International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064

An Approach for the Load Balancing in Cloud Based on the Dynamic Threshold

Pavan Gurve[1], Aishwarya Mishra[2]

[1]M.Tech student, Department of CSE, IES Institute of Technology and Management, Bhopal, India

[2]Associate Professor, Department of CSE, IES Collage of Technology, Bhopal, India

Abstract: ***Cloud computing offer IT services to the users worldwide on the basis of pay-as-you-go model. Now a day cloud is so famous because of their attractive features such as universal accessibility, availability, cheap, pay as you go etc. Cloud is a business model, where the cloud provider, gives the resources on rent basis. Cloud provider wants to maximum profit, which can be achieved by the proper resource utilization and reduction in energy consumption. Energy consumption is depends on the number active server [1] says that a host consume 70% (of their total energy consume when they are fully utilize) even when the host is completely idle. So energy consumption can be reduced by minimizing the number of active server. Load balancing is one of the critical issue in the cloud, because it affect the user performance. In this paper we proposed a adaptive threshold based load balancing approach which increase the resource utilization and reduce the energy consumption. This approach uses two thresholds that are upper and lower threshold for the load balancing. Load of the physical machine (PM) must be within the lower and upper threshold rang. When the load on the host is above/below the threshold, some VM will be migrated to balance the load on the host.***

Keywords: Virtual machines, Energy consumption, server consolidation, load balancing, VM migration.

1. Introduction

Cloud is the fastest growing technology in IT industry because of their attractive features [2, 3]. 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 [4]. It can be deploy in three different way i.e. private, public and protected and support three type of services Software as Services (SaaS), Platform as Services (PaaS) and Infrastructure as Services (IaaS) [4, 5]. Cloud computing supports on-demand service in which shared resources (i.e. CPU, memory and bandwidth), information, software, hardware, network and other devices are provided according to the client requirement. Main advantage of the cloud services is that these services can be access anywhere in the world at any time through the internet and pay as you used.

Virtualization [6, 7, 8] is the key technology in the cloud computing. Virtualization is the technology which allows the sharing of resources by dividing the physical resource. It is implemented through the hypervisor also known as virtual machine monitor which is small software reside between the hardware and guest OS. When the user request for the resources, hypervisor create the VM according to the user requirement, bind the service to the VM and assign to the user. One or more VM can be assign to the one user and each VM can run one or more application. Virtualization can be done in two different ways i.e. Full Virtualization and Para Virtualization. In full virtualization technique complete installation of one machine is run on another. The result is a system in which all software running on the server is within a virtual machine (VM). In this virtualization guest OS not need to be modify. But it slow compare to the paravirtualization. While in paravirtualization only its management module operates with an operating system that has been adjusted to work in a virtual machine. Paravirtualization typically runs better than the full virtualization model, simply because in a fully virtualized deployment, all elements must be emulated. VM migration is a core function provided by the virtualization.

Virtual machine migration [9, 10] can be defined as the movement of VM from one host to another host as show in figure-1 where one virtual machine (VM) move from one host to another host. There are number of situation in the real world where the VM are required such as load balancing, server consolidation, hot spot mitigation, host maintenance, server failure etc.

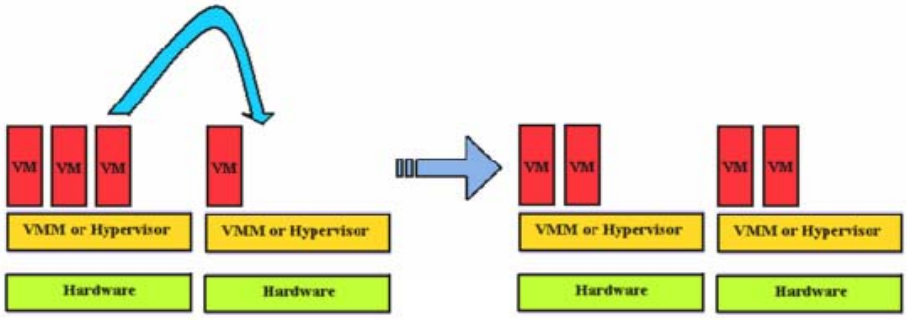


Figure 1: Virtual Machine migration

Load balancing- When the load on the host is greater than upper threshold, host call to be a overloaded. So to balance the load some VM are migrated from overloaded host to the other host.

Server consolidation:- A host call to be a underloaded when the load on the host is less than the lower threshold. In this case all the VM running onto that host must be migrated to the other host. So it will reduce the number of active server.
Hot spot mitigation:- When the VM running into the host, there may be a situation where the resource required by the VM are not fulfill the PM in which VM are running. This situation is known as hot spot. To mitigate the hot spot some VM have to be migrated.

For the effective utilization of the resources there should be a load balancing algorithm that equally distributed the load on the host. Migrations degrade the performance of the VM. VM performances degrade with number of migration. In this paper we proposed the energy efficient load balancing algorithm based on the threshold.

2. Related Work

VMware Inc. et al. [10], proposed a load balancing approach which is based on the static threshold. They use 45 as a lower threshold and 85 as a upper threshold to define the overloaded and underloaded condition. Load of the host must be in the range of lower and upper threshold. Problem with this approach is that they are using fix lower and upper threshold, which is not suitable for the cloud environment where the resource required by the VM change dynamically.

J. Hu et al. [11] proposed a load balancing approach based on the historical data. They use the genetic approach for balancing the load on the host and improve the performance of the system. Historical data are used to predict the future load on the system. This prediction reduces the number of migration and improves the performance.

Y. Fang et al. [12], proposed a load balancing approach, which schedule VM in two steps. 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.

M. Mishra et al. [13], Proposed a VM placement approach based on Vector theory. They used Total capacity vector (TCV), Resource utilization vector (RUV), Remaining capacity vector (RCV) and Resource requirement vector (RRV) to represent the system. All vector are represent in 3D cube as show in figure 4. In this approach firstly they calculate the resource requirement of the VM and resource utilization vector of PM. Then place the VM where the RRV of VM and RRV of PM are complementary to each other.

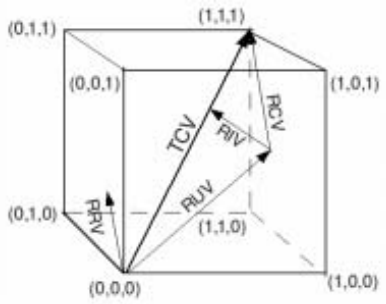


Figure 4: Representation of various vector

A Jain et al. [14], proposed a load balancing approach based on the threshold. They use the single dynamic threshold i.e. upper threshold, which change with the load. Load balancing done on the basis of this dynamic threshold. VM allocated to the host till the upper threshold. A host is called to be overloaded when the load on the host is greater than the upper threshold. Host CPU utilization are used to calculate the upper threshold. For calculating the threshold following formula are use

$$\theta = \frac{\alpha_1 + \alpha_2 + \alpha_3 - - - - \alpha_n}{n}$$

Where α_i is the CPU utilization of the i^{th} host and n is the total number of host in the cluster. This approach increase the resource utilization, but number of problem associate with this approach.
i. This approach not support to the sever consolidation. So it will increase the number of active server.
ii. Only CPU utilization is use to calculate the threshold of the host.

3. Proposed Work

Load balancing is one of the challenging task in the cloud computing because VM can changed our resource required dynamically and multiple resources (CPU, memory, bandwidth etc.) are used by the VM. To balance the load on the host we are using lower and upper threshold. Lower threshold are use to find the server consolidation and upper threshold is use to find the overloaded situation. A host is called to be overloaded when the load on the host is greater than upper threshold and if the load is below the lower threshold than the host is underloaded. When the host is overloaded we migrates one or more VM and if the host is underloaded then all VM running into the host move to the other host. Successive migration can affect the system performance. So main objective of our approach is to increase the resource utilization, reduce the energy consumption and minimize the number of migration. Following steps are involved into the proposed load balancing algorithm.

1) Calculate load on the VM
2) Calculate load on the PM
3) Calculate lower and upper threshold
4) Select the host for the VM selection
5) Select the VM from the host
6) Select the target PM

3.1 Calculate load on the VM

VM behave like a PM. So each VM have its own CPU, memory and bandwidth. Therefore these three parameters are use for calculating the load on the VM. CPU, memory and bandwidth used by the i[th] VM can be define as

$$VM_i^{cpu} = \frac{\text{totalRequestedMips by the VM}}{\text{total MIPS capacity of the PM}}$$

$$VM_i^{bw} = \frac{\text{Bandwidth use by the VM}}{\text{Total bandwidth capacity of the host}}$$

$$VM_i^{ram} = \frac{\text{ram use by the VM}}{\text{Total ram capacity of the host}}$$

Load on the i[th] VM can be calculated as

$$VM_{load}^{i} = VM_i^{cpu}$$

3.2 Calculate load on the PM

Since number of VM are running into each host. So total load on the PM is the summation of all VMs load.

$$VM_{load}^{m} = \frac{\sum_{i=1}^{n} VM_{load}^{i}}{n}$$

Where n is the number of VM into the m[th] host.

3.3Threshold calculation

For calculating the lower and upper threshold we gives the equal weight to all three resources. During the experiment it's observed that as the threshold increased performance of the system is decreased. So we design formula which decreases the upper threshold with the load.

$$\text{Temp} = \sum_{i=1}^{n} \frac{VM_i^{cpu} + VM_i^{ram} + VM_i^{bw}}{n}$$

$$T_{upper} = 1 - x * \text{Temp}$$

$T_{lower} = 0.3$

3.4 Select the host for the VM selection

Load on the host must be in the range between the lower and upper threshold. Host which load is beyond this limit is selected for transferring the VM.

3.5 Select the VM from the host

Numbers of VM are running into each host. So it is very difficult to decide which VM is migrated. Wrong VM selection can increase the number of migration. In the server consolidation all VM running into that host is migrated and in the load balancing VM whose size is greater than or equal to the difference between the upper threshold and the host utilization is the best VM for the migration.

1. Input: hostList, vmList Output: migrationList
2. Arrange each host into decreasing order of their utilization
3. for each h in hostList do
4. hostUtil ←host.util()
5. bestVm ← 250
6. while hostUtil > host.upThresh() do
7. for each vm in vmList do
8. if vm.util() > hostUtil ─ host.upThresh() then
9. temp ← vm.util() ─ (hostUtil ─ host.upThresh())
10. if temp < bestVm then
11. bestVm ← temp
12. bestVm ←vm
13. else
14. if bestFitUtil = 250 then
15. bestVm ← vm
16. break
17. hostUtil hostUtil ─ bestVm.util()
18. migrationList.add(bestVm)
19. vmList.remove(vm)

Algo for the consolidation

1. if hotUtil < lowThresh() then
2. migrationList.add(h.getVmList())
3. vmList.remove(h.getVmList())
4. return migrationList

3.6 Select the target PM

In our approach we select the minimum utilized host for placing VM. All host in the data center arrange into the decreasing order of their utilization and place the VM which is less utilize and consume less power.

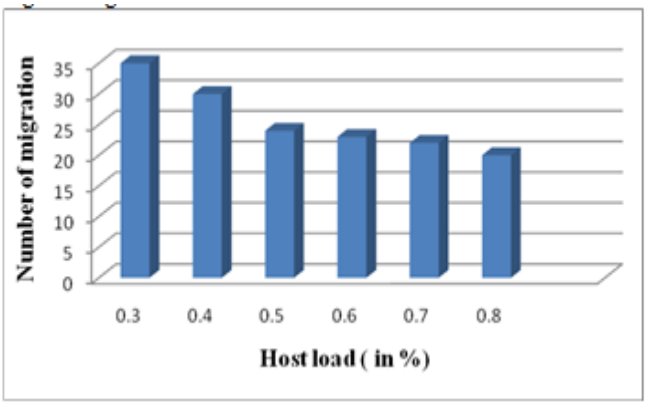
1. Input: hostList, vmList Output: allocation of VMs
2. Sort all host into the decreasing order of their utilization
3. foreach vm in vmList do
4. foreach host in hostList do
5. if Host_Load<=H_UTD)
6. Calculate the power after allocating the VM to the each host
7. if power < minPower then
8. allocatedHost ← host
9. minPower ← power
10. if allocatedHost≠ NULL then
11. allocate vm to allocatedHost

4. Experimental Result

CloudSim [15] simulator is used to implement our load balancing approach. CloudSim provide the cloud environment. It contained java library file for each entity such as VM, host data center Cloudlet etc. We have simulated a data center comprising 10 host and 20 VM. For the implementation of our approach we configure each host with 1000 MIPS, 100000 bit/sec bandwidth and 10000 MB of ram. Each VM having 2500 bit/sec bandwidth and 128 MB of ram. Random function are use to assign the MIPS for each VM. Random functions assign MIPS for each VM between 1-250.

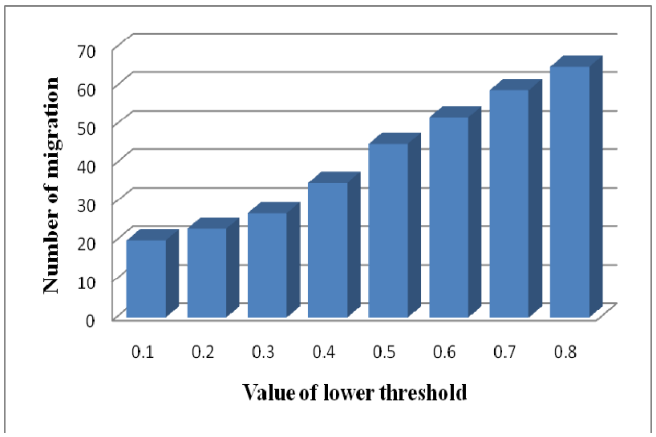
To find the upper threshold we execute our load balancing method 10 times for the different value of x=0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and observed number of migration. Number of

migration increase while value of x decrease and vice versa. We notice that our algorithm gives best result at x=.05.

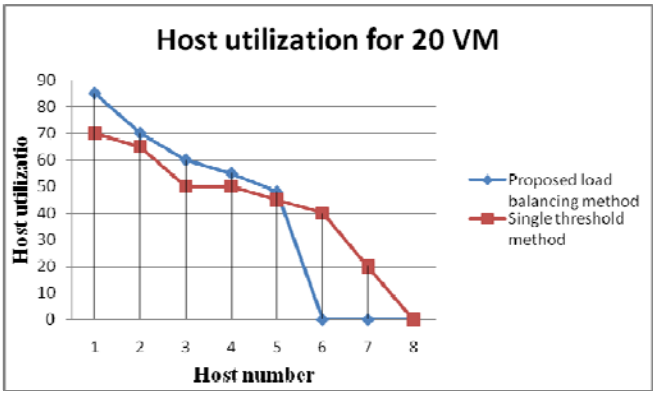


Graph 1: Number of migration on different value of x

Here we are using static lower threshold. So to find the best value of lower threshold we execute our load balancing method 10 times for the different value of lower threshold and observed the number of migration. We found that number of migration increased with the value of lower threshold. Its gives the best result at 0.3.



Energy consumption is depends on the number of active host. Energy consumption increased with the active host. Number of active host can be minimized by the proper utilization of the host resources. So we consider host utilization parameter to compare our method with the threshold based method. We compare both the approach for 20VM and found that threshold method required 7 host to place 20 virtual machine (VM), while our method required only 5 host. So our approach gives better result, because it required less number of physical machines to place 20 virtual machines.



Graph 3: Host utilization for 20 VM

5. Conclusion and Future Scope

Energy consumption and resource utilization is a challenge task in the cloud because the cost of energy is increased day by day. Energy consumption and resource utilization is depends on the load balancing. By the efficient load balancing resource utilization can be decreased. In this paper we proposed a double threshold based load balancing approach. Load on host must be belong between the lower and upper threshold. Host utilization is used for calculating the upper threshold. Experiment result show that our method reduce the number of active server and increase the host utilization.

Main objective of approach is reduced the energy consumption and load balancing. As a future work prediction technique can also be used along with approach to minimize the number of migration. Response time can also be use to minimize the number of SLA violation.

References

[1] L. Grit, D. Irwin, A. Yumerefendi. and J. Chase, "Virtual Machine Hosting for Networked Clusters: Building the Foundations for Autonomic Orchestration," in DistributedComputing (VTDC), November 2006.
[2] L. Cherkasova, D. Gupta, and A. Vahdat, "When virtual is harder than real: Resource allocation challenges in virtual machine based it environments," Technical Report HPL 2007-25, February 2007.
[3] M. Armbrust et al. ", Above the Clouds: A Berkeley View of Cloud Computing". In: EECS Department, University of California, Berkeley (2009)
[4] Kim et al., "Experimental Study to Improve Resource Utilization and Performance of Cloud Systems based on Grid Middleware" proceeding in first International Conference on Internet 2009.
[5] Andrzej Kochut, "Power and Performance Modeling of Virtualized Desktop Systems", in Proceedings of MASCOT 2009.
[6] VMware Inc. "VMware distributed power management concepts and use", 2010.
[7] Anton Beloglazov et al., "Adaptive Threshold-Based Approach for Energy-Efficient Consolidation of Virtual Machines in Cloud Data Centers", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.

[8] Yiqiu Fang et al., "**A** Task Scheduling Algorithm Based on Load Balancing in Cloud Computing" , In ICS '08: Proceedings of the 22nd annual international conference on Supercomputing, Springer, 2010.
[9] Y. Wu et al. "Performance modeling of Virtual Machine Live Migration," IEEE/ACM 4th International conference on cloud computing, pp. 492-499, July 2011.
[10] Mayank Mishra and Anirudha Sahoo, "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE 4th international conference on cloud computing in 2011.
[11] R. Calheiros, R Ranjan, César A. F. De Rose, R. Buyya, "CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services", 2011.
[12] Sosinsky et al." Cloud Computing Bible", Wiley Publishing Inc 2012
[13] A Jain et al., "A Threshold Band Based Model For Automatic Load Balancing in Cloud Environment", in proc. of IEEE International Conference on Cloud Computing in Emerging Markets, pp 1-7, 2013.
[14] RK Gupta et al., "A Complete Theoretical Review on Virtual Machine Migration in Cloud Environment", International Journal of Cloud Computing and Services Science (IJ-CLOSER), Vol.3, No.3, June 2014, pp. 172~178.
[15] R. Gupta et al." Survey on Virtual Machine Placement Techniques in Cloud Computing Environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 4, August 2014, pp. 1 -7.

Author Profile

Pavan Gurve pursuing M.Tech. degree in Computer science and Engineering from IES Institute of Technology and Management Bhopal, India

Aishwarya Mishra working as Associate Professor in Computer Science Engineering from IES Collage of Technology Bhopal ,Madhya Pradesh, India