Scheduling Virtual Machines for Load balancing in Cloud Computing Platform

Supreeth S [1], Shobha Biradar [2]

[1, 2]Department of Computer Science and Engineering, Reva Institute of Technology and Management
Yelahanka, Bangalore, Karnataka, India

Abstract*: Cloud computing enables developers to automatically deploy applications during task allocation and storage distribution by using distributed computing technologies in numerous servers. To gain the maximum benefit from cloud computing, developers must design mechanisms that optimize the use of architectural and deployment paradigms. The role of Virtual Machine's (VMs) has emerged as an important issue because, through virtualization technology, it makes cloud computing infrastructures to be scalable. Therefore developing on optimal scheduling of virtual machines is an important issue. In this paper a analysis of different existing Virtual Machine's (VM's) scheduling algorithms are done and proposed a weighted Round Robin algorithm over Round Robin algorithm in Virtual Machine environment of cloud computing in order to achieve better overall response time and processing time. The simulation results show the weighted round robin algorithm shows better improvements over Round-Robin algorithm. Then Comparison between Round Robin and Weighted Round Robin algorithm shows there is a improvement in Weighted Round Robin algorithm.*

Keywords- Scheduling of virtual machines, cloud computing, Weighted Round Robin.

1. Introduction

Cloud computing is the delivery of computer resources through a Web service interface (e.g., SOAP or REST) on an as needed basis. The term "cloud" refers to the organization of the underlying physical infrastructure remaining opaque (not visible) to the end user. In other words, cloud computing gives a user access to computer resources (i.e. machines, storage, operating systems, application development environments, application programs) Over a network through Web services, while the actual physical location and organization of the equipment hosting these resources—be it in the next room or spread across the globe—is not necessarily known to the user. As such, these resources appear to the user as being "in the cloud." The cloud computing will not merely become an enormous data storage, but it can achieve high-performance and high-computing capability. The cloud computing platform guarantees subscribers that it sticks to the service level agreement (SLA) by providing resources as service and by needs based on the broker policy. Cloud computing enables developers to automatically deploy applications during task allocation and storage distribution by using distributed computing technologies in numerous servers [1],[2]. Figure 1 shows the cloud computing architecture.

To gain the maximum benefit from cloud computing, developers must design mechanisms that optimize the use of architectural and deployment paradigms. The role of Virtual Machine's (VMs) has emerged as an important issue because, through virtualization technology, it makes cloud computing infrastructures to be scalable. Therefore developing on optimal scheduling of virtual machines is an important issue. In this paper a analysis of existing Virtual Machine's (VM's) scheduling algorithms are done and proposed a weighted Round Robin algorithm over Round Robin algorithm in Virtual Machine environment of cloud computing in order to achieve better overall response time and processing time. The results shows the weighted round robin algorithm shows better improvements over Round-Robin algorithm by comparing overall response time and Data Centre processing time.

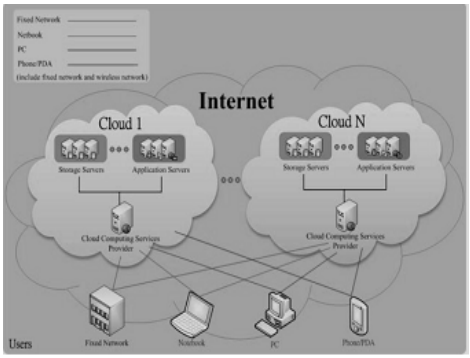

Figure 1: Cloud Computing Architecture

The cloud computing architecture is divided into three layers and it is shown in Figure 2:

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service(SaaS)

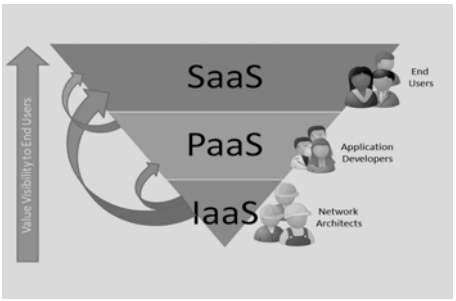

Figure 2: Cloud computing services

- The bottom layer is Infrastructure as a Service (IaaS), which has a service-oriented architecture. Provides access to virtualized computer hardware resources, including machines, network resources, and storage. The most famous service provider is Amazon EC2/S3.

- The middle layer is Platform as a Service (PaaS), a service platform that developers can use to deploy their own applications. It Provides network accessible access to a programming or runtime environment with scalable compute and data structures embedded in It. Well-known PaaS service providers include Amazon Web Services and Google App Engine.
- The top layer is Software as a Service (SaaS), which enables each user to access services according to his or her requirements. Provides network accessible access to software application programs. Examples of SaaS service providers are Microsoft's online update service, Trend Micro Internet Security and so on.

The rest of the paper is organized as follows. In next section Literature Survey about different scheduling algorithms of Virtual machine in cloud are discussed. Section 3 describes Existing Scheduling Algorithm in Cloud Computing. Section 4 discusses The Proposed Scheduling Algorithm. Section 5 discusses Experimental Setup and Result's are analysed. Conclusions are discussed in section 6.

2. Literature Survey

In [3] Dynamic Priority algorithm is discussed. Mainly this scheduling a virtual machine in Eucalyptus platform and it will work under various circumstances. But this algorithm does not handle certain cases because of failure of nodes. Also the uptime and downtime of nodes have not been measured.

In [4] Genetic Algorithm is discussed, in genetic algorithm the problem is the load balancing. So, the strategy for scheduling the VM resources on load balancing is based on the genetic algorithm. According to historical data and the current state of the system through the genetic algorithm, this scheduling strategy computes the needed VM resources after the deployment and chooses the least-affective solution through which it achieves the best load balancing and avoids or reduces the dynamic migration. In the genetic algorithm, the resources are deployed and are arranged to every physical node. By this way it solves the problems. The genetic algorithm introduces an average load distance in order to measure the overall load balancing effect of the algorithm. Virtual machine (VM) migration is used to avoid the conflicts on traditional systems like CPU and memory, micro-architectural resources such as shared caches, memory controllers, and non uniform memory access (NUMA).These relied on intra-system scheduling reduce contentions. In Architectural Shared Resources [5] it shows the live VM migration which is used to mitigate the contentions on micro-architecture resources. This reduces conflicts. It shows the evaluation of two-cluster level virtual machine scheduling techniques for cache sharing and it does not require any prior knowledge on the behaviours of VMs.

In Broker Virtual Machine Communication Framework [6] they have proposed an efficient algorithm to provide an effective and fast execution of the task assigned by the user. So there is an effective communication framework between broker and virtual machine for assigning the task and fetching the results in optimum time and cost using Broker Virtual Machine Communication Framework (BVCF). Prioritizing the VM and cloudlet scheduling through FCFS, and Round Robin, this is implemented over the cloudsim under the VM scheduling policies by the modification which is based on virtual machine cost. The basic building block are scheduling over Virtual Machine as well as over Cloudlets and Retransmission of Cloudlets. Round Robin algorithm helps in the Fast Execution due to Round Robin Scheduling Policy applied on the equally sized cloudlets. All cloudlets will execute as after each and every successfully received cloudlet VM sends the acknowledgement and for the unsuccessful cloudlets sends the retransmit message. It also results into Lower Cost as the VM's are prioritized according to its Cost only. Execution of cloudlets is being analysed over Round Robin and FCFS scheduling policy.

3. Existing Scheduling Algorithm in Cloud Computing

Scheduling in Eucalyptus determines the method by which Virtual Machines are allocated to the nodes. This is done to balance the load on all the nodes effectively and to achieve a target quality of service. The need for a good scheduling algorithm arises from the requirement for it to perform multitasking and multiplexing. The scheduling algorithm in Eucalyptus is concerned mainly with:

- **Throughput** - number of VMs that are successfully allocated per time unit.
- **Response time** - amount of time it takes from when a request was submitted until the first response is produced.
- **Fairness / Waiting Time** – All the requests for an allocation of a node should be treated in the same manner without any bias.

1. Greedy Algorithm*:* The Greedy algorithm is the default algorithm used for scheduling of Virtual Machines in Eucalyptus. The Greedy algorithm [3] is very simple and straight forward. As a matter of fact, it was the only scheduling policy which was in use for a long time. Only after the cloud started evolving, more complex scheduling policies came into effect. The greedy algorithm uses the first node that it finds with suitable resources for running the VM that is to be allocated. The first node that is identified is allocated the VM. This means that the greedy algorithm exhausts a node before it goes on to the next node.

Advantage:

- The main advantage of the Greedy algorithm is its simplicity.
- It is both simple to implement and also the allocation of VMs do not require any complex processing.

Drawback:

- The major drawback would be the low utilization of the available resources.

The drawback of the Greedy algorithm is overcome by the Round Robin algorithm.

2. Round Robin Algorithm: The Round Robin algorithm [3] mainly focuses on distributing the load equally to all the nodes. Using this algorithm, the scheduler allocates one VM to a node in a cyclic manner. The round robin scheduling in the cloud is very similar to the round robin scheduling used in the process scheduling. The scheduler starts with a node

and moves on to the next node, after a VM is assigned to that node. This is repeated until all the nodes have been allocated at least one VM and then the scheduler returns to the first node again. Hence, in this case, the scheduler does not wait for the exhaustion of the resources of a node before moving on to the next.

Advantage:

- The main advantage of this algorithm is that it utilizes all the resources in a balanced order.
- An equal number of VMs are allocated to all the nodes which ensure fairness.

Disadvantage:

- In this method it considers current load on each virtual machine.

4. The Proposed Scheduling Algorithm

The proposed algorithm is weighted round robin algorithm [7] with changes to existing round robin algorithm.

4.1 System Design

The bottom of cloud computing is composed with the virtual machines. When the user catches the large distributed data in cloud computing, the operation will affect the performance of virtual machines. In order to improve the virtual machine's processing in cloud computing infrastructure, this study proposes System for improving the performance of virtual machines in cloud computing infrastructure. To improve the VMs capability to offer reliable services in Cloud computing platform, this study proposes a cloud computing system with different master and slave system architecture.

Figure 3 illustrates the system architecture in which the system consists of cloud controller; cloud controller is the entry-point on system. The CLC is also responsible for managing the underlying virtualized resources like servers, storages and networks. The cloud controller is a collection of services which are grouped by the three categories: resource services, data services and interface services. It can also handle protocol translation and provide public system management.

The Broker policies: This component models the service brokers that handle traffic routing between user bases and data centers. The default routing policy routes traffic to the closest data center in terms of network latency from the source user base. In addition an experimental brokerage policy for peak load to share the load of a data center with other data centers when the original data center's performance degrades above a pre-defined threshold.

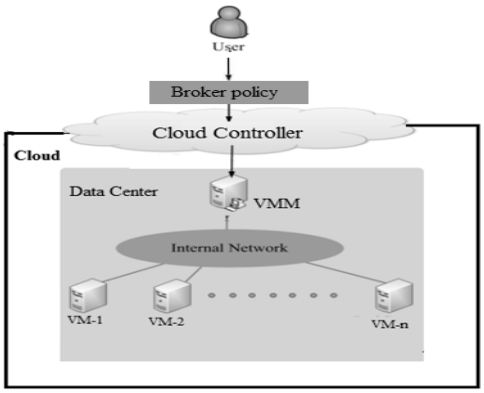

Figure 3: System Architecture

The proposed system which implements weighted round robin method. In this a weighted round robin algorithm, which allocates all incoming requests to the available virtual machines in round robin fashion based on the weight's without considering the current load on each virtual machine.

Steps for scheduling are as follows

Step 1: Master system (VMM) receives information regarding virtual machine from slave (VM-1….n). If the master node capability doesn't catch the data, it will determine the virtual machine to be dead. This study proposed by parameter W.

1. *If* W=0 is set up, it will define the virtual machine to be working and still alive now.
2. If W=1 then node is dead.
3. If W=2 then node is in previous state.

Step 2: If Master node receives the data from slave, then it gets the information's regarding data(memory used, cpu time etc..)

Step 3: Then Master node builds the weighted table containing the details which is collected from step 2.

Step 4: Then the master node sorts(Round-robin method) all the virtual machine's according to their performance. which is $1 \leqq i \leqq N$. Where N is the number of the virtual machines.

Step 5: The scheduling capability generates the weighted table.

Step 6: The virtual machine control capability receives the weighted table from the Step 5, and distributes the task to the virtual machines according to the weighted value.

Algorithm for Scheduling of Virtual Machine

```
j=0, w=0;
for (index =0; index<N; index++)
 {
 j=(j+1) mod N
 if (j==0)
 {
 w = w – gcd (V);
 if (w <= 0)
 {
 w = max(W(V));
 if (w == 0)
 return 0;
 }
 }
 else if (W(Vj) >= w)
 return Vj;
 }
```

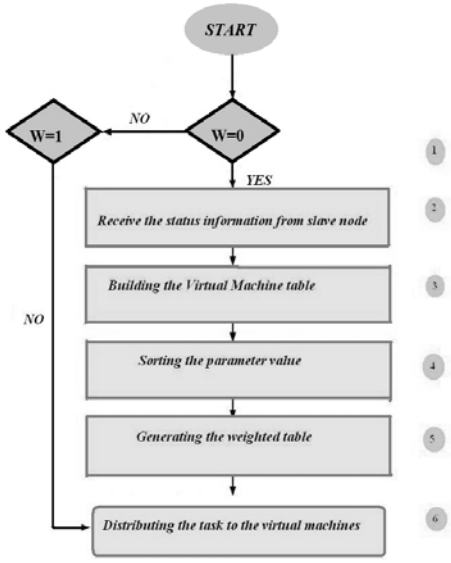

Figure 4: Flowchart of weighted round robin scheduling.

5. Experimental Setup and Result analysis

The proposed algorithm and existing round robin algorithm implemented like graphical simulation. Java language is used for implementing VM load balancing algorithm. Assuming the application is deployed in one data center having virtual machines (with 2048Mb of memory in each VM running on physical processors capable of speeds of 100 MIPS) and Parameter Values are as under

Table 1 discuss the Parameter value's which are used for experiment

Table 1: Parameter Value

Parameter	*Value*
Data Center OS	Linux
VM Memory	2048mb
Data Center Architecture	X86
Service Broker Policy	Optimize Response Time
VM Bandwidth	1000

Table 2 shows the results based on Round robin algorithm for Overall response time of the cloud. In this min(ms) time max(ms) time to different number of virtual machine's are analyzed. Table 3 shows the results based on Round Robin(RR) algorithm for Data Center processing time of the cloud, In this min(ms) time, max(ms) time to different number of virtual machine's are analyzed.

Table 2: For Overall response time for RR

# of VM's	*Avg (ms)*	*Min (ms)*	*Max (ms)*
10	307.94	271.64	346.648
20	304.686	264.265	369.265
30	303.48	240.389	369.39
40	302.304	240.514	369.515
50	301.286	240.639	369.64

Table 3: Data Center processing time for RR

# of VM's	*Avg (ms)*	*Min (ms)*	*Max (ms)*
10	.43	.127	.635
20	.526	.209	.76
30	.604	.309	.885
40	.733	.409	1.01
50	.846	.509	1.135

Table 4 shows the results based on Weighted Round robin algorithm for Overall response time of the cloud. In this min (ms) time max (ms) time to different number of virtual machine's are analyzed. Table 5 shows the results based on Weighted Round Robin(WRR) algorithm for Data Center processing time of the cloud, In this min(ms) time, max(ms) time to different number of virtual machine's are analyzed.

Table 4: For Overall response time for WRR

# of VM's	*Avg (ms)*	*Min (ms)*	*Max (ms)*
10	300.08	237.069	369.14
20	300.2	237.119	369.265
30	300.32	237.169	369.39
40	300.44	237.219	369.515
50	300.56	237.269	369.64

Table 5: Data Center processing time for WRR

# of VM's	*Avg (ms)*	*Min (ms)*	*Max (ms)*
10	.366	.022	.637
20	.486	.034	.762
30	.606	.047	.887
40	.726	.059	1.012
50	.836	.072	1.137

Table 6 shows the compares the results between Round Robin and weighted round robin For Overall response time. Table 7 shows the comparison between Round Robin and Weighted Round Robin For Data Center processing time. Comparison shows Weighted Round Robin method consumes less time for overall response time and Data Center processing time over Round Robin method.

Table 6: Comparison of results between Round Robin and weighted round robin For Overall response time

# of VM's	*Avg (ms)by RR*	*Avg(ms)by weighted RR*
10	307.94	300.08
20	304.686	300.2
30	303.48	300.32
40	302.304	300.44
50	301.286	300.56

Table 7: Comparison of results between Round Robin and Weighted Round Robin For Data Center processing time

# of VM's	*Avg (ms)*	*Avg (ms)*
10	.43	.366
20	.526	.486
30	.604	.606
40	.733	.726
50	.846	.836

The above experimental results show the weighted round robin method consumes less time for responding over round robin method.

6. Conclusion

A Virtual Machine is an abstraction of computer hardware within software. Virtual machine executes programs as if they were actual physical machines. In this paper it gives the detailed review on existing scheduling algorithms with their advantages and drawbacks. The proposed weighted round-robin scheduling method and existing round robin algorithm implemented Java language for implementing VM scheduling algorithm. Assuming the application is deployed in one data center having virtual machines (with 2048Mb of memory in each VM running on physical processors capable of speeds of 1000MIPS). These experimental results shows that weighted round robin method improves the performance by consuming less time for scheduling virtual machines.

References

[1] M.D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis and A. Vakali, ”Cloud Computing: Distributed Internet Computing for IT and Scientific Research,” IEEE Internet Computing, Vol.13, No.5, pp.10-13, 2009.220.

[2] B. Ahlgren, P.A. Aranda, P. Chemouil, S. Oueslati, L.M. Correia, H. Karl, M. Sollner and A. Welin, ”Content, Connectivity and Cloud: Ingredients for the Network of the Future,” IEEE Communications Magazine, Vol.49, No pp.62-70, 2011.

[3] Subramanian S, Nitish Krishna G, Kiran Kumar M, Sreesh P4and G R Karpagam, “An Adaptive Algorithm For Dynamic Priority Based Virtual Machine Scheduling In Cloud” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

[4] Jianhua Gu, Jinhua Hu, Tianhai Zhao, Guofei Sun, “A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment”, Journal Of Computers, Vol. 7, No. 1, January 2012.

[5] Jeongseob Ahn, Changdae Kim, Jaeung Han, Young-ri Choi, and Jaehyuk Huh, “Dynamic Virtual Machine Scheduling in Clouds for Architectural Shared Resources”,

[6] Gaurav Raj and Sonika Setia, ”Effective Cost Mechanism for Cloudlet Retransmission and Prioritized VM Scheduling Mechanism over Broker Virtual Machine Communication Framework”, International Journal on Cloud Computing: Services and Architecture(IJCCSA),Vol.2, No.3, June 2012.

[7] Jiann-Liang Chen, Yanuarius Teofilus Larosa and Pei-Jia Yang, “Optimal QoS Load Balancing Mechanism for Virtual Machines Scheduling in Eucalyptus Cloud Computing Platform”, 2012 2nd Baltic Congress on Future Internet Communications.

Authors Profile

Supreeth S received B.E. degree in Computer science and engineering in 2011 from SJCIT, Chikkaballapura. Currently he is pursuing his MTech degree at Department of Computer Science and Engineering from Reva Institute of Technology and Management, Visvesvaraya Technological University (VTU), Bangalore, Karnataka. His research interests include Cloud computing and its applications, Image processing and real-time applications.

Shobha Biradar received B E and MTech degree in computer science and engineering from Visvesvaraya Technological University (VTU), Bangalore, Karnataka. She is currently working as a assistant professor at Reva Institute of Technology and Management, Bangalore. Her research interests include Cloud computing and its applications, Storage area networks.