A Modified Performance Oriented Approach for Load Balancing in Cloud Computing

Anterpreet Kaur
M.Tech. Student
Computer Science Department
Geeta Institute of Management & Technology

Anurag Jain
Assistant Professor
Computer Science Department
Geeta Institute of Management & Technology

ABSTRACT
Cloud computing is advancing with a great pace. It has been already adopted over a large user base. Easy to use and anywhere access like capabilities of cloud computing has made it attractive among other technologies. It has not only reduced the deployment cost on user side but also allowed the big companies to sell their infrastructure to reduce the installation cost for the small organization. Roots of cloud computing extends to Grid computing. Along with the features of its ancestor technologies it also carries the loopholes present in those technologies. Some of these are identified and corrected in the recent times but still there remains the need for the improvement in cloud computing. These improvements can be divided widely into two categories viz. performance and security. The work done in this paper is devoted to performance enhancement for the user of existing cloud system. There are many simulation environments available to simulate this type of work. Among these available tools cloud Analyst is having attractive GUI along with transparent control of parameters. The Cloud Analyst is used to simulate the enhancements done in the existing system. Among various components load balancing task is used for the modification and enhancement aiming to reduce the response time, computational time and cost included. The results are analyzed and compared with the existing algorithms and as observed, proposed work is one step ahead of existing techniques.

Keywords
Cloud Computing, Virtual Machines, Cloud Service Provider, Load Balancing, Cloud Analyst.

1. INTRODUCTION
Enterprises have been striving to trim down computing costs and for that reason most of them start strengthening their IT operations and later using virtualization technologies. Cloud Computing has raised the enterprises search to a new level and helped them to further reduce costs through enhanced utilization, reduced administration and infrastructure cost and quicker deployment cycles [1]. Cloud Computing is a term used to indicate both a platform and application. As a platform it supplies, configures and reconfigures server. Here, these servers can be physical machines or virtual machines. On the other hand, Cloud Computing describes applications which are accessible through an internet and for this purpose large data centers and powerful servers are used to host the web applications and web services [2]. The cloud is a metaphor for the Internet and it is an abstraction for the complex set up. Cloud Computing differs from traditional computing paradigms as it is scalable, & provide different level of services to the clients, driven by economies of scale and the services are dynamically configurable [3].

Load balancing is used to distribute the workloads across the different. With the increase in user base for the services of cloud computing , a huge amount of user are there on internet; to manage such number of users it becomes essential to distribute the load evenly according to the requirement of task. So there is a need of efficient load balancing technique [4].

In section 2, basics of load balancing has been discussed. Section 3 covers the related work which is the source of motivation. Proposed load balancing algorithm has been given in the section 4. Section 5 includes the configuration of simulation environment while results have been given in section 6. It is followed by conclusion in section 7.

2. LOAD BALANCING
Load balancing is used to distribute the workloads across the different computing resources, such as computers, computer cluster, disk drivers, network links, CPU. It is the process of distribution of tasks among available nodes of a distributed system to get better both resource utilization and job response time. It also avoids situation where some nodes are overloaded while other nodes under loaded [5].

With the increase in user base for the services of cloud computing , a huge amount of user are there on internet , to manage such number of users it becomes essential to ensure they get the workload spread to all available servers for the achievement of aim of high user satisfaction. Load balancer is also available for other technologies and other services along with the cloud computing. This is done at different levels

- At VM level, the scale is mapped in such a way that the mapping can be done to have the VM with available physical computers to balance the load of several applications.
- At Host level, the scale is mapped to have the required coordination between the virtual machine and host resources available to perform different tasks coming from the application.

2.1 Need of Load Balancing in Cloud computing
Load balancing in clouds is a process that distributes the excess dynamic local workload equally across all the nodes. It is used to attain a high user satisfaction and resource utilization ratio

2.2 Classification of Load Balancing Algorithm
The available algorithms are classified into two classes which are based on the nature of decision making process: Static and dynamic load balancing [7].

2.2.1 Static Load Balancing Algorithm

Static approach is much simpler as compared with dynamic approach. Static approach requires the prior knowledge of the status of distributed system globally. It does not consider the present state or scenario of a node while allocating the load. It divides the traffic in equal to all available servers.

2.2.2 Dynamic Load Balancing Algorithm

Dynamic load balancing is done with the knowledge of the present state of a node while the allocation process is going on. It does not require the history information of global parameters. It is more convenient and comfortable for a distributed system that is widely spread over different regions. It divides the incoming requests according to the available capacity of available resources. In case of failure, it will provide the robustness to system by replacing it with the active and live nodes.

3. RELATED WORK

- Kumar V. et. al. in [7] have surveyed the load balancing issue in cloud computing and analyzed various techniques used in load balancing. They have proposed a new load balancing approach and compared it with the existing one.
- Kapgate D. et. al. in [8] have discussed the scenario of mobile cloud computing, and the need for load balancing in it. They have discussed how load balancing helps to achieve a higher user satisfaction and resource utilization ratio. Further, the authors purposed predictive load balancing strategy for reduction of computational latency in mobile cloud computing.
- Singh A. et. al. in [9] have simulated the existing load balancing algorithm on cloud analyst simulator. Load indices used for these algorithms are Data Center, VMs, Image Size, Memory, and Bandwidth. The results show the impact of choosing load balancing algorithm on the performance of server.
- Kansal N. et. al. in [10] have discussed the existing load balancing techniques in cloud computing and further compares them based on various parameters like performance, scalability, associated overhead etc. They have also discussed these techniques from energy consumption and carbon emission perspective.
- Mohapatra S. et. al. in [11] have simulated four different load balancing algorithms Each algorithm is observed and their balancing criteria like average response time, data center service time and total cost of different data centers are found. According to the experiment and analysis throttled algorithm has the best integrate performance among all of them.

4. PROPOSED WORK

After the literature analysis of existing work related to load balancing, authors have proposed a new approach for load balancing algorithm. Proposed approach is modified version of throttled load balancing approach.

Algorithm Proposed _Greedy()

{

Step 1: Get the list of available virtual machines.

Step 2: Pick the current task from the task queue.

Step 3: Get the virtual machine having lowestresponse time or the virtual machine whose characteristics suits the requirements of task..

Step 4: If (virtual machine is available)

{

Assign the task to that virtual machine.

}

Else

{

Wait till no such virtual machine is available.

}

Step 5: Whenever the task is completed, do the following:

a. Free the virtual machine.

b. Get the response time parameter.

c. Sort the available virtual machine according to efficiency based on response time and maintain the temporary list.

Step 6: If task list is empty then exit otherwise go to step 1.

}

Proposed algorithm consists of all the steps which include more than one place of code for the implementation in Cloud Analyst simulator. The implementation is done with the help of existing libraries and adding further the needed code of the above proposed algorithm. Then this code is further extended by implementing a new series of experiments and having the custom scenario for the simulation and the results are obtained with the help of the console output that is formatted and designed in such a way to provide us the complete view of the algorithm performance.

5. SIMULATOR CONFIGURATION

Simulation Scenario provides the view and information about the scale and area we considered for our study. Scenario provides the information about the resources and available structure of the network to replicate and get the results.

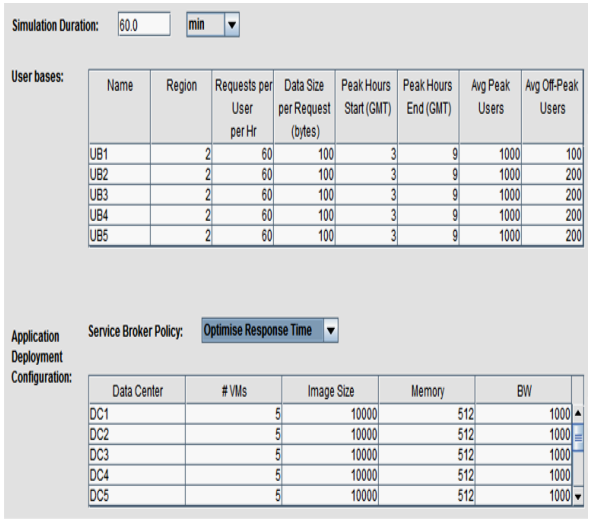

Figure 1: Configuration of User Base and Broker Policy

User base is used to generate the required traffic and the request for the services. Parameters for this include Region, request/hour, Data size, Peak hours, Average no. of users present in peak hours and Avg. Off-Peak users.

Name	Region	OS	VMM	Cost per Vm in $/Hr	Memory cost in $/second	Memory cost in $/second	Data Transfer cost in $/GB	Physical Harware Unit
DC 1	5	Linux	Xen	0.1	0.05	0.1	0.1	2
DC 2	5	Linux	Xen	0.1	0.05	0.1	0.1	2
DC 3	4	Linux	Xen	0.1	0.05	0.1	0.1	3
DC 4	4	Linux	Xen	0.1	0.05	0.1	0.1	3
DC 5	5	Linux	Xen	0.1	0.05	0.1	0.1	2
DC 6	5	Linux	Xen	0.1	0.05	0.1	0.1	2
DC 7	3	Linux	Xen	0.1	0.05	0.1	0.1	4
DC 8	3	Linux	Xen	0.1	0.05	0.1	0.1	4
DC 9	3	Linux	Xen	0.1	0.05	0.1	0.1	4
DC 10	5	Linux	Xen	0.1	0.05	0.1	0.1	3

Figure 2: Data Centers Configuration

Datacenter lists the available resources including hardware, their availability in a region, Architecture, Operating system, VMM, Cost and physical hardware units.

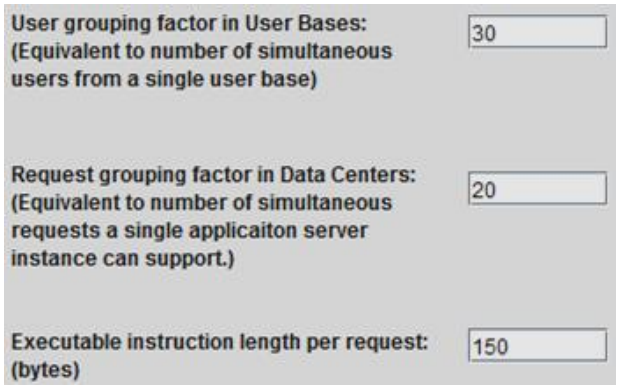

Figure 3: Other parameters for simulation

Other parameters include the user grouping factor used for them, request grouping factor and executable instruction length /request.

6. RESULTS & ANALYSIS

Quantitative analysis of algorithm is presented in figure 4, 5 & 6. For each algorithm, their response time, processing time & cost for jobs is depicted with different colors. A proposed greedy algorithm is shown in blue color, round robin algorithm in red color and throttled in green color and equally spread algorithm is shown in yellow color. Firstly for response time, each algorithm results are depicted below.

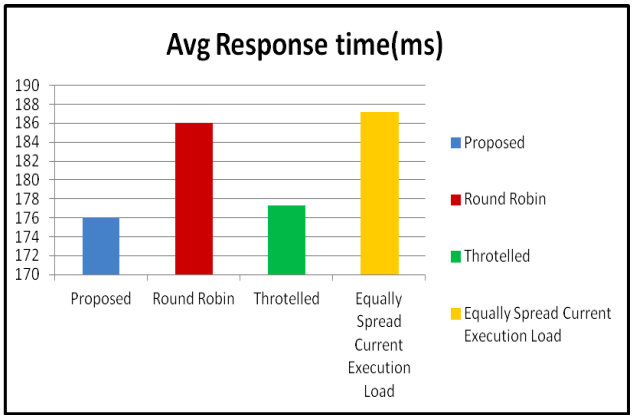

Figure 4: Average Response time comparison

For data processing time, each algorithm results are depicted below.

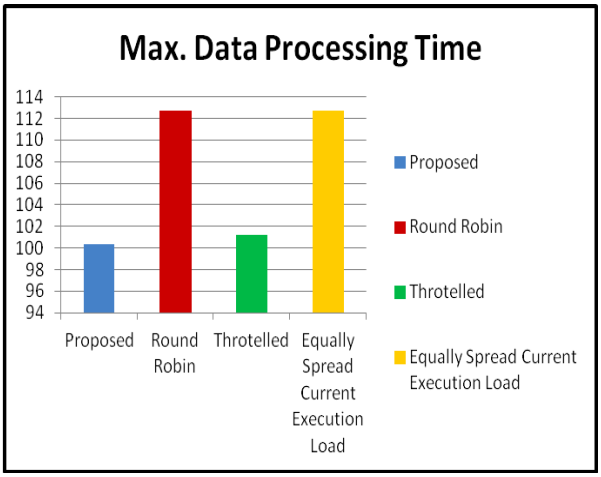

Figure 5: Data Processing Time comparison

For cost, each algorithm results are depicted below.

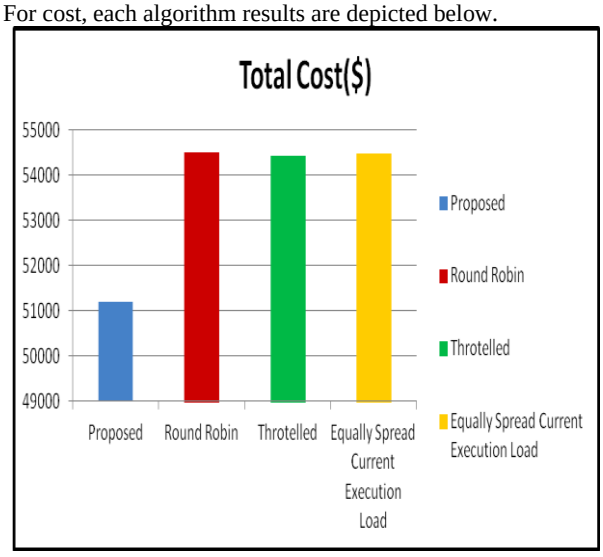

Figure 6: Cost comparison

6.1 Comparative Analysis

Comparative analysis of throttled load balancing approach with the modified approach is given in table 1.

Table 1: Comparative Analysis

Sr. No.	Throttled Load Balancing Algorithm	Proposed Greedy Approach
1	Job manager reference to virtual machine's indexed list is limited to first index only.	The reference to virtual machine's indexed list is not limited to first index, it go through the complete list when required.
2	Every time it picks the first VM from the indexed list of job manager.	Every time it picks the VM that is having higher performance metric.
3	If job manager doesn't have the information about the available vm, then it is put in the queue.	Due to greedy nature it keeps the dynamic information about the VMs present in indexed list.
4	Dependability on job manager makes it vulnerable to delay.	Any dependency on any index is not present to avoid any such delay.

7. CONCLUSION

By the nature of the problem the adopted strategy applied in the proposed algorithm has the improved results as compared to the other three i.e. round robin, equally spread current execution load and the throttled. The throttled is known best load balancing strategy among these. As a result of the comparison, it is accompanied with the conventional round robin strategy. As the result of the comparison it is clear that the performance for the proposed algorithm is better than the existing load balancing strategy. When it is compared by user bases the individual result may vary according the favorable situations. The overall comparison is done from the average values.

The values attained overall by the proposed algorithm are best in terms of average values. The minimum values for the response

time and computational time matches with the performance of throttled sometimes. The values attained by these parameters lacks only in terms of Maximum time in some cases. Overall performance, of the proposed algorithm is better than previous best performing algorithm i.e. throttled.

8. REFERENCES

[1] Patidar, S., Rane, D., Jain P. (2012), 'A Survey Paper on Cloud Computing', Second International Conference on Advanced Computing & Communication Technologies, IEEE Computer Society.

[2] Harauz, J., Kaufinan, L., Potter, B. (2009), 'Data Security in the World of Cloud Computing', IEEE Security & Privacy and Reliability Societies.

[3] Jain, A., Kumar, R. (2014), 'A Taxonomy of Cloud Computing', International Journal of Scientific and Research, Volume 4, Issue 7, pp. 1-5.

[4] Oliveira, C. (2012), 'Load Balancing on Virtualized Web Servers', IEEE Computer Society, pp. 1-6.

[5] Randles, M., David, A., Bendiab, T. (2010), 'A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing', IEEE Computer Society, pp. 1-5.

[6] Shah, P., Shah, S. (2010), 'Load Balancing in Distributed System', International Journal of computer Applications, pp. 1-4.

[7] Kumar, V., Prakash, S. (2014), 'A Load Balancing Based Cloud Computing Techniques and Challenges', International Journal of scientific research and management (IJSRM), Volume 2, Issue5, pp. 1-8.

[8] Kapgate, D., Narnaware, M. (2013), 'Predictive Load Balancing Strategy for reduction of Latency in Mobile Cloud Computing', International Journal of Computer & Communication Engineering Research (IJCCER), Volume 1, Issue 1, pp. 1-5.

[9] Singh, A., Bedi, R., Gupta S. , 'Comparative Analysis of Load Balancing Algorithms in cloud Computing', International Journal of Advanced Technology & Engineering Research (IJATER), Volume 4, Issue 2, pp. 1-4.

[10] Kansal, N., Chana, I. (2012), 'Cloud Load Balancing Techniques: A Step Towards Green Computing', IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1

[11] Mohapatra, S., Rekha, K., Mohanty S., 'A Comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing', International Journal of Computer Applications, Volume 68, pp. 1-5.