Selective Heuristic Load Balancing In Cloud Computing

[1]Pooja Mangla, [2]Dr. Sandip Kr. Goyal
[1]*PhD Scholar*, [2]*Guide and Head*
[1]*CSE Department*,
[1]*M.M. University, Mullana, Ambala, Haryana, India*

Abstract

Cloud Cloud computing is the provision of dynamically scalable and often virtualized resources as a service over the Internet (public cloud) or intranet (private cloud). Load Balancing is the term used for distributing a larger processing load to smaller nodes for getting the working done efficiently. In this work, a new heuristic approach has been implemented, which is known as MinMin-MinMax Selective and the results are shown. Comparison with existing Max-Min and Min-Min approaches have been shown on the basis of few parameters like makespan, load balancing level etc..

Index Terms – *Cloud Computing, Load Balancing, Heuristic, MaxMin, MinMin***.**

I. INTRODUCTION

Cloud computing is a new technology which uses virtual machine instead of physical machine to host, store and network the different components. Cloud Computing refers to manipulating, configuring, and accessing the applications online. It offers online data storage, infrastructure and application. A Cloud is the parallel and distributed system possessing a group of inter-connected and virtualized computers that are dynamically scheduled and highlighted as one or more unified computing resources based on service-level agreements established through conciliation between the service provider and consumers [1]. It is a way of managing large number of highly virtualized resources such that, from a management perspective, they resemble a single large resource. There is a greater need for IT to help address business challenges and cloud computing can help you do all of these:

A. Doing more with Less: Reduce capital expenditures and operational expenses.

B. Higher Quality Services: Improve quality of services and deliver new services that help the businesses to grow and reduce costs.

C. Reducing Risk: Ensure the right levels of security and resiliency across all businesses data and processes.

D. Breakthrough Agility: Increase ability to quickly deliver new services to capitalizon opportunities while containing costs and managing risks.

A. Model of Cloud Computing

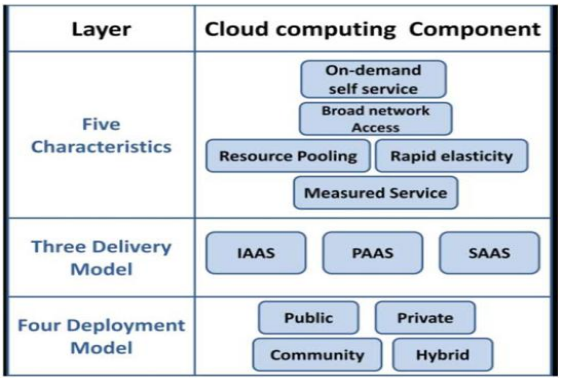


Figure 1.1: Model for Cloud Computing

II. LOAD BALANCING

Load balancing is a generic term used for distributing a larger processing load to smaller processing nodes for enhancing the overall performance of system. Load Balancing is done with the help of load balancers where each incoming request is redirected and is transparent to client who makes the request.

Load Balancing helps in improving the performance substantially, maintenance of system stability, cost effectiveness, etc. Load balancing is a process of reassigning the total load:

- to the individual nodes of the collective system
- to make resource utilization effective
- to improve the response time of the job
- simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded.

III. HEURISTIC LOAD BALANCING IN CLOUD COMPUTING

In order to better use tremendous capabilities of this large scale distributed system, effective and efficient scheduling algorithms are needed. Various algorithms have been designed and implemented and so as we also has introduced a new scheduling algorithm based on two conventional scheduling algorithms, Min-Min and Max-Min, to use their various cons and at the same time, tried to overcome their pros. This heuristic scheduling algorithm, called

min-min min-max selective, is evaluated using a Cloud simulator called CloudSim. Min-Min and Max-Min heuristics are reviewed and it is seen that depending on the length of unassigned tasks in MT (meta-task), one of these heuristics has better results than the other one [7]. For example, if there is only one long task and too many short tasks, Max-Min will execute long task first and allows short tasks to be executed concurrently with the long task, resulting better makespan and even better resource utilization rate and load balancing level, compared to Min-Min that executes all short tasks first and then executes the long task.

A. Basic Steps for Proposed Algorithm

In this the advantages and disadvantages of Min-Min and Max-Min Algorithms are considered and a new algorithm is proposed in which firstly all the tasks in Meta-Task MT are sorted and their completion time is calculated. Then for all tasks, machine having that tasks, having minimum completion time is selected. If more than one such machines are obtained then the machine having lesser resources is selected and the standard deviation is calculated. This is compared with the threshold value so as to assign the particular task to a particular machine. As soon as task is allotted to the machine, it is removed form the Meta-task list.

Following is the algorithm which is implemented:

for all tasks T_i in the Meta-set M_i on all the resources R_j
calculate the average task length avgTL
do until all tasks in M_i are mapped
for each task in Meta-Set M_i
if t_i.getMI() <= avgTL
create a sub-list of tasks t_k
for all resources r_j
find the task t_k in the meta-set with earliest completion time and the machine that obtains it
update C_{ij} for all other tasks in Meta-Set
else
create a sub-list of tasks t_k
for all the resources r_j
find the minimum completion time of i^{th} task in the Meta-Set
find the task t_k in the meta-set with earliest completion time and the machine that obtains it
assign the task t_k to the machine r_j which gives maximum minimum completion time
delete this task from the Meta-Set
update C_{ij} for all other tasks in Meta-Set
end do

B. Performance Metrics

A. Makespan: Makespan is a measure of the throughput of the heterogeneous computing systems, such as cloud. It can be calculated as the following relation:

makespan = max(CTi) , where ti belongs to MT

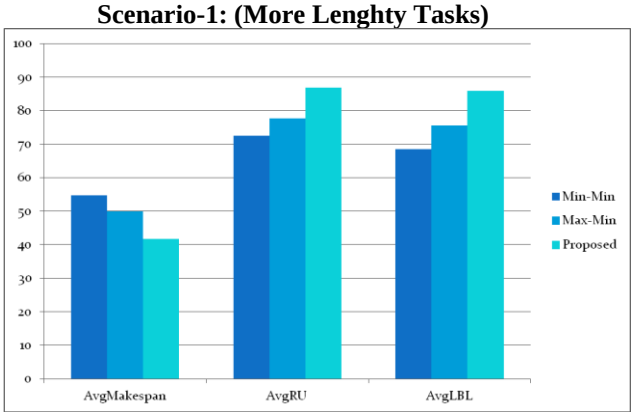
The less the makespan of a scheduling algorithm, the better it works.

B. Average Resource Utilization Rate: It is one of the metrics that is used in [14]. Average resource utilization of each resource can be calculated.

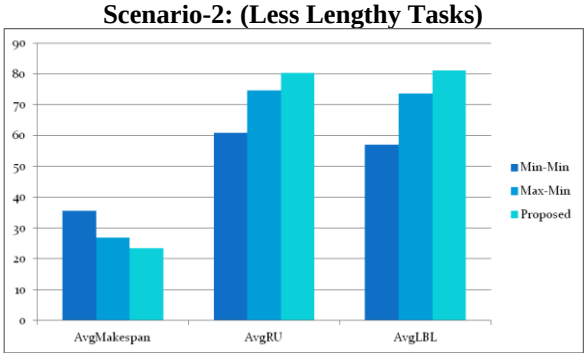
***C. Load Balancing Level*:** The mean square deviation of *ru is defined as :* The best and most efficient load balancing level is achieved if d equals zero and β equals 1. So, scheduling algorithm will have better performance if d is close to 0 and β is close to 1. It is the other metric that is used in [14].

C. Results and Discussion

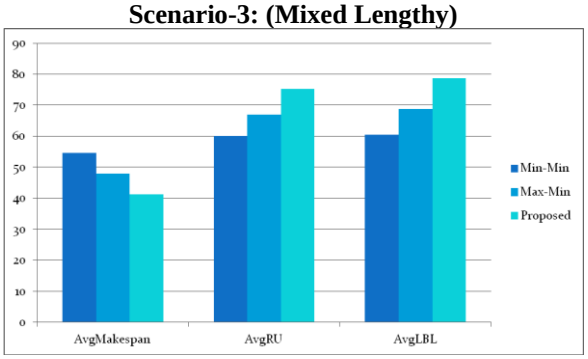
Scenario-1: (More Lenghty Tasks)



Scenario-2: (Less Lengthy Tasks)



Scenario-3: (Mixed Lengthy)



IV. CONCLUSION

In this research work, a new model is developed and analyzed. In this method, three parameters i.e., makespan, average resource utilization and load balancing level, are calculated and the corresponding results are shown. It has been observed that the results generated from proposed model are better than the existing algorithms.

V. ACKNOWLEDGMENT

First of all I would like to thank Dr. Sandip Kr. Goyal for helping me in completing my work. Then I would like to thank the University for providing the required resources on time. Last but not the least, I am grateful to almighty God and my family members for their belief in me.

REFERENCES

[1] Ali, A. 2001.Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. Journal of Empirical finance, 5(3): 221–240.

[2] Basu, S. 1997. The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. Journal of Finance, 33(3): 663-682.

[3] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model.Evidence from KSE-Pakistan.European Journal of Economics, Finance and Administrative Science, 3 (20).