ISSN:0975-9646

Energy Efficient Dynamic Threshold Based Load Balancing Technique in Cloud Computing Environment

Shivani Gupta[#1], Damodar Tiwari[*2], Shailendra Singh[#3]

Department of Computer Science & Engineering
Rajiv Gandhi Proudyogiki Vishwavidyala, Bhopal
Madhya Pradesh (India)

Abstract— **Cloud computing is the new emerging technology in the field of research and IT industries. So the demand for the computational power required by the business, web and scientific application are rapidly increased in recent past year. To fulfill this demand large scale data center are created which consume enormous amount of electric power. Energy consume by the data center can be reduced by minimizing the number of active server which is known as server consolidation and by the proper load balancing. To achieve the server consolidation and the load balancing VM migration is used. VM migration is an important feature provided by the virtualization. It is the process of transferring the VM from one host to the host. VM migration is the costly operation which degrades the system performance. Proper load balancing can help to minimize the number migration as well as energy consumption. This paper present a load balancing approach based on the Virtual Machine migration. Lower and upper threshold are use for the load balancing. When the load on server is above the upper threshold or below the lower threshold, system is unbalance and some VM has to be migrated. Our result show that the proposed method increase the resource utilization by applying the dynamic lower and upper threshold compare to the traditional VM migration algorithm.**

Keywords— Put your keywords here, keywords are separated by comma.

I. INTRODUCTION

Cloud computing is emerging as a new technology in the field of computer science [1,2]. It become so famous because of their attractive features such as availability, metered services, cheap and online accessibility anywhere in the world. Cloud support three type of services (Software as a services (SaaS), platform as a services (PaaS) and infrastructure as a services (IaaS)) and can be deploy in three different way (private , public and protected) [3,4]. User can use these services anywhere at any time using pay-as-you-use model. In cloud environment each data center having a number of host. So the load balancing among the different host and the resource utilization are the most challenging task in the cloud. Virtualization is used to increase the resource utilization.

Virtualization [5, 6] is the key technology behind the cloud computing. It is a process which divides the physical resources and allows running multiple OS on a single physical machine at the same time.

Virtualization is implemented through the hypervisor also known as virtual machine monitor (VVM), which is a small layer between the physical hardware and operating. VMM is responsible for all the operation related to the VM such as scheduling, VM creation Destruction etc. Each host can create number of virtual machine as show in figure 1. And these virtual machines are provided to the client to run their application as pay per use model. When any request come from any client for cloud service then VMM create new virtual machine and assign the required resources to that virtual machine. The actual execution of the cloud services is on host server but client see the separate virtual machine with their required resources. So there is an illusion that the user is not sharing the computing resources.

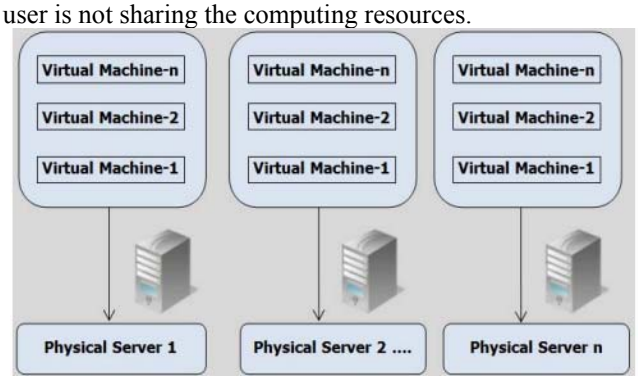


Figure-1: Virtualization

In cloud environment each host can have number of VM and one VM can run multiple applications. Resources required by the VM are dynamic in nature. It means VM can increase or decrease their resource demand according to the application requirement. If VM decrease their resource requirement then VM releases extra resources. If the resource required by the is VM increase the VMM allocate the extra resources to the VM, but what happen if the resource requirement of the VM can't be fulfilled by the host in which VM is running. To deal with this situation VM migration [7] are used.

VM migration is the process of transferring the VM from one host to another host without the compromising the application performance. VM are required to deal with load balancing, hot spot mitigation, server consolidation, server failure etc.

Server consolidation- VMs on lightly loaded hosts can be "Packed" onto fewer machines to meet resource requirements. The freed-up PMs can either be switched off (to save power) or represent higher-resource availability bins for new VMs.

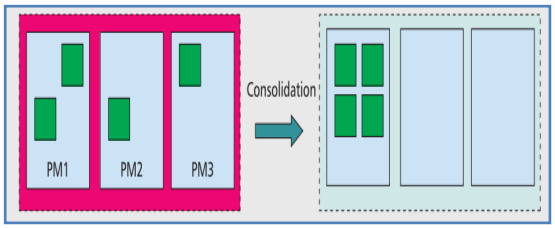


Figure 2: Server consolidation

Hot spot mitigation- It is a situation where the resource required by the VM are not fulfil by the PM in which VM are running currently. Under such conditions, additional resources can be allocated either locally (on the same PM) or within the set of PMs available for provisioning. When local resources are not sufficient to remove the hot spot, VMs can be migrated to another host to make the resources required available to mitigate the hot spot.

Load Balancing- It is a process that distributes the dynamic load evenly across all nodes in the cloud, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded.

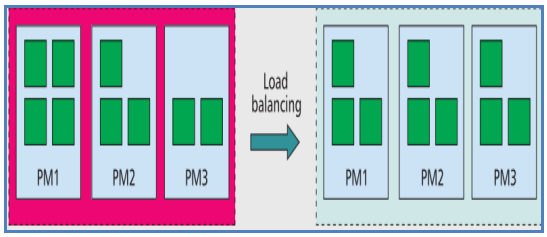


Figure 3: Load balancing

VM migration is the important feature of the virtualization. It allows the flexibility to move VM, but some cost associate with migration. Excessive number of migration can degrade the performance of the VM, because migration required some proceeding power. So number of migration should be less as possible. Number of migration can be reduced by the proper load balancing

II. RELATED WORK

VMware Inc. et al. [8], proposed double threshold based approach. Two thresholds that is upper threshold and the lower threshold with values 45% and 81% are used for the load balancing. The upper threshold will let know the machines when to migrate the selected VM and lower threshold will let know when to migrate all VM. Problem with this approach is that they are using static threshold. Since resource requirement of the VM change dynamically so static threshold is not a good solution.

Mayank Mishra et al. [9], Proposed an energy efficient VM placement approach for the cloud. This approach is based on the vector theory. All the quantities related to various resources i.e. total capacity of the physical and virtual machine, resource utilization vector etc. as vectors. Static lower and upper threshold are used to define the underloaded and overloaded situation respectively. When a PM runs on low utilization, the VMs on the server can be migrated to other PMs so that this PM can be taken off line, which is known as a serve consolidation. Main objective of this approach is to balance the resources such that CPU, memory and bandwidth etc. To achieve this goal they place the VM to the PM where the resource required by the VM and the remaining capacity of the PM is complementary to each other. This approach reduced the resource leakage but they use the static threshold, which increase the human monitoring.

Y. Song et al. [10], proposed a RAINBOW model for the resource allocation. Resource are allocated on the base of priority. VM whish having the higher priority firstly assign to the host compare to the VM which have a lower priority. This approach seems good, but there may be starvation if VM with higher priority comes regularly. This approach also not support to the VM migration.

T. Wood et al. [11], proposed an approach for the hot spot mitigation know as sandpiper. Sandpiper use black and gray box approach to monitor the host. They use the Xen hypervisor. The monitoring engine is responsible for tracking the processor, network and memory usage of each virtual server. It also tracks the total resource usage on each physical server by aggregating the usages of resident VMs. Problem with this approach is that they only consider the cpu load to calculate the load on host.

A Jain et al. [12], proposed a dynamic threshold based approach for the load balancing. Upper threshold are used to decide the overloaded condition. When the load on the host is more than the upper threshold some virtual machine has to be migrated. For calculating the threshold they consider the average CPU utilization of the host. This approach seems good because they are using the dynamic threshold, but this problem has number of anomalies. First they are using only CPU utilization, second average CPU utilization are used to calculate the threshold and last problem is that they are not using the lower threshold for the server consolidation.

III. PROPOSED ALGORITHM

In cloud environment each PM can host number of VM and number of application can run on a single host. So load on the VM is depends on the size and type of the application. Load on the host is the load of all VM running into that host, which can be change dynamically. So we proposed a dynamic threshold based approach. Our approach use two thresholds that is upper and lower threshold to define the overloaded and undreloaded condition. When the load on the host is below the lower threshold, all VM running into that host will be migrated which is known as server consolidation. When the load on the host is above the upper threshold we migrate some VM so that host serve normal load. Four steps are involved in the VM migration. Main goal of our approach is to reduce the number of migration and reduce the energy consumption that can we achieved through the proper load balancing. To achieve our goal following steps are performed

i. Calculate load on the PM and VM.
ii. Calculate the upper and lower threshold to find the overloaded and undreloaded condition.
iii. Select the best VM for the migration
iv. Select the best host to place the selected VM

a. CALCULATING LOAD ON THE PM AND VM.

Three parameter CPU, RAM and bandwidth are the mostly used resources in the cloud environment. Therefore we use these three parameters for calculating the load on the physical and virtual machine

$$VM_{cpu} = \frac{\text{totalRequestedMips}}{\text{total MIPS of the PM}}$$

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

$$VM_{ram=} \frac{\text{ram use by the VM}}{\text{Total ram of the host}}$$

CPU utilization plays an important role in the virtual machine load. So CPU used by the VM can be considered as a VM load.

$$VL_{load} = = \frac{\text{totalRequestedMips}}{\text{total MIPS of the PM}}$$

Number of VM are running in a single host and each VM executing some application. These VM use the PM resources. So load on the physical is equal to the summation of all VM load running on that physical machine. If there are n VM on p^{th} host then average load on the p^{th} host can be calculated by the given equation

$$HL_p = \frac{\sum_{i=1}^{n} VL_i}{n}$$

b. Lower and upper threshold calculation

It is found during the experiment that number of migration increase with the threshold. That means probability of the migration is high on the higher threshold. We are using this information to calculate the threshold.

$$VmTotal = \sum_{i=0}^{p} \frac{Requested\ MIPS\ in\ VM}{Available\ MIPS\ in\ VM}$$

$$Temp = VmTotal + \sum_{i=0}^{p} \frac{VM_i^{ram}}{Total\ ram} + \sum_{i=0}^{p} \frac{VM_i^{bw}}{Total\ bandwidth}$$

$$Th_{upper} = 1 - x*temp$$

Where Th_{upper} is a upper threshold and x is the percentages of the host load, which is based on the experiment. After the experiment value of x=5%.

Lower threshold play an vital role in the load balancing approach. If the lower threshold value is minimum is will increase the number of active server and if the value is high it will increase the number of migration. So we use the static value for the lower threshold that is 30%. of the total capacity of the host.

$Th_{lower} = 30\%$

C. VIRTUAL MACHINE SELECTION

Numbers of virtual machine are running in a single host. So which virtual machine is migrating is also effect the system performance. Down time and the total migration time are two the parameters that can be consider for choosing the suitable virtual machine. Numbers of virtual machine need to be migrated if choose small virtual machine and if we choose the large virtual machine it may increase the down time as well as total migration time. So in our approach initially we find the difference between the host utilization and the upper threshold and then select the VM which has a size greater than or equal to this difference.

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. while hostUtil > H_UTD do
6. worstVmUtil ← First VM utilization
7. for each vm in vmList do
8. if (Vm.Util() > hostUtil ─ H_UTD) then
9. temp ← Vm.Util() ─ (hostUtil ─ H_UTD)
10. if temp < worstVm then
11. worstVmUtil ← temp
12. worstVm ←vm
13. else
14. worstFitUtil = First VM utilization
15. worstVm ← vm
16. end for

Algorithm for the consolidation

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

d. SELECT THE HOST TO PLACE THE SELECTED VM

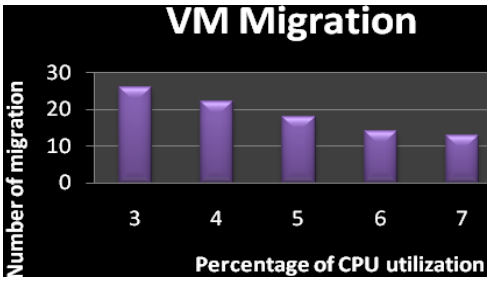
Suitable PM selection is one of the most challenging tasks in load balancing approach. A wrong physical machine selection can degrade the performance of the host. Energy consumption metric is used to place the VM. In our approach we place the VM to the host which consume minimum power among all available hosts in the data center.

1. Input: hostList, vmList Output: allocation of VMs
2. Sort all VM according their utilization
3. for each vm in vmList do
4. minPower ← Max power define by the provider
5. allocatedHost ← NULL
6. for each host in hostList do
7. if ((hostUtil<=H_UTD) && (hostUtil>=H_LTD))
8. Calculate the power after allocating the VM to the each host
9. if(power < minPower) then
10. allocatedHost ← host
11. minPower ← power
12. if (allocatedHost≠ NULL) then
13. allocate vm to allocatedHost
14. else
15. Active new host

IV. EXPERIMENTAL RESULTS

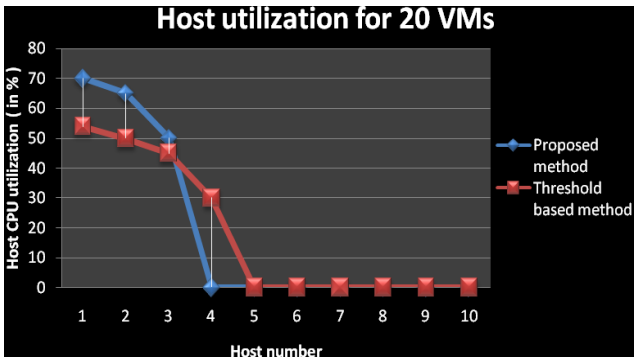
To implement our approach CloudSim [13] simulator is used. It is a collection of java files for the host, VM, allocation policy, data center etc. In our experiment we take 10 hosts, 20 VM and one data center. Each host having 1000 MIPS, 10000 MB ram and 10000 bit/sec bandwidth speed. Each VM configure with 128 MB ram and 2500 bit/sec bandwidth speed. MIPS for each VM allocated randomly between 1-250. This experiment is performed 10 times.

To calculate the upper threshold we need value of x. We run this program 10 times for the value of x= 3, 4, 5, 6, 7 and check the number of migration. As the value of x is increase number of migration decrease as show in graph 1. We observed that number of migration on the value of x=6 and x=7 is almost same. So we choose x=.06.

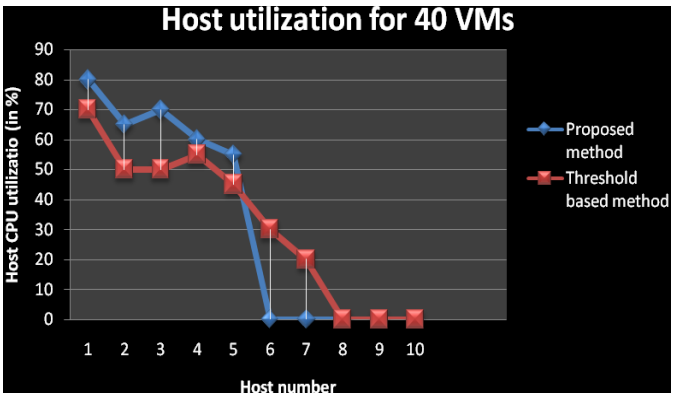


Graph- 1Number of VM migration

Main objective of our approach is to increase the resource utilization. We execute this approach for 20 and 40 VM and found that our method gives better result as show in graph 2 and graph 3



Graph-2 Host utilization for 20 VM



Graph-3 Host utilization for 40 VM

V. CONCLUSION

In the cloud environment resource are provided to the clients on the basis of pay as you go model. User need to pay only for the resources which he used. Cost of the resources charged by the provider can be minimized by the proper utilization of the resource. Load balancing in the cloud is a very challenging task because of their dynamic nature. In this paper we present a load balancing method which use lower and upper threshold. Experiment result show that our method increased the resource utilization as compare to the threshold based method.

REFERENCES

[1]. R. Buyya et al., "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Future Generation Computer Systems, vol. 25, no. 6,june 2011, pp 599-616

[2] Mell, P. et al., "The NIST Definition of Cloud Computing". NIST Special Publication, 2011.

[3] Sosinsk et al. "Cloud Computing Bible", 1st ed., USA : Wiley Publishing Inc.,2012

[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 pp. 25-30, December 2009.

[5] Borja Sotomayor et al., "Enabling cost-effective resource leases with virtual machines," Research Gate article may 2014..

[6] L. Cherkasova et al. "When virtual is harder than real: Resource allocation challenges in virtual machine based it environments" in proc. 10th conference on hot topic in operating system, Vol. 10, pp.20-20, 2005.

[7] Y. Wu et al. "Performance modeling of Virtual Machine Live Migration," IEEE/ACM 4th International conference on cloud computing, pp. 492-499, July 2011.

[8] VMware Inc. "VMware distributed power management concepts and use", 2010.

[9] Mayank Mishra et al., "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE/ACM 4th international conference on cloud computing, pp 275-282, July 2011.

[10] Y. Song, "Multi-Tiered On-Demand resource scheduling for VM-Based data center" In Proc. of the 2009 9th IEEE/ACM Intl. Symp. on Cluster Computing,155, 2009.

[11] T. Wood et al., "Black-Box and Gray-Box strategies for virtual machine migration", NSDI'07 Proceedings of the 4th USENIX conference on Networked systems design & implementation, pp. 7-17, 2007

[12] 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.

[13] 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.