Weighted VM Load Balancer in Cloud Environment

Sushma Rani[1] Pardeep Kumar[2] Shalini Aggarwal[3]

[1]M.Tech Student,Department of Computer Science and Applications, Kurukshetra University Kurukshetra, Haryana, India
[2]Associate Professor, Department of Computer Science and Applications, Kurukshetra University Kurukshetra, Haryana, India
[3]Teacher Fellow, Department of Computer Science and Applications, Kurukshetra University Kurukshetra, Haryana, India

ABSTRACT

Cloud computing is the evolution in the IT world where the end users provides the computing resources and services on pay per use model. It provides the virtual environment to the simultaneous access of multiple users through virtualization which is the process of maintaining and creating the virtual machines. Cloud is made up of large number of datacenters on which virtual machines are created that are able to handle the large number of users but as the number of users are increasing day by day there is a need to balance the load. To distribute the load evenly among virtual machines various techniques are weighted which minimizes the response time, cost, optimize resource use, bottlenecks and hence maximize performance. In this paper the weighted VM load balancer technique has been discussed which minimizes the response time and have better resource utilization of virtual machines.

Keywords: *Cloud computing, Cloud analyst, Dispatcher, Weight, Load balancer.*

1. INTRODUCTION

Cloud computing using the internet provides the distributed computing to the multiple users with distributed access to scalable virtualized software and hardware infrastructure. Cloud computing is simply packaging of computing resources such as computation, services and storage as metered service. Cloud computing refers to the deliverance of computing resources over the Internet. Instead of keeping data on your own hard drive or updating applications time-to-time according to our needs. One can use a service over the Internet, at another location, to store our information and use its updated applications. Cloud computing is the delivery of computing services over the Internet. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. Examples of cloud services include online file storage, social networking sites, webmail, and online business applications. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available.

Services that are provides by the cloud computing are:
a) Infrastructure as a service (IaaS): User doesn't need to procure the infrastructure and maintain it. The entire infrastructure can be taken care of by the cloud itself in form of virtual machines.
b) Software as a service (SaaS): In this service user not need to purchase the software and install it he can directly use it from cloud provider e.g. Google+, Gmail.
c) Platform as a service (PaaS): It provides a platform where user can run and deploy their applications.
d) Human as a service (HaaS): Here the human is given some task to do the work e.g voting, rating.

Characteristics of cloud computing:

- On-demand self-service: Computing resources are used automatically by the cloud without the human intervention whenever needed by any task
- Broad network access: Using the large network connectivity computing resources can be shared with the different client applications and heterogeneous platforms.

- Resource pooling: Computing resources are 'pooled' together to serve multiple consumers using either the multi-tenancy (multiple users can access at same time) or the virtualization model.
- Rapid elasticity: Computing resources can be scaled up whenever client need.
- Measure Service: Cloud has the mechanism of measuring the usage of computing resource that is consumed by the user through its metering service.

Organizations can make their own cloud according to their need. Cloud deployment model is divided in to four categories viz. public, private, hybrid and community. Public cloud can be accessible by the all clients but the private cloud is accessible by only the authorized clients of one organization. Hybrid cloud is the combination of private and public cloud. Community cloud is only accessible by the particular community of organizations like doctors, teachers community etc [10]. As Cloud computing data is increasing day by day therefore to maintain it we need balancer. In the section 3, weighted balancer is discussed.

2. LITERATURE REVIEW

Cloud computing is based on the concept of virtualization. Virtualization is the process of creating and maintaining virtual machines on physical servers. Cloud is made up of large number of datacenters on which virtual machines are created that are able to handle the large number of users but as the number of users increasing day by day there is need for balancing the load. Load balancing is the mechanism that decides which requesting nodes/client will use the virtual machine and which requesting machines will be put on hold. Load balancing can be done in dividually as well as on grouping basis. For balancing the load various load balancing have been weighted such as round robin, throttled, active monitoring, CLBVM etc. Y. Lua et al. [15] weighted a load balancing algorithm Join-Idle-Queue in which distributed dispatcher makes the queue of idle processors so that no communication overhead occurs. M. Ajit et al.[7] presents Weighted Signature based load balancing (WSLB) algorithm to minimize user response time. It proposed the VM load balancing algorithm at the VM level where, individual task is assigned to VM mapped on host with different computing power. Poonam devi et al. [8] addressed a Shortest job scheduling algorithm it follows the approach shortest job complete the execution first to utilize the resources in completion of heavy jobs. Gulsan soni et al. [3] recommended clbvm which improve the performance of the system but fault tolerance problem is not resolved in distributed environment. Amandeep kaur et al. [1] introduced Round robin algorithm in this request assigned to the VMs on rotating manner. Weikun Wang et al.[9] define a Weighted round robin algorithm in it the weights are assigned to the each VM which improves the performance. Tanvee Ahmed et al.[13] weighted Equally spread current execution algorithm in which load is distributed evenly among the VM's. Jasmine James et al. [5] define a Throttled load balancing algorithm that maintains a list of virtual states i.e. busy/idle. Hemant S. Mahalle et al. [4] suggested Active monitoring load balancing algorithm that maintains a list contain the number of request currently allocated to VMs request assigned to least loaded VM . Meenakshi Sharma et al.[6] recommended Efficient Response Time Load Balancer algorithm is based on lower response time for the allocation of the virtual machine. Reena Panwar et al. [11] define dynamic load management algorithm in which incoming request is distributes among the virtual machines efficiently and resource utilization is proper. Shridhar G. Domanal et al.[12] proposed Modified throttled algorithm maintains an index table of virtual machines and also the state of VMs similar to the Throttled algorithm. There has been an attempt made to improve the response time and achieve efficient usage of available virtual machines.

3. PROPOSED WORK

The proposed weighted load balancing algorithm will improve the resource utilization virtual machines using the dispatchers that are maintaining the idle virtual machines in the queue. In the weighted algorithm user bases send the request to the data center controller. Data center controller then send the request to the WeightedLoadBalancer which maintain a list of dispatcher. Dispatcher which has the maximum queue length is selected. If there is no element in the queue then VM is select on the basis of weight factor. Weighted algorithm calculates the weight using the bandwidth, memory and number of CPU.

W[i] = (BW[i]*Memory[i]*CPU[i])/Queuelength
where,
i = id of virtual machine

W[i] = weight of virtual machine

BW[i] = bandwidth of virtual machine

Memory[i] = memory of virtual machine

CPU[i] = number of CPU

Queuelength = number of element in the queue

WEIGHTED ALGORITHM:

1. WeightedLoadBalancer maintains a list contain dispatcher which maintain the idle virtual machines queue. (Initially all VM are available)
2. Data center controller receives a new request
3. Data Center Controller queries the weighted Load balancer for next allocation.
4. Weighted put the VM in the queue as

a) If (Policy==SQ)
VM put in the queue with minimum size
else
put randomly

5. WeightedLoadBalancer pick the dispatcher with maximum queue length

a) If (size==0)
VM having the maximum weight picked up and increase the counter
else
return the first idle

6. If (counter ==1)
pick randomly
else
return VM which has maximum weight

7. WeightedLoadBalancer returns the VM id
8. Datacenter controller gets the VM id and assigns the request to it and update the status.
9. When the VM finish the processing of request it informs the datacenter controller for deallocation and update the status.

Working of the above algorithm can also be shown using following flowchart:

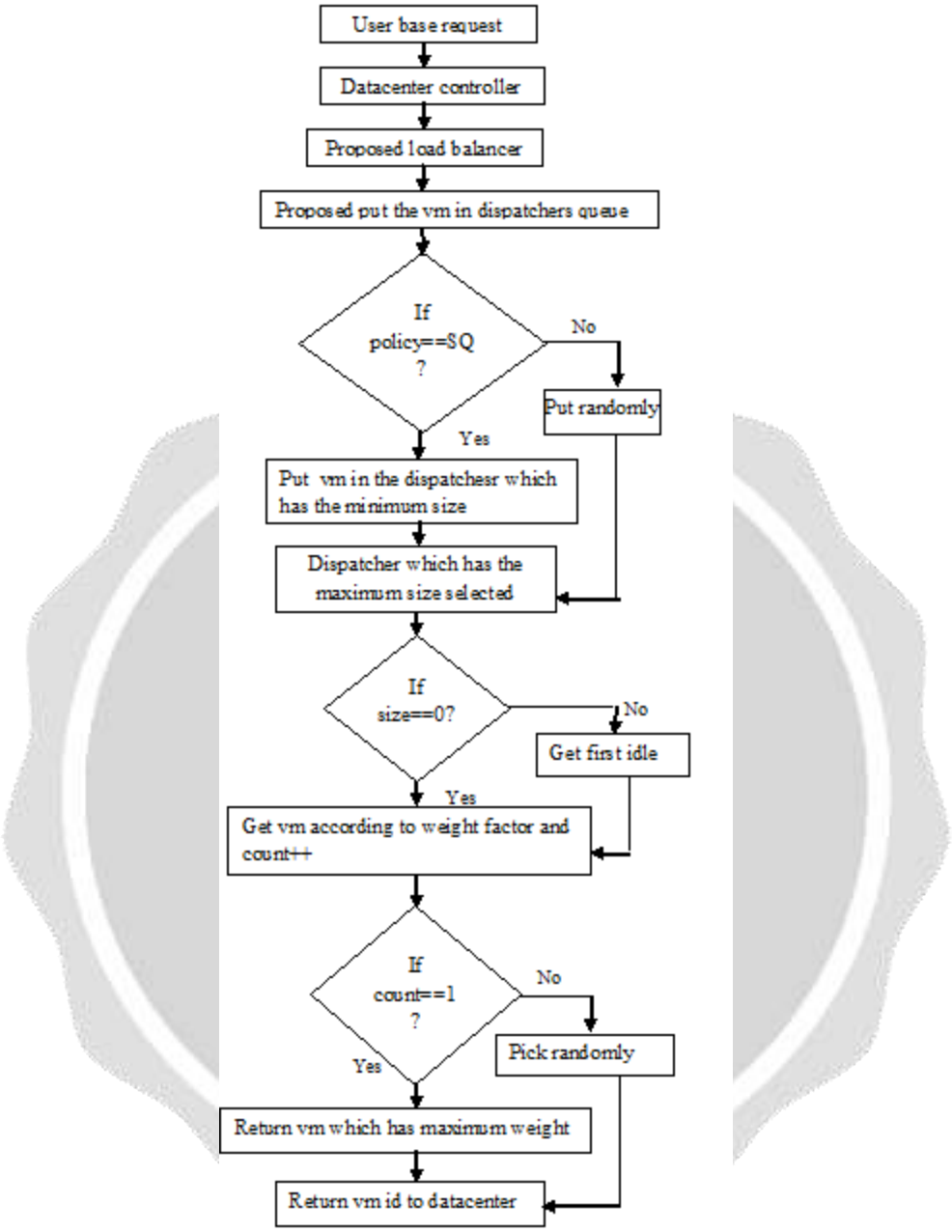

Figure 1: Flowchart for Weighted algorithm

4. EXPERIMENTAL SETUP

The weighted algorithm is implemented on Cloud Analyst tool [14].
User base table: It represents the user requests that are coming from different regions.

Table: 1 Define 6 user bases

Name	Region	Request per User Per Hr	Data Size per Request (Bytes)	Peak Hrs start (GMT)	Peak Hrs End (GMT)	Avg Peak Users	Avg off-Peak Users
UB1	0	60	100	14	24	100000	10000
UB2	1	60	100	15	23	200000	20000
UB3	2	60	100	16	22	300000	30000
UB4	3	60	100	17	21	250000	25000
UB5	4	60	100	13	20	150000	15000
UB6	5	60	100	12	18	240000	24000

Service broker policy: Optimize response time

Table: 2 Parameter used

Parameter	Value used
VM Image Size	10000
VM Memory	512 Mb
VM Bandwidth	1000
Data Center – Architecture	X86
Data Center- OS	Linux
Data Center- VMM	Xen
Data Center- Number of Machines	2
Data Center- Memory for Machines 1	204800
Data Center- Memory for Machines 2	104800
Data Center- Storage for Machine1	100000000
Data Center- Storage for Machine2	400000000
Data Center- Available BW per Machine	1000000
Data Center- Number of Processors for machine 1	4
Data Center- Number of Processors for machine 2	2
Data Center- Processor Speed for machine 1	10000 MIPS
Data Center- Processor Speed for machine 2	20000 MIPS
Data Center- VM Policy	Time shared
User Grouping factor	1000
Request Grouping Factor	100
Execution Instruction Length	250

5. SIMULATION RESULT

The weighted algorithm is tested on the data sets given in section IV. Comparison of the weighted algorithm with the existing [4] algorithm has been shown in the following figures:

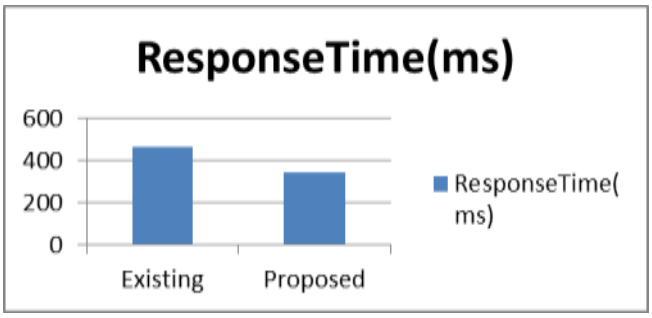

Chart 1: Comparison of Response Time

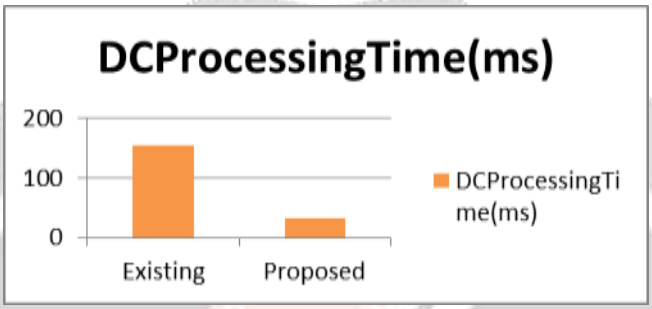

Chart 2: Comparison of Processing Time

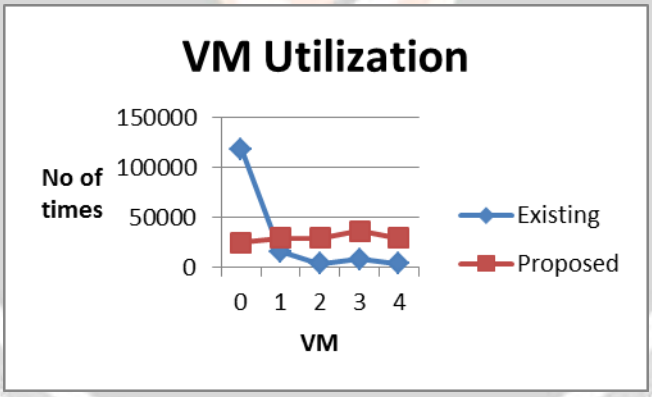

Chart 3: Comparison of VM Utilization

6. CONCLUSION AND FUTURE WORK

In this paper, a new algorithm for load balancing has been proposed, which basically balance the load among various VM, available. It has been identified that the new weighted algorithm performs far better than the currently existing algorithm [3] in terms of better response time, better data center processing time and better VM utilization. This has been proved after performing experiment using cloud analyst tool .Therefore it can be concluded that the proposed algorithm can balance the load in better way and proved to be a beneficially in cloud computing. So it can be used for balancing the load to get the better results. In future work weighted algorithm can be tested on heterogeneous environment and virtual machine migration is another issue that also can be considered.

7. REFERENCES

[1] Amandeep Kaur Sidhu and Supriya Kinger, "Analysis of Load Balancing Techniques in Cloud Computing", International Journal of Computers & Technology, Volume 4 No. 2, pp: 737-741, March-April, 2013.
[2] Bhasker Prasad rimal, eunmi choic and lan lump "A Taxonomy and Survey of Cloud Computing", IEEE, pp: 44-51,2009

[3] Gulshan Soni and Mala Kalra,"A Novel Approach for Load Balancing in Cloud Data Center", International Advance Computing Conference (IACC) IEEE, pp.807-812, 2014.
[4] Hemant S. Mahalle, Parag R. Kaveri and Vinay Chavan, "Load Balancing on Cloud Data Centers", International Journal of Advanced Research in Computer Science Load and Software Engineering, Volume3, issue1, January2013.
[5] Jasmin James and Dr. Bhupendra Verma, "Efficient VM Balancing Algorithm for a Cloud Computing Environment", International Journal on Computer Science and Engineering (IJCSE), Volume 4 No., pp.1658-1663, 09 Sep2012.
[6] Meenakshi Sharma, Pankaj Sharma and Dr. Sandeep Sharma, "Efficient Load Balancing Algorithm in VM Cloud Environment", International Journal on Computer Science and Technology, volume3,pp.439-44I,Jan-March-2012.
[7] M. Ajit and G. Vidya, "VM Level Load Balancing in Cloud Environment", 4th ICCCNT 2013, Tiruchengode, India, July 4-6, 2013.
[8] Poonam devi and Mr.Trilok Gaba "Imlementation of cloud computing by shortest job scheduling", International journal of advanced research in computer science and software engineering.
[9] Weikun Wang and Giuliano Casale, "Evaluating Weighted Round Robin Load Balancing for Cloud Web Services", April 2010.
[10] Qi Zhang, Lu Cheng and Raouf Boutaba, "Cloud computing: sate-of-art and research challenges", Published online: 20th April 2010, Copyright: The Brazillian Computer Society 2010.
[11] Reena Panwar and Dr.Bhawna Mallick, "Load balancing in cloud computing using dynamic load management algorithm" International Conference on Green Computing and Internet of Things (ICGCloT) ieee , june 2015.
[12] Shridhar G. Domanal and G. Ram Mohana Reddy, "Load Balancing in Cloud Computing Using Modified Throttled Algorithm", IEEE, International conference CCEM 2013.
[13] Tanvee Ahmed and Yogendra Singh, "Analytic Study of Load Balncing Techniques Using Tool Cloud Analyst", International Journal of Engineering Research and Applications, 2012, pp.1027-1030.
[14] W. Bhathiya, "CloudAnalyst a CloudSim-based tool for modelling and analysis of large scale cloud computing environments", MEDC Project, Cloud Computing and Distributed Systems Laboratory, University of Melbourne, Australia, pp. 1-44, June 2009.
[15] Y. Lua, Q. Xiea, G. Kliotb,A. Gellerb, J. R. Larusb, and A. Greenber, "Join-Idle-Queue: A novel load balancing algorithm for dynamically Scalable web services", An international Journal on Performance evaluation, August 2011.