A Research Paper on Efficient Usage of MIPS to Allocate Virtual Machine to Cloudlets

Kuldip Kaur*, Er. Jasdeep Mann
Department of CSE, Bhai Maha Singh College of Engineering, Shri Muktsar Sahib, Punjab, India

DOI: 10.23956/ijarcsse/V7I6/0249

Abstract: Cloud computing is distributed computing, storing, sharing and accessing data over the Internet. It provides a pool of shared resources to the users available on the basis of pay as you go service that means users pay only for those services which are used by him according to their access times. In this paper, we have proposed a new load-balancing algorithm with efficient usage of virtual machines in a cloud environment by computing the utilization power and capacity of each virtual machine. Multiple numbers of experiments have been conducted by taking different configurations of cloudlets and virtual machine. Various parameters like waiting time, processing time and the usage cost have been computed inside the cloud sim environment to demonstrate the results. We will be implementing the proposed work in cloud sim simulation tool

Keywords: Cloud computing, Datacenter, Virtual machine, load balancing, cloudlets.

I. INTRODUCTION

Cloud computing is a combination of many computing fields and has gained much popularity in the recent years. Cloud computing provides computing, storage, services, and applications over the Internet. Moreover, cloud computing facilitates to reduce capital cost, decouple services from the underlying technology, and provides flexibility in terms of resource provisioning. Cloud computing has become very beneficial for business services, applications and other types of consumer requirements. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing has claimed to jump the enterprise business to a brand-new level and permits them to cut back all the prices through improved production, reduced administration and infrastructure, architecture price and quicker preparation cycles. Cloud computing is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In cloud computing, the word cloud(also phrased as "the cloud") is used as a metaphor for "the Internet", so the phrase cloud computing means "a type of Internet-based computing", where different services such as servers, storage and applications are delivered to an organization’s computers and devices through the Internet. The Cloud Computing may be a term that describes the infrastructure, platform, services and different kind of applications. As this is a platform it reconfigures servers or applications where the server can be a physical machines or virtual display machines.

II. RELATED WORK

Al-Rayis et al. [1] explains that basically, load balancers can be deployed based on three different architectures. The centralized load balancing architecture which includes a central load balancer to make the decision for the entire system regarding which cloud resource should take what workload and based on which algorithm(s). Bhoi et al. [2] discussed that in enhanced Max-Min Task Scheduling Algorithm in cloud computing helps in supplying a high performance computing based on protocols which allowed shared computation and storage over long distances. It depends upon expected execution time instead of completion time. Max-Min algorithm assign task with maximum execution time to resource produces minimum completion time while Enhanced Max-min assign task with average execution time to resource produces minimum execution time.

Bendiab et al. [3] introduced the Map Reduced based Entity Resolution load balancing technique in networking which is based on large datasets. In this technique, two main tasks are done: Map task and Reduce task which the author has described. For mapping task, the PART method is executed where the request entity is partitioned into parts. And then COMP method is used to compare the parts and finally similar entities are grouped by GROUP method and by using Reduce task. Map task reads the entities in parallel and process them, so that overloading of the task is reduced.

Buzato et al. [4] proposed Bee Life algorithm which was used for scheduling in Cloud computing. Bee Life algorithm is inspired by the behavior and reproduction of bee to find food source. The algorithm evaluated the performance of the resources and it has the aim to reduce time and complexity of work.

Babu et al. [5] proposed a Honey Bee Behavior inspired Load Balancing [HBB-LB] technique which helps to achieve even load balancing across virtual machine to maximize throughput. It considers the priority of task waiting in queue for execution in virtual machines. After that work load on VM calculated decides whether the system is overloaded, under loaded or balanced. And based on this VMs are grouped. New according to load on VM the task is scheduled on VMs.
Task which is removed earlier. To find the correct low loaded VM for current task, tasks which are removed earlier from over loaded VM are helpful. Forager bee is used as a Scout bee in the next steps.

Dorigo et al. [6] has proposed a load balancing technique called colony of cooperating agents in ants based on soft computing for solving the optimization problem. This technique solves the problem with high probability. It is a simple loop moving in direction of increasing value which is uphill. And this make minor change in to original assignment according to some criteria

Deldari et al. [7] proposed a novel load balancing algorithm called VectorDot in intelligent ants. It handles the hierarchical complexity of the datacenter and multidimensionality of resource loads across servers, network switches, and storage in an agile data center that has integrated server and storage virtualization technologies.

Desai et al. [8] discusses about the emerging technology i.e. a new standard of large scale distributed computing and parallel computing. It provides shared resources, information or other resources as per clients’ requirements at specific times. For better management of available good load balancing techniques are required. And through better load balancing in cloud, performance is increased and user gets better services. So in this author has discussed many different load balancing techniques used to solve the issue in cloud computing environment.

Elzeki et al. [9] discussed in Improved Max-Min Algorithm in Cloud Computing that focuses on the cloud computing which further deals with the allocation of the tasks to the resources while observing different parameters like waiting time, Average waiting time, Turn Around time, processing cost. So, an algorithm named as Max-Min in improved manner from load balancing has been shown to overcome such kinds of problems.

Fahringer et al. [10] introduced a static load balancing technique called Ant Colony Optimization. In this technique, an ant starts the movement as the request is initiated. This technique uses the Ants behavior to collect information of cloud node to assign task to the particular node. In this technique, once the request is initiated, the ant and the pheromone starts the forward movement in the pathway from the “head” node.

Fang et al. [11] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud computing environment.

Gellerb et al. [12] introduced a static well-known load balancing technique called Round Robin, in which all processes are divided amid all available processors. The allocation order of processes is maintained locally which is independent of the allocation from the remote processor. In this technique, the request is sent to the node having least number of connections, and because of this at some point of time, some node may be heavily loaded and other remain idle. This problem is solved by CLBDM (Central Load Balancing Decision Model).

III. OBJECTIVES

The primary purpose of the cloud system is that its client can utilize the resources to have economic benefits. A resource allocation management process is required to avoid underutilization or overutilization of the resources which may affect the services of the cloud.

- To study and evaluate the performance of existing load balancing algorithm.
- To take the heterogeneous Virtual machines of different capacities inside different hosts of multiple datacenters.
- To take into account the MIPS of VMs along with the Utilization Power and capacity of each VM.
- To create the cloudlets of different requirements and services.
- To allocate the virtual machines to the cloudlets by matching their specifications.

IV. PROPOSED WORK

RESEARCH METHODOLOGY

Cloud services provide computing on demand in real time. Number of users accessing cloud environment are always more than that were using it on previous day. Cloud has application areas for developing applications, providing and managing infrastructure, patching applications. Users and their requests for accessing cloud infrastructure are highly dynamic and loading servers running in data center. We need efficient strategy to balance load on these servers so that the servers don’t get crash and they can persist long. Precisely Objective is to achieve accuracy, performance of servers and the cloud environment can be maintained.

1. Initialize the Cloud Sim in Java by creating different data centers having multiple number of hosts of different configurations.
2. Each Host in the datacenter will have the different numbers of Virtual machines of different capacities.
3. Then we will create the Cloudlets of varying length and size of different users.
4. Arrange the virtual machines in descending order of their MIPS (Million Instructions per Second). The virtual machine having more MIPS is more powerful.
5. The HashMap will contain the list of all the Virtual Machines will their status and allocation count. The status will specify whether the virtual machine is busy or idle. The allocation entry in the table will define the total number of tasks submitted to this particular virtual machine.
6. The list containing the Virtual machines and Cloudlets will be given to the Data Center Broker (DCB).
7. Arrange the cloudlets in descending order of their instruction size.
8. DCB will fetch the first cloudlet from the list and will check the requirements of that virtual machine.
9. Fetch that virtual machine which is having more power than another virtual machines and is under loaded.
10. If the multiple virtual machines are satisfying the requirement, then the machine which is available at the first index will be used for executing the given task.
11. Change the status of that virtual machine from idle to busy.
12. Dispatch our cloudlet to that virtual machine and we will modify the rating of that particular virtual machine.
13. Repeat the same procedure for all the cloudlets.

V. EXPERIMENTAL RESULTS

We need efficient strategy to balance load on these servers so that the servers don’t get crash and they can persist long. Precisely Objective is to achieve accuracy, performance of servers and the cloud environment can be maintained. Various experiments have been conducted and the results of existing work and the proposed work have been mentioned in the Table 1 and Table 2 and Table 3.

PARAMETERS USED
I. Total Processing Time
II. Total Processing Cost
III. Total Waiting Time
I) TOTAL PROCESSING TIME:
It is defined as the time interval between the request sent and the response received by the cloud user/consumer. Overall processing time is calculated as given as follow:
Processing Time (PT) = FT - ST
Where,
FT=finish time of execution
ST=start time of execution

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>No. of Cloudlets</th>
<th>Total Processing Time of Existing work</th>
<th>Total Processing Time of Proposed work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3000</td>
<td>1350000</td>
<td>949381</td>
</tr>
<tr>
<td>2</td>
<td>4500</td>
<td>2025000</td>
<td>1424072</td>
</tr>
<tr>
<td>3</td>
<td>5000</td>
<td>2249600</td>
<td>1582019</td>
</tr>
<tr>
<td>4</td>
<td>6500</td>
<td>2924600</td>
<td>2056707</td>
</tr>
<tr>
<td>5</td>
<td>8000</td>
<td>3599600</td>
<td>2531396</td>
</tr>
<tr>
<td>6</td>
<td>9000</td>
<td>4050000</td>
<td>2848140</td>
</tr>
<tr>
<td>7</td>
<td>10000</td>
<td>4499600</td>
<td>3164314</td>
</tr>
</tbody>
</table>

The data, thus collected was used to plot a graph consisting of number of jobs.

![TOTAL PROCESSING TIME](image1)

Figure 2. Total Processing Time

For the above bar chart, it is clear that total processing time has been reduced.

II) TOTAL PROCESSING COST
It is obtained by addition of cost per storage, cost per memory.
Processing cost=RT* unit cost.
Where,
RT=response time
Unit cost=cost per unit time

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>No. of Cloudlets</th>
<th>Total Processing Cost of Existing work</th>
<th>Total Processing Cost of Proposed work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3000</td>
<td>4118850</td>
<td>2896561</td>
</tr>
<tr>
<td>2</td>
<td>4500</td>
<td>6178275</td>
<td>4344844</td>
</tr>
<tr>
<td>3</td>
<td>5000</td>
<td>6863530</td>
<td>4826741</td>
</tr>
<tr>
<td>4</td>
<td>6500</td>
<td>8922955</td>
<td>6275015</td>
</tr>
<tr>
<td>5</td>
<td>8000</td>
<td>10982380</td>
<td>7723289</td>
</tr>
<tr>
<td>6</td>
<td>9000</td>
<td>12356550</td>
<td>8689676</td>
</tr>
<tr>
<td>7</td>
<td>10000</td>
<td>13728280</td>
<td>9654321</td>
</tr>
</tbody>
</table>
For the above bar chart, it is clear that total processing cost has been reduced.

### III) TOTAL WAITING TIME
Waiting time = Allocation Time - generation Time
Fig illustrates that waiting time for the base work and proposed work

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>No. of Cloudlets</th>
<th>Total waiting Time of Existing work</th>
<th>Total waiting Time of Proposed work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3000</td>
<td>80189300</td>
<td>68795098</td>
</tr>
<tr>
<td>2</td>
<td>4500</td>
<td>181033300</td>
<td>155145134</td>
</tr>
<tr>
<td>3</td>
<td>5000</td>
<td>223648100</td>
<td>191590453</td>
</tr>
<tr>
<td>4</td>
<td>6500</td>
<td>378492100</td>
<td>324109701</td>
</tr>
<tr>
<td>5</td>
<td>8000</td>
<td>573836100</td>
<td>491263949</td>
</tr>
<tr>
<td>6</td>
<td>9000</td>
<td>726565300</td>
<td>622003766</td>
</tr>
<tr>
<td>7</td>
<td>10000</td>
<td>897294600</td>
<td>768034814</td>
</tr>
</tbody>
</table>

For the above bar chart, it is clear that total waiting has been reduced.

### VI. CONCLUSION AND FUTURE SCOPE
Cloud computing involves the sharing of resources, software, information via internet with an objective to reduced capital and operational cost, better performance in terms of waiting time and data processing time. In this paper, we have
proposed a new load balancing mechanism by computing the utilization power and capacity of each virtual machine before assigning the task to the virtual machine. The model proposed is based on efficient usage of MIPS to allocate virtual machine to cloudlets. In this research work processing time, processing cost and waiting time has been reduced. The processing cost, Total processing time and waiting time are calculated using various number of experiments. The experiments conducted are compared with existing algorithms. The experimental results are obtained by applying the new planned algorithmic program within the Cloud Sim simulator developed in java programming language. The deadline of each task, arriving rate of tasks, cost of the task execution on each of the resource, cost of communication and many other cases that can be a topic of research are not considered. This work can be further extended by dividing the incoming jobs into different classes according to the job’s nature. The fault tolerance capacities of VM can be added and the jobs can be migrated to another VM/host if the virtual machine fails are also the considerable factors that can be explored in the future work.

REFERENCES