International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 2 No. 4 Apr. 2013, pp. 484-487

http://www.issr-journals.org/ijias/

Impact of arrival rate of job / resource and Dual queues in the Matchmaking process in Grid Environment

Japhynth Jacob[1], Dr. R. Elijah Blessing[2], and Dr. J. R. Isaac Balasingh[3]

[1]Associate professor,
Dr. G. U. Pope College of Engineering,
Sawyerpuram, TamilNadu, India

[2]Director,
KSCST/Karunya University,
Coimbatore, India

[3]Dean,
Dr. G. U. Pope College of Engineering,
Sawyerpuram, TamilNadu, India

ABSTRACT: A grid is an open system, a large collection of autonomous systems giving individual users the image of a single virtual machine with a rich set of hardware and software resources. The crest aim of a Grid is to allocate best resource to a job by comparing the user requirements with the resource capabiliities. The process of selecting resources based on user requirements is called “resource matching”. In Grid environment the resource pool is highly dynamic. The user behavior in the Grid environment is also cannot be predicted. While matching the job with the resources, the average response time to find the best match depends upon the arrival rate of the job and the resources. The average response time to find the best match can be increased by increasing the service rate of the Grid. The service rate of the Grid can be increased by having two queues in the Grid. This paper deals with the impact of time varying arrival rate of job, a large time varying arrival rate of resource, number of queues, in matching the user requirements with the resource capabilities. The arrival rate of job and the resource would require to be in a fashion that the average response time of the job should get minimized.

KEYWORDS: Guaranteed time, matching, average response time, arrival rate, service time.

1 INTRODUCTION

In most organizations there are large amounts of underutilized computing resources. Most desktops are busy less than 5% of time [1]. Often machines have enormous unused disk drive capacity. Grid computing provides a framework for exploiting these underutilized resources by matching the user requirement with the resource capabilities before selecting a resource for a job. The Grid can help in enforcing security rules and implement policies, which can resolve priorities for both job and resources. The job and the resource arrival rate in the Grid environment usually differs from one grid site to another. The job and the resource arrival in the Grid follows Poisson distribution function [2]. The rest of the paper explains the impact of the arrival pattern of job and the resource in successfully completing the job.

Corresponding Author: Japhynth Jacob (jafijacob@yahoo.co.in)

2 JOB AND THE RESOURCE ARRIVAL PATTERN

The job/resource distribution in Grid follows poison distribution function [2]. The total time interval is divided into small fixed time intervals. Each "small fixed time intervals" can have only 0 or 1 occurrences of jobs. The Poisson arrival pattern is as follows:

$$p(y)=\frac{n!}{y!(n-y)!}p^y(1-p)^{n-y}=\frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^y(1-\frac{\lambda}{n})^{n-y}$$

Taking limit as: $n\to\infty$

$$\lim_{n\to\infty}p(y)=\lim_{n\to\infty}\frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^y(1-\frac{\lambda}{n})^{n-y}=\frac{\lambda^y}{y!}\lim_{n\to\infty}\frac{n(n-1)-(n-y+1)(n-y)!}{n^y(n-y)!}\left(1-\frac{\lambda}{n}\right)^n(\frac{n-\lambda}{n})^{-y}$$

$$=-\frac{\lambda^y}{y!}\lim_{n\to\infty}\frac{n(n-1)-(n-y+1)}{(n-\lambda)^y}\left(1-\frac{\lambda}{n}\right)^n-\frac{\lambda^y}{y!}\lim_{n\to\infty}\left(\frac{n}{n-\lambda}\right)\left(\frac{n-1}{n-\lambda}\right)-\left(\frac{n-y+1}{n-\lambda}\right)\left(1-\frac{\lambda}{n}\right)^n$$

Note: $\lim_{n\to\infty}\left(\frac{n}{n-\lambda}\right)=\lim_{n\to\infty}\left(\frac{n-y+1}{n-\lambda}\right)=1$ for all fixed y

$$\lim_{n\to\infty}p(y)=\frac{\lambda^y}{y!}\lim_{n\to\infty}(1-\frac{\lambda}{n})^n$$

From calculus we get: $\lim_{n\to\infty}(1-\frac{a}{n})^{\alpha}=e^{\alpha}$

$$\Rightarrow\lim_{n\to\infty}p(y)=\frac{\lambda^y}{y!}e^{-\lambda}=\frac{e^{-\lambda}\lambda^y}{y!}\qquad y=0,1,2,....$$

3 IMPACT OF ARRIVAL RATE OF JOB AND RESOURCE IN MATCHMAKING PROCESS

Let $job(j)$ denotes $j-t$h job in the order of jobs. Let $start(j)$ denotes the time that the job j enter into the Grid site for matchmaking process. Let $Finish(j)$ be the time that the job has been finished allocated to a resource. Let $start(j)$ be the time that the job has started its execution with the selected resource. The $latency\ L(j)$ of job j with its initial resource r_{init} be

$$L(j)\ =\ start(j)-\ Finish(j) \qquad (1)$$

Let $Lcmax$ denote the maximum acceptable latency $L(j)$ of job (j) for user c, If the arrival rate of resource is less than the arrival rate of job.

$$L(j)>Lmax(j) \qquad (2)$$

Let $J(t)$ be the set of all jobs that arrive at time t for user c, then using the helper function

I(X,Y)= 1 if X>y

I(X,Y) = 0 otherwise

$$Lapsed\ jobs\ =\ (L(j),Lcmax(j)) \qquad (3)$$

Hence it is inferred that as the lapsed job decreases the average response time of job increases.

4 IMPACT OF DUAL QUEUE IN THE MATCHMAKING PROCESS

Traditional matchmaking model has a single queue of jobs which are ready to be processed. Jobs may experience delays inside the single queue which results in increase in the average response time. This motivates to design the Dual Queue Model *(DQM)* which maintains two queues viz*., new_job* queue (for new jobs) and *mature_job* queue (for old jobs). In traditional matchmaking system, all jobs sink into a single sink.

In *Dual Queue Model*, job entering, sinks either into the *mature_job* queue or into the *new_job* queue [3]. The mean service rate of a system with more number of sinks is higher than the mean service rate of a system with less number of sink. By maintaining arrival rate constant and the mean service rate of a system high, the waiting time of a job in the queue is reduced. Thus the average waiting time of a job is reduced in the system with more number of sinks and increased in the system with less number of sinks.

5 EXPERIMENTAL RESULTS AND DISCUSSIONS

The system was simulated using Gridsim [4] with discrete event simulation in Java. The matchmaking process is represented as a sequence of events. Each event occurs at an instant of time and marks a change in the state of the Grid system. The grid workload comprises single-processor jobs, which are sent to the grid in batches. A batch submission [5] is the set of jobs ordered at the time of arrival into the Grid site. New allocation procedures are integrated into Grid by extending the AllocPolicy class. The simulated environment encompasses the DAS-3 grids, for a total of 5 grid sites and over 225 processors. Each sites of the combined system receives independent stream of jobs. The experiments were conducted with 75 service providers with varying number of resources. The experimental results are the average of the 10 sets of synthetic job streams for each load levels viz. 20%, 30%, 40%, 50%, and 60%. The resource ingress is entry of resources and resource egress is exit of resources. The following graph shows the impact of arrival rate of resources into the grid site in successfully completing the jobs.

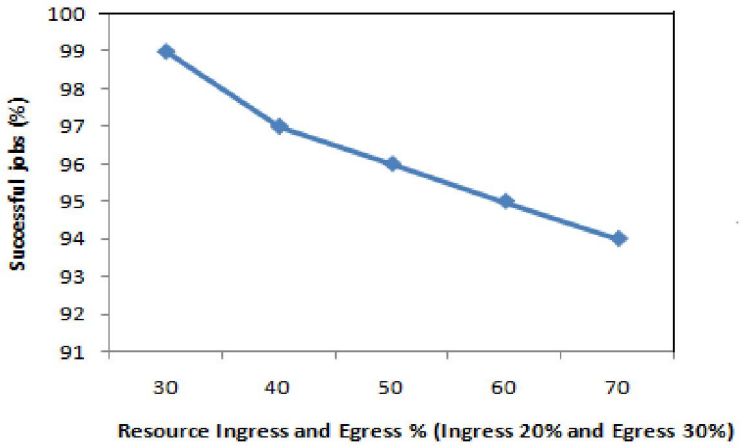

Fig. 1. Successful jobs (%) Vs Resource Ingress and Egress (%)

The experiment is conducted to compare the average response time of single queue system with Dual Queue system. The experiment is conducted with 75 service providers evenly distributed in all grid sites. The number of jobs submitted ranges from 900 to 1800.

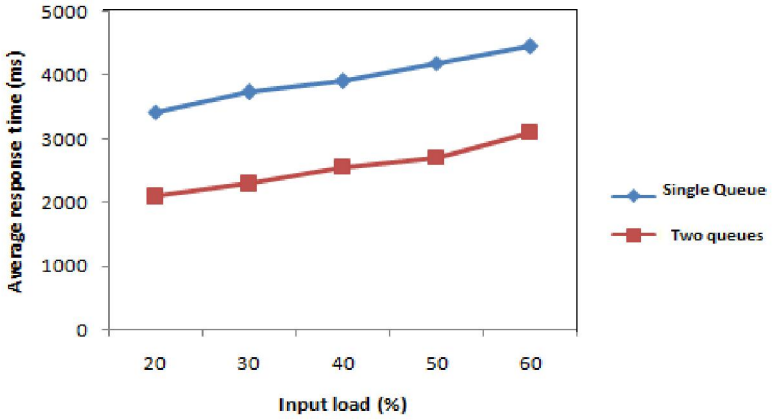

Fig. 2. Average response time (s) Vs Input load (%) for single queue and two queues

The Figure 2 shows the average response time when the Grid uses single queue and two queues. It has been observed that, the average response time when two queues are used is relatively lower than the average response time when single queue is used. This is due to the fact that the dual queue reduces the average waiting time of the job in the queue which in turn reduces the average response time.

6 CONCLUSION

Matchmaking is a process of evaluating the degree of similarity between any two objects. The impact of arrival rate in decreasing the average response time and the impact of two having two queues at the grid site in minimizing the average response time is analyzed. As the average response time is decreased the Grid can complete more number of jobs within the guaranteed time.

REFERENCES

[1] Bart Jacob, Michael Brown, Kentaro Fukui, Nihar Trivedi, "Introduction to Grid Computing, *IBM*, First Edition, December 2005.

[2] Haight, Frank A., "Handbook of the Poisson Distribution", John *Wiley & Sons, Inc.,* New York, 1967.

[3] Rajiv Ranjan, Aaron Harwood and Rajkumar Buyya, "Case for Cooperative and incentive Based Federation of Distributed Clusters", *Future Generation Computer Systems,* Vol.24, pp. 280-295, May 2007.

[4] Simscript: a simulation language for building large-scale, complex simulation models. [Online] Available: http://www.simscript.org, 2008.

[5] Z/OS, "VIRI 2.0 MVS Planning Global Resource Serialization", SA22-7600-09, Copyright IBM Corporation, Edition 10, September 2009.