Load Balancing in Cloud Computing System

Rashmi Sharma and Abhishek Kumar

Department of CSE, ABES Engineering College, Ghaziabad, Uttar Pradesh, India

E-mail: abhishek221196@gmail.com

(Received on 10 August 2012 and accepted on 15 October 2012)

***Abstract* - The conception of Cloud computing has not only reshaped the field of distributed systems but also fundamentally changed how businesses potential extend today. Load balancing is a core and challenging issue in Cloud Computing. How to use Cloud computing resources efficiently and gain the maximum profits with efficient load balancing algorithm is one of the Cloud computing service providers' ultimate goals. In this paper firstly a analysis of different Virtual Machine (VM) load balancing algorithms was done a new VM load balancing algorithm has been proposed and Implemented in Virtual Machine environment of cloud computing in order to achieve better response time.**

Keywords : Virtual Machine, Load Balancing, Cloudsim

I. Introduction

Cloud computing is a fast rising area in computing research and industry today. It has the potential to make the new idea of 'computing as a utility' in the near future. The Internet is often represented as a cloud and the term "cloud computing" arises from that analogy. Cloud computing is the dynamic provisioning of IT capabilities (hardware, software, or services) from third parties over a network [7]. It is generally supposed that there are three basic types of cloud computing: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [1]. In IaaS grids or clusters, virtualized servers, memory, networks, storage and systems software are delivered as a service. Perhaps the best known example is Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3), IaaS Provide access to computational resources, i.e. CPUs. And also provide (managed and scalable) resources as services to the user [7]. PaaS typically makes use of dedicated APIs to control the behavior of a server hosting engine which executes and replicates the execution according to user requests. Example: Force.com, Google App Engine. Software as a Service (SaaS) Standard application software functionality is offered within a cloud. Examples: Google Docs, SAP Business by design. Load balancing is one of prerequisites to utilize the full resources of parallel and distributed systems. Load balancing mechanisms can be broadly categorized as centralized or decentralized, dynamic or static, and periodic or non-periodic. Physical resources can be split into a number of logical slices called Virtual Machines (VMs). All VM load balancing methods are designed to determine which Virtual Machine assigned to the next cloudlet [11]. This document introduce a new VM load balancing algorithm and compare the performance of this algorithms with the already existing algorithms like Round robin and Equally spreaded current execution Load VM load balancer [11]. Section III introduces the problem formation, section IV includes the proposed algorithm of the problem section V includes the experimental setup section VI shows the result and the last section VII includes the conclusion.

II. Existing VM Load Balancer

Virtual machine enables the abstraction of an OS and Application running on it from the hardware. The interior hardware infrastructure services interrelated to the Clouds are modeled in the simulator by a Datacenter element for handling service requests. These requests are application elements sandboxed within VMs, which need to be allocated a share of processing power on Datacenter's host components. Datacenter object manages the data center management activities such as VM creation and destruction and does the routing of user requests received from User Bases via the Internet to the VMs The Data Center Controller [11], uses a VmLoadBalancer to determine which VM should be assigned the next request for processing. Most common VmLoadBalancer are Round Robin and Equally spreaded current execution load balancing algorithms.

A. Round Robin

It is one of the simplest scheduling techniques that utilize the principle of time slices. Here the time is divided into multiple slices and each node is given a particular time slice or time interval i.e. it utilizes the principle of time scheduling. Each node is given a quantum and in this quantum the node will perform its operations. The resources of the service provider are provided to the requesting client on the basis of this time slice. Though the algorithm is very simple but there is an additional load on the scheduler to decide the size of quantum. It is shown in fig. 1.

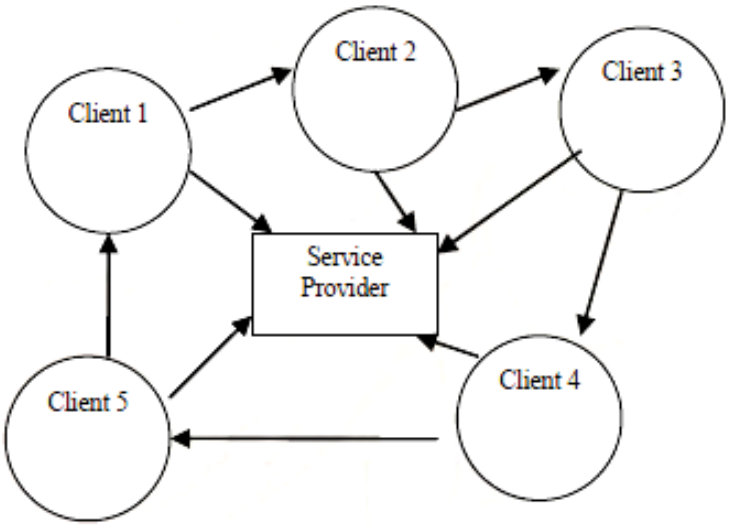

Fig. 1 Round Robin Algorithm

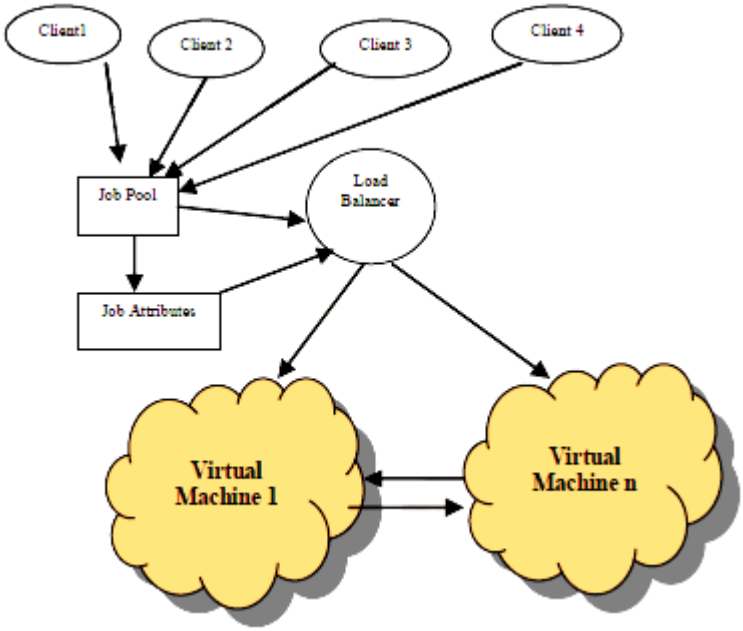

Fig. 2 ESCE Algorithm

B. Equally Spread Current Execution Load

It is spread spectrum technique in which the load balancer spread the load of the job in hand into multiple virtual machines. The load balancer maintains a queue of the jobs that need to use and are currently using the services of the virtual machine. The balancer then continuously scans this queue and the list of virtual machines. If there is a VM available that can handle request of the node/client, the VM is allocated to that request. If however there is a VM that is free and there is another VM that needs to be freed of the load, then the balancer distributes some of the tasks of that VM to the free one so as to reduce the overhead of the former VM. Figure2. better explains the working of the ESCE algorithm. The jobs are submitted to the VM manager, the load also maintains a list of the jobs, their size and the resources requested. The balancer selects the job that matches the criteria for execution at the present time. Though there algorithm offers better

results as shown in further section, it however requires a lot of computational overhead.

III. Problem Formulation

In this paper a study of various load balancing algorithms in cloud computing was done. The algorithms are round robin, ESCEL load balancer. A new algorithm has been proposed from modifying these loads balancing algorithm in Virtual Machine environment of cloud computing in order to achieve better response time, and processing time.

IV. Proposed VM Load balancing Algorithm

The Proposed Load balancing algorithm is divided into three parts. The first phase is the initialization phase. In the first phase, the expected response time of each VM is to be found. In second Phase find the efficient VM, in Last Phase return the ID of efficient VM.

1. This VM Load Balancer find expected response time of each Virtual machine.
2. When a request to allocate a new VM from the Data Center Controller arrives, A VM Load Balancer find the most efficient VM (efficient VM having least loaded, minimum expected response time) for allocation.
3. The VM Load Balancer will return the id of the efficient VM to the Datacenter Controller.
4. Datacenter Controller notifies the new allocation.
5. Proposed VM Load Balancer updates the allocation table increasing the allocations count for That VM.
6. When the VM finishes processing the request and the Datacenter Controller receives the Response. Datacenter controller notifies the VM Load Balancer for the VM de-allocation.

The proposed algorithm find the expected Response Time of each Virtual Machine at the Datacenter controller because virtual machine can be of heterogeneous platform, the expected response time can be find with the help of the following formulas.

$$\text{ResponseTime} = \text{Fint} - \text{Arrt} \quad (1)$$

Where, Arrt is the arrival time of user request and Fint is the finish time of user request after servicing the request at datacenter the result will be transmitted at the requested UserBase. So the transmission delay can be determined using the following formulas

$$\text{TDelay} = \text{Tlatency} + \text{ResponseTime} \quad (2)$$

Where, TDelay is the transmission delay Tlatency is the network latency (Round Trip) time taken to transfer the size of data of a single request (D) from source location to destination and destination to source and ResponseTime is the time taken to service the request at the datacenter.

V. Experimental Setup

The proposed algorithm implemented through simulation packages like CloudSim and cloudsim based tool [11]. Java language is used for implementing VM load balancing algorithm. Assuming the application is deployed in one data center having 50 virtual machines (with 1024Mb of memory in each VM running on physical processors capable of speeds of 100 MIPS) and Parameter Values are as under.

Table I Parameter Value

Parameter	Value
Data Center OS	Linux
VM Memory	1024MB
Datacenter Architecture	X86
Service Broker Policy	Optimize response Time
VM Bandwidth	1000

VI. Results

After performing the simulation the result computed by cloud analyst is as shown in the following figures. We have used the above defined configuration for each load balancing policy one by one and depending on that the result calculated for the metrics like response time, request processing time in fulfilling the request has been shown. overall response time calculated by the cloud analyst for each loading policy has been shown in the figure 3, 4 and 5 respectively. As can be seen from the figure the overall response time of Round Robin policy and ESCEL policy is almost same while that of the proposed VM Load Balancing Policy is very much low as compared to other two policies.

Overall Response Time Summary

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	123.85	50.17	350.85
Data Center processing time:	32.84	0.23	76.35

Fig. 3 Response time using Round Robin Policy.

Overall Response Time Summary

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	123.98	50.66	350.85
Data Center processing time:	32.90	0.23	74.42

Fig. 4 Response time using ESCEL Policy.

Overall Response Time Summary

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	109.54	50.49	350.85
Data Center processing time:	16.84	0.23	65.89

Fig. 5 Response time using VM Load Policy.

VII. Conclusion

In this paper a new VM load balancing algorithm was proposed and then implements in CloudSim cloud computing environment using java language. Proposed algorithm find the Expected response time of each resource (VM) and Send the ID of virtual machine having minimum response time to the data center controller for allocation to the new request, according to this experiment we conclude that if we select a efficient virtual machine then it effect the overall performance of the cloud Environment and also decrease the average response time is decrease.

References

[1] Cary Landis, Dan Blacharski,"Cloud Computing Made Easy", Version 0.3

[2] Ioannis Psoroulas, Ioannis Anagnostopoulos, Vassili Loumos, Eleftherios Kayafas,"A Study of the Parameters Concerning Load Balancing Algorithms", *IJCSNS International Journal of Computer Science and Network Security,* Vol. 7, No. 4, 2007, pp. 202-214 .

[3] Sandeep Sharma, Sarabjit Singh, Meenakshi Sharma "Performance Analysis of Load Balancing Algorithms", *World Academy of Science, Engineering and Technology,* 38, 2008 pp. 269- 272.

[4] Luqun Li,"An Optimistic Differentiated Service Job Scheduling System for Cloud Computing Service Users and Providers", *Third International Conference on Multimedia and Ubiquitous Engineering, IEEE Explore,* 4-6, June 2009, pp. 295-299.

[5] Rodrigo N. Calheiros, Rajiv Ranjan, César A. F. De Rose, Rajkumar Buyya,"Modeling and Simulation of Scalable Cloud Computing Environment and CloudSim Toolkit:Challenges and opportunities", *IEEE Explore International Conference on high performance Computing and simulation,* 21-24 June 2009, pp. 1-11.

[6] QI CAO, ZHI-BO WEI, WEN-MAO GONG,"An Optimized Algorithm for Task Scheduling Based On Activity Based Costing in Cloud Computing". *International Conference on Bioinformatics and Biomedical Engineering,IEEE Explore* 14 July 2009, pp. 1-3.

[7] Bhasker Prasad Rimal, Eummi Choi, Lan Lump "A Taxonomy and Survey of Cloud Computing System", *5th International Joint Conference on INC, IMS and IDC, IEEE Explore* 25-27 Aug 2009, pp. 44-51.

[8] Yi Zhao, Wenlong Huang,"Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud", *IEEE 5th International Joint Conference on INC,IMS and IDC,* 25-27 Aug 2009, pp. 170-176

[9] Huai Zhang, Shufen Zhang, Xuebin Chen, Xiuzhen Huo,"Cloud Computing Research and Development Trend" *Second International Conference on Future Networks,* 22-24 Jan 2010, pp. 93-97.

[10] Martin Randles, David Lamb, A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", *IEEE 24th International Conference on Advanced Information Networking and Applications Workshops,* 20-23, April 2010, pp. 551-556.

[11] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, Rajkumar Buyya,"CloudAnalyst: A CloudSim-based Visual modeler For analyzing Cloud Computing Environments and Applications", 20-23, April 2010, pp. 446-452