Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.

ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-26

# Enhanced Cost Estimation Model for Replica Selection

[ Kama Hazira Abdul Kadir, Gian Chand Sodhy ]

***Abstract*—A data grid allows large amount of data to be stored and shared among users at geographically different locations. However, when required, data has to be copied to the location where it is needed. This means a lot of network bandwidth will be used for the data transfer. One strategy employed in data grids is to make multiple copies of the data, called replicas, and place them in strategic locations so that the data is not too far from where it might be needed. When data is required, a user now has to choose among the different replicas. Hence, the need for best replica selection to reduce time, effort and resources needed to bring that data to where it is required. Best replica selection tries to estimate the cost involved in copying the data from the alternative replica sites. Most replica cost estimation models use only a few parameters to make the calculations, hence limiting their usage for specific purposes. In this work, we include multiple parameters, taking into account the characteristics of the replica sites as well the network links between the user site and replica sites. We combine the parameters into a formula, and test various scenarios in a simulator. Our results show that this enhanced model can consistently select best replica, and is comparable to other models.**

***Keywords*—replica selection, cost estimation model, data replication, data grids**

## I. Introduction

A data grid connects a collection of hundreds of geographically distributed computers and storage resource locations in different parts of the world to facilitate sharing of data and resources [1]. Grids are made up of a diverse set of machines and instruments with different capacities and purposes. A typical grid job requires three types of resources: computing facilities, data access and storage, and network connectivity [2].

When large amount of data needs to be transferred on the grid, the network will face congestion and latency. Replication is needed to place original copies of data in various storage elements in a grid system to overcome these problems. Replication is a strategy in which multiple copies of some data are stored at multiple sites [3].

Kama Hazira Abdul Kadir
Politeknik Ungku Omar
Malaysia

Gian Chand Sodhy
Universiti Sains Malaysia
Malaysia

Replica selection is of great importance to data intensive scientific computing targeted by many data grid projects and the enabling of virtual data technology. Reference [4] defines replica selection as a process of choosing a replica from among those spread across the grid, based on some characteristic specified by the application. The key to replica selection is the prediction of the file transfer time, which depends on different factors including characteristics of transfer, network status, server load and disk I/O information. A replica management system is normally used to create a replica, locate it at a site, select the best replica and delete or update the replica within the specific sites.

Cost estimation model is a parametric equation used to estimate a specific cost of products or projects. In the replica management system, the cost estimation model is driven by the estimation cost of each replica. The cost calculations are based on many factors such as network latency, bandwidth, replica size and accumulated run-time read/write statistics [5]. Usually a model will be developed to test the replica management system.

Several grid projects have implemented data replication systems that minimized the access time for the largest data volumes. The access time of data intensive jobs in such a heterogeneous environment is difficult to predict without specific monitoring tools. Every replica management system will have a cost estimation model that focuses on a few factors such as network bandwidth, file transfer time or response time. Many other factors (such as storage speed, CPU load, file size, network latency, network utilization, I/O) are usually not considered in these cost estimation models.

To identify the best replica that will be selected by replica management system, cost estimation model is used to calculate the cost of each replica.

## II. Related Work

There are many replica cost estimation and best replica selection models proposed. Each model uses an equation of specific factors that have been identified as cost parameters that contribute towards selecting the best replica. Some of the factors considered are bandwidth cost, network cost, input/output state, CPU load and read/write statistics.

Reference [6] uses a cost model with three significant parameters – network bandwidth, CPU load and I/O state. Their work focuses on GridFTP protocol to improve data transfer. Their cost model can provide users or applications the best choice mechanism for replica selection.

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.

ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-26

Reference [7] uses three parameters - network transfer time, storage access latency and request waiting time in the queue. Their work focuses on how to run the jobs with minimum response time that can be estimated when selecting the best replica. Minimum response time is considered as a criterion for selecting the best replica location in underlying grids.

A cost estimation model proposed in reference [5] tries to estimate data access gains and the maintenance cost of the replica. The cost calculations are based on network latency, bandwidth, replica size, and accumulated run-time read/write statistics. It is basically a data transfer cost. They calculate the cost to optimize the transfer associated with the read/write cost of placing the replica. The formula used can minimize the read/write cost.

Reference [8] uses a system model to calculate total cost of reads based on number of servers, objects to be read, level of replication and total number of bytes that need to be transferred. Their model minimizes the time for standalone computers, but cannot minimize overall time or bandwidth requirements.

Reference [9] selects a replica based on factors such as network latency, storage performance (read/write tests) and estimated file access time. However, the performance depends on network architecture.

Reference [10] proposes a cost model for mobile data management. Their model compares cost saving of allocating a replica with that of replica maintenance cost. The access cost is calculated based on network transmission cost. The average response time for read request is calculated, and the results of access cost and response time are based on network distance. The formula depends on network transmission cost. If the network cost is high, it is not sufficient to implement the formula. However, the performance of their algorithm can be improved if the replication scheme is integrated with other mechanisms such as caching, pre-fetching and data broadcasting.

# III. Enhanced Cost Estimation Model

To improve best replica selection process, we decided to incorporate more parameters into a formula that calculates the costs of a using a replica. These include properties of the site that currently holds the replica as well as properties of the link between the replica site and user site (where the replica will be used). Table I shows the parameters used, their definitions and measurement units.

We justify our choice of parameters as following.

When a user wants to use a data file that is currently located on another site, that file has to be transferred/copied to the user site. This involves two components – the replica site and the network connection between replica site and user site. In order to evaluate which replica site should be used for a particular data file, the properties of the replica site as well as the connection between the replica site and user site are important considerations.

CPU load is dynamic and depends on the loading. If the CPU load is heavy, then it will affect the downloading process of the replica from the site.

Input and output states are busy doing a heavy workload, therefore I/O idle ratio will affect the data transfer process.

Storage speed depends on the hard disk being used. A faster hard disk (with higher RPM) will give a better performance. Bigger drives are also denser, which means the head has to travel a shorter distance between data bits. This speeds up the throughput and bigger drives will have less data fragmentation, since there is more room to write files contiguously.

File size is another factor that also needs to be considered as it can affect the data transfer process. A larger file obviously will take some time to transfer, compared to a small file.

TABLE I. PARAMETER DEFINITIONS AND MEASUREMENT UNITS

| Parameter | Definition | Measurement unit |
|---|---|---|
| CPU load (CPL) | Processor utilization on the site that currently holds the replica. | % |
| I/O idle (IO) | Percentage of time the processor (on the site that currently holds the replica) is waiting for disk input/output operations. | % |
| Storage speed (SP) | The speed of a data storage device on the site that currently holds the replica. | revolutions per minute (RPM) |
| File size (FS) | The size of the data file (or replica) that user wants to access. | MBytes |
| Network bandwidth (NB) | Maximum transmission capacity of the connection between replica site and user site. | Mbps |
| Network utilization (NWU) | Percentage of actual current traffic against maximum capacity of the connection between replica site and user site. | % |
| Network latency (NWL) | Amount of time delay encountered by data travelling on the connection between replica site and user site. | ms |

Network bandwidth is a significant factor to consider in replica selection. It can directly influence the data transfer process. The higher the network bandwidth connecting a replica site with the user site, the faster it will be to download the replica.

Network utilization gives a more accurate picture of the current state of a network. A high percentage value (e.g. 90% of bandwidth) means that the network is currently overwhelmed, hence not desirable for data transfer. Selecting another replica site with a lower network utilization value might be more beneficial since the network will not be congested.

Network latency is another factor that contributes to network speed. High latency will create bottlenecks that decrease the effective bandwidth. Even when the bandwidth is large, percentage of latency can reduce the performance of the bandwidth.

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.

ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-26

Of course there are many other parameters that can be considered. However, we decide to concentrate on these seven parameters that capture the essence of replica site as well as the network link connecting it to the user site. Furthermore, these parameters are easily obtained from system.

Next, we put together these parameters into a formula to calculate the cost of assessing multiple replicas from a user site, which helps decide which replica will be selected.

Our cost estimation formula is as following:

$$Cost_{s-d} = w.CPL + w.IO + w.SP + w.FS + w.NB + w.NWU + w.NWL$$

where

$Cost_{s-d}$ is the score of cost from a source node (s) to a destination node (d),
w is weightage (measured in %),
CPL is CPU load (measured in %),
IO is I/O idle (measured in %),
SP is storage speed (measured in RPM),
FS is file size (measured in MBytes),
NB is network bandwidth (measured in Mbps),
NWU is network utilization (measured in %),
NWL is network latency (measured in ms).

A weightage value is assigned to each parameter in the formula. These weightages can be varied in order to increase or decrease a particular parameter's influence in the calculating of the cost. This will allow us to test different scenarios in our experiment.

# IV. Experiment and Simulation

In order to test our formula, we built our own Java-based simulator, based on OptorSim [2], to specifically test replica selection costs using the parameters described.

The simulator creates a grid topology of a fixed number of nodes and connections. Values are randomly assigned to each node (i.e. CPU load, I/O idle, storage speed, file size) and connection (network bandwidth, network utilization, network latency). These values are then normalized (into a scalar) before multiplied with the specific weightage.

A node is selected from the topology, this becomes a source node. All nodes directly connected to the source node become possible replica sites (called destination nodes). Each destination node's cost total is calculated based on the parameter values assigned. Destination side with the highest score is deemed to be the best replica site for the said source node. In other words, the simulator will try to find the best replica site for each node in the topology.

The simulator uses three configuration files:

- topology – for storing the grid topology,
- node properties – for storing CPU load, I/O idle, storage speed of a node and file size,
- link node properties – for storing network bandwidth, network utilization, network latency for a link.

The algorithm of the simulator is as following:

1. Assign random values to the configuration files (topology, node properties, link node properties).
2. Select one source node.
3. Find out other nodes connected to the source node.
4. Read the values of the destination node.
5. Read the values of the link between source node and destination node.
6. Normalize the parameter values into a scalar.
7. Multiply the normalized values with respective weightages.
8. Sum up the total costs.
9. Compare the total costs.
10. Destination node with the highest total is returned as the best replica.
11. Repeat step 2 until all the source nodes have been processed.

We run a total of 4 experiments, with different weightages for each parameter. At the same time, we also run the same experiment using C.T.Yang's model [6], which only uses three parameters (network bandwidth, CPU load and I/O) with fixed weightages, as shown in Table II.

For each experiment, random values are generated for each parameter and the total cost is calculated according to the formula described earlier. The same parameter values are also used to calculate the total cost for C.T.Yang's model. This allows us to compare our results with C.T.Yang's model in terms of selecting the best replica for each node.

Each experiment is run 25 times, with different random values for each parameter. We store how frequently a particular node is selected as best replica for a given source node, both for our model as well as C.T.Yang's.

# V. Results

Table III shows how many times a particular site/node is selected as best replica in the 4 experiments we conducted (repeated over 25 times), for our model (ECEM) as well as C.T.Yang's (CTY).

For example, in experiment 1, for site 0 (with 3 direct neighbours 1, 2 and 4), our model (1 ECEM) selects site 4 as the best replica (11 times) for it, compared to other candidate sites 1 and 2 (7 times each). When the same experiment is repeated using C.T.Yang's model (1 CTY), the same figures are obtained. This means that site 4 is the best replica for site 0 (under experiment 1 conditions).

Results from experiment 1 shows that our model (ECEM) and C.T.Yang's model (CTY) agree on the selection of best replica for sites 0, 1, 2 and 4. Only for site 3 there's a slight difference. This is expected as the weightages assigned to the parameters follow similar distribution in both the models, i.e. network bandwidth is given the highest weightage.

In experiment 2, the weightage for network bandwidth (NB) is reduced to 40% in our model, while the weightages of the other parameters are slightly increased to 10%. However, the weightages for C.T.Yang's model remains the same. Once again, both models select similar best replica for each site, the exception again being site 3.

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.

ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-26

TABLE II. WEIGHTAGE OF PARAMETERS USED IN DIFFERENT EXPERIMENTS

| Experiment | Enhanced Cost Estimation Model (ECEM) | | | | | | | C.T.Yang Model (CTY) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Parameter Weightage (w)* | | | | | | | *Parameter Weightage (w)* | | |
| | ***SP*** | ***CPL*** | ***IO*** | ***FS*** | ***NB*** | ***NWL*** | ***NWU*** | ***NB*** | ***CPL*** | ***IO*** |
| 1 | 7% | 7% | 7% | 7% | 58% | 7% | 7% | 80% | 10% | 10% |
| 2 | 10% | 10% | 10% | 10% | 40% | 10% | 10% | 80% | 10% | 10% |
| 3 | 20% | 10% | 10% | 10% | 30% | 10% | 10% | 80% | 10% | 10% |
| 4 | 25% | 10% | 10% | 10% | 25% | 10% | 10% | 80% | 10% | 10% |

TABLE III. FREQUENCY OF A SITE CHOSEN AS BEST REPLICA IN DIFFERENT EXPERIMENTS

| $Node_s$-$Node_d$ | Experiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ***1 ECEM*** | ***1 CTY*** | ***2 ECEM*** | ***2 CTY*** | ***3 ECEM*** | ***3 CTY*** | ***4 ECEM*** | ***4 CTY*** |
| Site0-Site1 | 7 | 7 | 6 | 7 | 6 | 7 | 7 | 7 |
| Site0-Site2 | 7 | 7 | 7 | 7 | 9 | 7 | 8 | 7 |
| Site0-Site4 | 11 | 11 | 12 | 11 | 10 | 11 | 10 | 11 |
| | | | | | | | | |
| Site1-Site2 | 5 | 4 | 5 | 4 | 5 | 4 | 6 | 4 |
| Site1-Site3 | 9 | 10 | 10 | 10 | 13 | 10 | 13 | 10 |
| Site1-Site4 | 11 | 11 | 10 | 11 | 7 | 11 | 6 | 11 |
| | | | | | | | | |
| Site2-Site0 | 7 | 8 | 7 | 8 | 8 | 8 | 7 | 8 |
| Site2-Site1 | 6 | 7 | 6 | 7 | 7 | 7 | 6 | 7 |
| Site2-Site3 | 12 | 10 | 12 | 10 | 10 | 10 | 12 | 10 |
| | | | | | | | | |
| Site3-Site1 | 8 | 10 | 8 | 10 | 9 | 10 | 9 | 10 |
| Site3-Site2 | 11 | 8 | 11 | 8 | 9 | 8 | 9 | 8 |
| Site3-Site4 | 6 | 7 | 6 | 7 | 7 | 7 | 7 | 7 |
| | | | | | | | | |
| Site4-Site0 | 10 | 9 | 9 | 9 | 9 | 9 | 8 | 9 |
| Site4-Site1 | 5 | 4 | 5 | 4 | 4 | 4 | 4 | 4 |
| Site4-Site2 | 5 | 4 | 5 | 4 | 6 | 4 | 7 | 4 |
| Site4-Site3 | 5 | 8 | 6 | 8 | 6 | 8 | 6 | 8 |
| | | | | | | | | |
| Site5-Site1 | 4 | 2 | 4 | 2 | 4 | 2 | 4 | 2 |
| Site5-Site3 | 10 | 12 | 10 | 12 | 10 | 12 | 11 | 12 |
| Site5-Site4 | 11 | 11 | 11 | 11 | 11 | 11 | 10 | 11 |

In experiment 3, the weightage for network bandwidth (NB) is reduced to 30% while for storage speed (SP) the weightage is increased to 20% in our model. The weightages of the other parameters are maintained at 10%. The selection of best replica for sites 0, 2, 3 and 4 matches the choices made by each model. Only for site 1, our model differs slightly from the choice made by C.T.Yang's.

In experiment 4, network bandwidth (NB) and storage speed (SP) are both assigned weightages of 25%, whereas other parameters remain at 10% in our model. Once again, both models agree on best replica selection for sites 0, 2, 3 and 4; with slight variation for site 1.

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.

ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-26

These experiments have shown that our enhanced cost estimation model is capable of selecting the best replica in a consistent manner. The results show that the same replica is given the highest score most of the time even when the parameter weightages are changed and random values assigned. The results also show that our model compares quite well with C.T Yang's, especially in experiment 1 when the parameter weightage distribution is quite similar.

This proves that our model is quite reliable in selecting the best replica in different circumstances, by incorporating more parameters.

# VI. Conclusion

In this work, we have incorporated multiple parameters in selecting a best replica, by taking into account the properties of the replica sites as well the network links among them. This can help to improve cost estimation models used by replica managers.

## *References*

[1] H. Lamehamedi, B. Szymanski, Z. Shentu, and E. Deelman, "Data replication strategies in grid environments," Proceedings of the. Fifth International Conference on Algorithms and Architectures for Parallel Processing, Beijing, China, Oct. 23-25, 2002, pp. 378–383.

[2] W. H. Bell, D. G. Cameron, A. P. Millar, L. Capozza, K. Stockinger, and Floriano Zini (2003), "Optorsim: A grid simulator for studying dynamic data replication strategies," International Journal of High Performance Computing Applications, 2003, 17(4), pp. 403-416.

[3] P. A. Bernstein, V. Hadzilacos and N. Goodman, Concurrency Control and Recovery in Database Systems, Addison-Wesley, 1987.

[4] S. Vazhkudai, S. Tuecke and I. Foster, "Replica selection in the globus data grid," Proceedings of the 1st International Symposium on Cluster Computing and the Grid, Brisbane, Australia, May 15-18, 2001, pp: 106-113.

[5] H. Lamehamedi, Z. Shentu, B. Szymanski, and E. Deelman, "Simulation of dynamic data replication strategies in Data Grids', Proceedings of the 17th International Symposium on Parallel and Distributed Processing, 2003. Proceedings. International, Nice, France, April 22-26, 2003, pp. 100.2.

[6] C. T. Yang, C. H. Chen, K. C. Li, and C. H. Hsu, 'Performance analysis of applying replica selection technology for data grid environments', Proceedings of the 8th international conference on Parallel Computing Technologies (PaCT'05), Krasnoyarsk, Russia, Sep 2005, pp. 278-287.

[7] H.H.E. Al-Mistarihi, and C.H. Yong, "Response time optimization for replica selection service in data grids", Journal of Computer Science, vol 4(6), 2008, pp. 487-493.

[8] T. Loukopoulos, and I. Ahmad, "Static and adaptive data replication algorithms for fast information access in large distributed systems," Proceedings of 20th International Conference on Distributed Computing Systems, Taipei, Taiwan, April 10-13, 2000, pp.385-392.

[9] R. Słota, D. Nikolow, L. Skital, and J. Kitowski, "Implementation of replication methods in the grid environment," Advances in Grid Computing - European Grid Conference, Amsterdam, Netherlands, Feb 14-16, 2005, pp. 474-484,

[10] S. Y. Wu and Y. T. Chang, "An active replication scheme for mobile data management," Proceedings of 6th International Conference on Database Systems for Advanced Applications, Hsinchu, Taiwan, Apr 19-21, 1999, pp. 143-150.