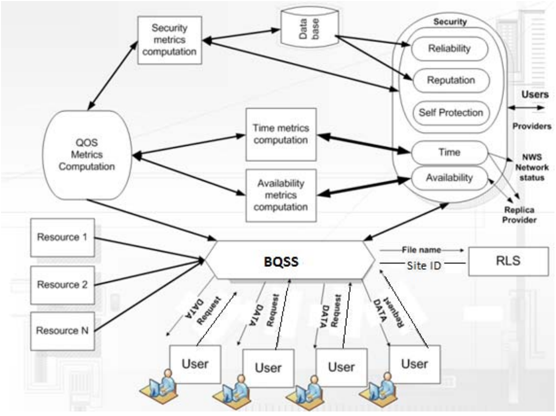


Figure 2. Overview of the proposed data grid model

3.1. Time Rating

In this work time is rated by extending the equations explained in a recent work proposed by E. Husni et al [22] which integrates replica requests that waiting in the storage queue. The response time Ti for a site is calculated as:

Ti = T1i +T2i +T3i (1)

T1: Transfer time, T2: Storage access latency, T3: Request waiting time in the queue.

T1 represents data transmission via a network which depends on network bandwidth and the size of the file. This is computed as follows:

T1= File Size (MB) / Bandwidth (MB/SEC) (2)

In general, the operating systems schedule the disk I/O requests in a manner that improves system performance [23]. Scheduling is implemented by maintaining a queue of requests for the storage device. Therefore, storage speed, the number of requests in queue and file size play major roles in the average response time experienced by applications. As a result, storage access latency (T2) is the time delay of the storage machines to serve earlier requests. Delay time depends on file size and storage type. Consequently, larger data files lead to a T2 increase. Moreover, different storage machines have diverse speeds (data transfer rates) for I/O disk operations. T2 is calculated as follows:

T2 = Files size / Storage Speed (3)

The storage machine receives requests but it can only serve one request at a time [22]. This leads to requests pending being placed in a queue. Input data transfers must be performed prior to an actual request; similarly, output data transfers must be completed after an actual write process request. This is known as a buffering technique [24]. Furthermore, a site will be busy during the period in which it transfers any replica from the storage machine to the network. Any new incoming data requests have to wait for completion of both; the running transaction and any requests that joined the queue prior to the current request [25]. Consequently, the new request should wait for all earlier requests in the storage queue. The waiting time is the sum of time from the first request in queue to the last. Each of these times is the storage access latency time T2. Therefore, the request waiting time in queue (T3i) is calculated as follows:

$$T3 = \sum_{i=1}^{n} T2i \quad (4)$$

n: Number of requests waiting in queue prior to the new request, hence:

Ti= File Size / (Bandwidth +Storage speed) + $\sum_{i=1}^{n} T2i$ (5)

Site rating based on Response Time (T0i) is calculated by:

T0i= Min $\{_{i=1}^{n}$ T0i}/ T0i ×100 (6)

More details are demonstrated in the case study next section.

3.2. Availability Rating

Site availability is the relationship between the operating time declared by the service provider to serve certain VOs and the requirements to transfer a file from the same provider during the replica selection process. Availability ratings require more attention and will be addressed in our future work.

3.3. Security rating

In general, the purpose of a security mechanism is to provide protection against malicious parties. Traditional security mechanisms typically protect the resources from malicious users by restricting access to only authorized users. However, in a multitude of situations within distributed applications one has to protect oneself from those who offer resources. For instance, a resource that provides information can act deceitfully by providing false or misleading information. The traditional mechanisms are unable to protect the users against these types of threats. Trust can be used to overcome such threats in a distributed grid system. Therefore, trust can be helpful to provide entry level security such as authentication and access control. Trust is specified in terms of the relationship between a trustor, trustee and the context in which the target entity is trusted. Our security model (S) uses the computing trust factor (TF) proposed by V. Vijayakumar et al [11]. According to them, TF consists of Self-Protection Capability (SPC) and reputation weightage. They defined the self-protection capability of a site to include its ability to detect intrusions, viruses and unauthorized access as well as having secured file storage and job completing abilities. Conversely, they defined reputation as a mechanism that offers a way to build trust through social control by using community based feedback about the past experiences of entities. Moreover, reliability [13] is also an important element to build trust as reliable sites should be more trusted than less reliable sites. Our security rating model extends beyond the V.Vijayakumar et al [11] model by integrating reliability proposed by [13] with some modifications. The details of the calculation are presented in the succeeding subsections.

3.4. Self-Protection Capability calculation

The Grid Organization Manager (GOM) maintains the self-protection capability of all entities in a grid organization. Every so often, each entity reports its self-protection capability trustfully and honestly to the GOM. The self-protection capability of an entity is calculated by aggregating the values of the security factors mentioned below. The values of these factors differ in range between 0 and 1.

- IDS Capabilities: - The ability of an entity to protect the system against host and network based intrusions.
- Anti-virus Capabilities: - The ability of an entity to defend against viruses and malicious codes.
- Firewall Capabilities: - The ability to protect the entity from other network accesses.
- Authentication Mechanism: - The ability of the mechanism to verify an identity claimed by or for system security.
- Secured File Storage Capabilities: - The ability of an entity to securely store the files needed for job.
- Interoperability: - The ability of an entity to restrict interfacing between concurrent jobs.
- Secured Job Execution: - The ability of an entity to securely execute the job.

The self-protection capability is calculated using the following formula:

$$SPC = \sum Wi \times Ai \qquad i=1,\ldots,n \tag{7}$$

Where n is the total number of factors, W is the weightage and A(i) is the value of the factor.

3.5. Reputation Calculation

Since reputation is a multi-faceted concept [26], it has many aspects (i.e. truthfulness, honesty, etc.). Reputation weightage is calculated via feedback analysis concerning a multitude of security characteristics derived from the previous experiences of the user community. After usage, the users will provide feedback on the attributes to the Reputation Manager (ReMg) based on their experience. The feedback is a value in the range between 0 and 1.The feedback is aggregated from all the users. The ReMg in the grid organization maintains the reputation weightage of all entities. The security attributes considered for reputation are as follows:

- Consistency: - The ability of an entity to perform functions under stated conditions for a specified period of time
- Confidentiality: - The ability to prevent the disclosure of information to unauthorized users
- Truthfulness: - The ability of the entity to protect against unauthorized data modifications
- Security: - The ability of the system to provide job execution and file storage protection
- Privacy: - maintaining information isolated just to oneself
- Non-repudiation: - The inability of something that has performed a particular action to later deny responsibility
- Authentication: - Defined as the process of verifying an identity claimed by or for a system entity.

Based on the aggregated feedback of all the security attributes of an entity, reputation weightage RpW (Ea) is calculated.

3.6. Reliability Calculation

Reliability in general is defined as the capability of a system or component to execute its requisite functions under stated conditions for a particular period. In this research reliability is the extent of confidence that selected replica will function properly with no failures or crashes [13].Reliability is calculated using the following formula:

$$Rb = Nt / Nc \times 100 \tag{8}$$

Where Nt = Number of times the resource is available to the grid resource provider (Rb) and Nc = Number of resource access attempts performed with the condition that both numbers were attempted during the operating hours declared by the resource owner.

3.7. Security Rating &Trust Factor Calculation

The Trust Factor (TF) of each Entity (E) or a grid node is calculated by utilizing the Self-Protection Capability (SPC), Reputation weightage (Rpw) and Reliability (Rbw) weightage calculated as in the following equation:

$$\text{TF (Ei)} = \text{SPC (Ei)} + \text{Rpw (Ei)} + \text{Rbw (Ei)} \tag{9}$$

From this equation we rate each site by assigning values from 0 to 100

$$\text{Si}= \text{TF (Ei)} / \text{Max } \{_{i=1}^{n} \text{TF(Ei)}\}, \quad \text{i=1….n} \tag{10}$$

3.8. Detailed Algorithm

Replica selection is an optimization grid high-level service, defined as the service layer of the Data Grid consisting of two sub layers (i.e. upper and lower). The upper layer provides a high-level service that utilizes the lower layer core services.

The proposed scheme in this paper (replica selection) is well thought out as a high level service that uses core services. The scheme receives user's requests from the resource broker (RB). The BQSS requests related physical file names and locations from the replica location service (RLS). The system obtains the site's related information and network status from the Grid Resource Information Service (GRIS) such as: Network Weather Service (NWS), Meta computing Directory Service (MDS) and the Grid File Transfer Protocol (GridFTP).

BQSS is an optimization dynamic replica selection which is a high-level service that locates the best replica location for grid users. The best replica now may not be the best replica later due to the dynamics of the grid resources. The number of requests, response time, security and availability are changing as time passes. The detailed scheme of BQSS is explained in the following steps:

Step 1: Collects the requests from the users or typically from the Resource Broker (RB).

Step 2: Collects the replica site's information from RLS and collects the existing criteria values from the NWS (i.e. network bandwidth).

Step 3: Rate the sites QoS parameters (equations section 2).

Step 4: Find the modified distance between the sites and the model value using the equation:

$$md = \sqrt{(100-T_o)^2+(100-A_o)^2+(100-S_o)^2} \Big/ \sqrt{3} + Sd(\mathrm{T_O,A_O,S_O}) \,/\beta \tag{11}$$

The distance proposed in D-system did not utilize standard deviation while ***md*** did to increase the distance if the parameters are divers which reduces their chance of being selected. A new metric ***BTAS*** proposed to measure site QoS.

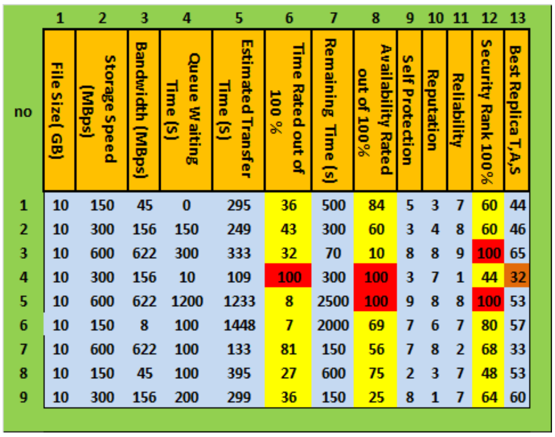
Step 5: Select the site with smallest ***md***.

Step 6: Utilize services like Grid FTP to transfer the replica.

4. Case Study

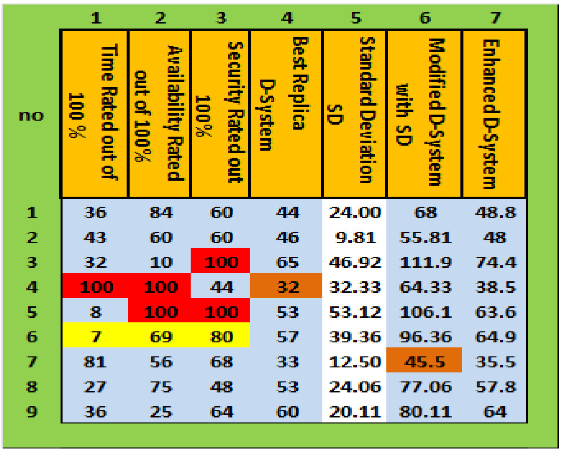
To clarify our approach we made a scenario of 9 sites as shown in Table I, Columns 1 to 4 are the parameters of Eq. 7. The estimated download time based on Eq. 7 from sites1, 2, 3 and 4 are 295s, 249s, 333s and 109s respectively. Site 4 displays the smallest download time so it is rated as a 100% site and site 2 is rated based on Eq. 8, 109/295×100=36%while site 3 is rated 109/249×100=43% and site 3 is rated. 109/333×100=32% as a result, all sites are rated based on estimated download time to make the selection decision in the next step feasible and easier. On the other hand the remaining operating times for the same sites respectively are 500s, 300s, 70s and 200s respectively. Assuming the value of α is 2, the site availability for site 1, is 500/(295×2)×100=84%, and the site availability for site 2 is 300/(249×2)×100=60%. The rest of the calculations are illustrated in tables I & II.

TABLE I. Sites with Different Parameters' Values



	1	2	3	4	5	6	7	8	9	10	11	12	13
no	File Size(GB)	Storage Speed (MBps)	Bandwidth (MBps)	Queue Waiting Time (s)	Estimated Transfer Time (s)	Time Rated out of 100 %	Remaining Time (s)	Availability Rated out of 100%	Self Protection	Reputation	Reliability	Security Rank 100%	Best Replica T,A,S
1	10	150	45	0	295	36	500	84	5	3	7	60	44
2	10	300	156	150	249	43	300	60	3	4	8	60	46
3	10	600	622	300	333	32	70	10	8	8	9	100	65
4	10	300	156	10	109	100	300	100	3	7	1	44	32
5	10	600	622	1200	1233	8	2500	100	9	8	8	100	53
6	10	150	8	100	1448	7	2000	69	7	6	7	80	57
7	10	600	622	100	133	81	150	56	7	8	2	68	33
8	10	150	45	100	395	27	600	75	2	3	7	48	53
9	10	300	156	200	299	36	150	25	8	1	7	64	60

Table II. Best Sites According to D-system and our Approach.



	1	2	3	4	5	6	7
no	Time Rated out of 100 %	Availability Rated out of 100%	Security Rated out 100%	Best Replica D-System	Standard Deviation SD	Modified D-System with SD	Enhanced D-System
1	36	84	60	44	24.00	68	48.8
2	43	60	60	46	9.81	55.81	48
3	32	10	100	65	46.92	111.9	74.4
4	100	100	44	32	32.33	64.33	38.5
5	8	100	100	53	53.12	106.1	63.6
6	7	69	80	57	39.36	96.36	64.9
7	81	56	68	33	12.50	45.5	35.5
8	27	75	48	53	24.06	77.06	57.8
9	36	25	64	60	20.11	80.11	64

However from table 2, column 4, we noticed that the D-system selected the combination in row 4, (100, 100, 44) the problem here is the first two parameters are excellent but the third is below the average which shows unbalanced solution extremes to some parameters over others. Other example can be noticed in rows 1 and 2 where D-system overweight the combination (36, 84, 60) over the combination (43, 60, 60) which is not logical. To overcome this problem the standard deviation is utilized Eq.11 and as shown in Table 2, the aforementioned drawback resolved and the best site now is in row 7, (81, 56, 68) the parameters as block better than row 4, no extremely high and extremely low parameters to somehow the discrepancy is acceptable and the value of the parameters are acceptable as well. Also the problem of rows 1 and 2 is resolved. On the hand even utilizing standard deviation resolved the previous problems however other side effects appeared which can be noticed in rows 5 and 6. Row 5 is better than 6 in all parameters but the standard deviation made it worse therefore the influence of the standard deviation should be reduced or tuned to reasonable values, our preliminary experiments concluded that it should be divided by 5 Eq.11 Enhanced D-System (BQSS)

5. Results and Discussion

A user-defined number of nodes, each with varying performance in time, availability and security, were simulated to compare and verify the results of our new approach. A number of requests for certain data were carried out by the proposed replica strategy. The same experiments were also performed with the same data on D-system. The performance of the new system showed the better results as the standard deviation of the QoS parameters of the new approach is less than D-system as shown in Fig. 3, On the other hand D-system shows slight better performance with respect to TAS metric.

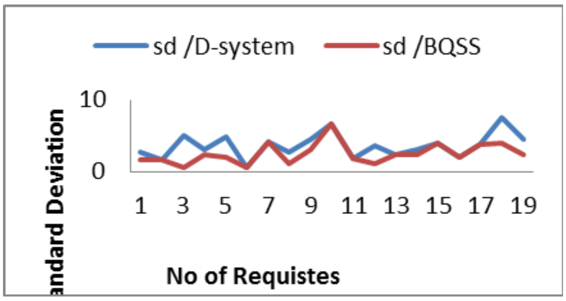


Figure 3. Standard Deviation of QoS Parameters

6. Conclusion

In this paper, we have introduced a new replica selection strategy to overcome D-system limitations. The solution was based on engaging the standard deviation concepts. Also we demonstrated a method of estimating the QoS parameters which was utilized in D-system. The strengths of the algorithm had been investigated and the results of our experiments were presented. The simulation results demonstrated that the new strategy for replica selection enhanced the performance of the grid environment. Results shows promising improvements as the standard deviation values of the three parameters decreased in most cases. In our future work we will focus in integrating users' preferences in the replica selection process.

7. References

[1] M. Lei, S.V. Vrbsky, Q. Zijie, in: Online Grid Replication Optimizers to Improve System Reliability. IPDPS 2007. IEEE International Long Beach, CA, June 2007 pp:1-8.

[2] R. Vingralek, Y. Breitbart, M. Sayal and P. Scheuermann, in: Web++: A System for Fast and Reliable Web Service. In Proceedings of the USENIX Annual Technical Conference,1999.USENIX Association.

[3] M. Sayal, Y. Breitbart, P. Scheuermann and Vingralek in: R. Selection Algorithms for Replicated Web Servers. In Proceedings of the Workshop on Internet Server Performance.1998-Madison.

[4] Load Balancing System, Chapter 6 in Intel Solutions Manual, Intel Corporation, pp. 49-67.

[5] Load Balancing in a Cluster, Web Logic Server 7.0, bea.

[6] S. Lewontin E. Martin in: Client Side Load Balancing for the Web. In Proceedings of 6th International World Wide Web Conference.(Santa Clara, California, April 7-11, 1997).

[7] Z. Fei, S. Bhattacharjee, E. Zegura and M. Ammar in: A Novel Server Selection Technique for Improving Response Time of a Replicated Service. In Proceedings IEEE INFOCOM 1998, pp. 783-791.

[8] Server Load Balancing. Tech Brief from Extreme Networks.

[9] M. Sayal, Y. Breitbart, P. Scheuermann, and R. Vingralek in: Selection Algorithms for Replicated Web Servers. In Proceedings of the Workshop on Internet Server Performance,Wisconsin, June 1998.

[10] T. Ceryen and M. Kevin in: Performance characterization of decentralized algorithms for replica selection in distributed object systems. WOSP 2005: 257-262

[11] V. Vijayakumar and R.WahidaBanu in: Security for Resource Selection in Grid Computing Based On Trust and Reputation Responsiveness', 2008, IJCSNS, VOL.8 No.11

[12] A. Farag and M. Maheswaran in: Evolving and managing trust in grid computing systems", Proceedings of the 2002 IEEE Canadian Conference on Electrical Computer Engineering, 2002, pp 1424-1429.

[13] S. ThamaraiSelvi, P. Balakrishnan, R. Kumar and K. Rajendar in : Trust Based Grid Scheduling Algorithm for Commercial Grids, iccima, vol. 1, pp.545-5581, 2007 International Conference on Computational Intelligence and Multimedia Applications, 2007

[14] G. Kavitha and V. Sankaranarayanan in: Secure Resource Selection in Computational Grid Based on Quantitative Execution Trust, World Academy of Science, Engineering and Technology 72 2010

[15] B. Zhang, X. Yang and X. Qiang in : Trust and Reputation Based Model Selection Mechanism for Decision-Making, Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference on Digital Object Identifier

[16] T. ShaikNaseera , Vivekanandan and K. Murthy, in: Data Replication Using Experience Based Trust in a Data Grid Environment, Proceedings of the 5th International Conference on Distributed Computing and Internet Technology, 2008, New Delhi, India

[17] A. Jaradat, R. Salleh and A. Abid, in: Imitating K-Means to Enhance Data Selection', Journal of Applied Sciences , 2009 , Volume: 9 ,Issue: 19 pp: 3569-3574.

[18] R. Almuttairi, R. Wankar, A. Negi and C. Rao, in: Smart Replica Selection for Data Grids using Rough Set Approximations(RSDG), 2010 IEEE International Conference on Computational Intelligence and Communication Networks, 26-28 Nov 2010, Bhopal, India,

[19] R. Almuttairi, R. Wankar, A. Negi and C Rao, in: Replica Selection in Data Grids Using Preconditioning of Decision Attributes by K-means Clustering (K-RSDG)," vcon, pp.18-23, 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems, 2010

[20] O. Jadaan, W. Abdulal, W. Hameed and M. Jabas, in: Enhancing Data Selection using Genetic Algorithm 2010 International Conference on Computational Intelligence and Communication Networks

[21] W. Changze, M. C. Ming and Y. Chunxiao in: Dynamic Replica Selection Services Based on State Evaluation Strategy', chinagrid, pp.116-119, 2009 Fourth China Grid Annual Conference, 2009

[22] A. Husni and H. Chan, in : Response Time Optimization for Replica Selection Service in Data Grids', Journal of Computer Science, 2008, 4 (6): 487-

[23] S. Aberham, G. Peter and G. Greg, in: Operating System Principles. 7th Edn., Wiley, New York, NY, USA. ISBN 0-471-69466-5

[24] Aberham, S., G. Peter Baer and G. Greg, 2006. Operating System Principles. 7th Edn., Wiley, NY, USA. ISBN 0-471-69466-5

[25] K. Ranganathan and I. Foster, in: Identifying dynamic replication strategies for a high-performance data grid. In: Proceedings of the Second International Workshop on Grid Computing, 1, pp: 75-86.

[26] Y. Wang and J. Vassileva, in: Trust and Reputation Model in Peer-to-Peer Networks. In Proceedings of the 3rd IEEE International Conference on Peer-to-Peer Computing.Linkŏping: IEEE Computer Society (2003), 150–158.