Modified VM Consolidation Approach Using Prediction of Host Utilization to Optimize Energy Consumption in Cloud

Dimpal Patel
PG student[1]
dimpipatel1406@yahoo.in

Prof. Nimisha Patel
Associate Professor[1]
Research Scholar[2]
nimishaa_25@yahoo.co.in

Dr. Hiren Patel,
Professor[3]
hbpatel1976@gmail.com

[1] Sankalchand Patel College of Engineering Visnagar-Gujarat, India
[2] Rai University, Ahmadabad-Gujarat, India
[3] LDRP Institute of Technology & Research, Gandhinagar-Gujarat, India

***Abstract*— Cloud computing is aimed at providing computing resources and services to the Cloud users on demand and on pay-per-use model with ease of flexibility, availability, reliability and elasticity. Increase usage of Cloud has lead to a concern of energy consumption by numerous Cloud data centers across globe. Dynamic virtual machine (VM) consolidation is best solutions to scale down energy consumption and scale up resource utilization. Maximum real works compromise with reduce number of hosts in data center without considering future resource requirements that may result into increase in unnecessary VM migration and service level agreement (SLA) violation. To address the issue, we recommend predicting host utilization and propose a modified utilization prediction aware best fit decreasing (MUP-BFD) algorithm to reduce energy consumption, decrease number of VM migrations and diminish number of SLA violation. The propose mechanism is planned to be implemented in CloudSim and the results are yet to be compared with existing mechanisms.**

Keywords— Cloud computing, Dynamic Consolidation, Utilization Prediction Model, Energy Efficiency, SLA, VM consolidation.

I. INTRODUCTION

Cloud computing is a type of Internet-based computing that provides shared computing resources order. It is model for enabling ubiquitous, on-demand access to shared pool of configurable resources (e.g., computer networks, servers, storage, applications and services) which can be rapidly provisioned and released with minimal management effort [1]. In recent time, Cloud computing is much popular and being used in IT industry. It is a hybrid technology which makes use of parallel computing, distributed computing, grid computing, cluster computing, utility computing on the top of virtualization. It offers various services such as Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS).

Having great popularity and wide adaptation, huge amount of energy consumption by Cloud datacenters is a darker side of it. The issue is being under consideration by many researchers in recent times. VM consolidation is one of the different ways to address the issue of energy efficiency. VM consolidation is approach for reducing power consumption by increasing effective resource utilization by decreasing number of resources required. In consolidation techniques VMs are map on hosts in such a way that least possible hosts remain alive at a given time and put the idle hosts in sleep mode so as to reduce power consumption [2]. Hence, energy consumption can be reduced using VM consolidation. During VM consolidation, we migrate few VMs from source host (normally considered to be underloaded or overloaded host) to target host in such a way that, the target host completes the task where as the source host becomes the vacant host and subsequently it is turned off. While doing so, we also need to consider few other facts such as number of VM migrations, service level agreement (SLA) violations etc. Hence, it is required to design an algorithm which not only reduces energy consumption but also increase the QoS in terms of SLA [2].

In VM consolidation method, detection and selection of overloaded and underloaded hosts is one of the challenging tasks. Further, selection and placement of VMs to be migrated from overloaded host is also an exigent task. Dynamic VM consolidation is a Bin Packing problem. Bin packing problem only deals with minimizing the number of bins used for a given a set of items [3]. Many heuristics algorithms use the concept of bin packing for solving the VM consolidation. However, they produces broad amount of worthlessVM migrations and increase the risk of SLA violation. Among the greater heuristic algorithms, First Fit (FF) algorithm which places all item into the first bin in where it will fit. The other heuristic algorithm is the Best Fit (BF) which puts all item into the filled bin in which is fits. Moreover, the FF and BF heuristics can be upgraded by applying a specific order of items such as First Fit Decreasing (FFD) and Best Fit Decreasing (BFD) algorithms. [3].

In paper [4], authors aim to propose a heuristics algorithm to minimize the VM migrations and SLA violation while achieving energy efficiency. In their work, the proposed dynamic VM consolidation (DVMC) algorithm is partitioning in two phases: (1) migrating some VMs from the hosts that are overloaded presently or may become overloaded in near future, (2) migrating most of VMs from underloaded hosts to moderately loaded hosts. The recommended dynamic VM consolidation allocates a VM to a hosts based on recent and future resource requirements [5][6] , the prediction model can predict the future resource utilization . To test this prediction model we achive the historical data by running various workloads present in Cloud environment [5].

The rest of the paper is organized as follows. Section II discusses about the background theory. Section III is about related work that has been carried out in the domain over recent years. Section IV discusses about our proposal based on VM consolidation. We conclude our work in section V and subsequently list the references used in section VI.

II. BACKGROUND THEORY

Dynamic VM consolidation techniques have been popularly used for developing resource utilization and energy-efficiency in Cloud data centers. A migration of VM and consolidation algorithm for reducing the rate of capacity required to support a specific rate of SLA violations for a given workload. Detecting when a host becomes overloaded directly impacts on rate of SLA violation in terms of QoS, because if the resource capacity is completely utilized, it is highly likely that the applications are experiencing resource shortage and performance degradation calling for the migration of some VMs from it [7]. At a time a overloaded host has been detected, it is basic requirement to determine which VM are selected for migration. This problem is solved by many heuristics VM selection algorithms [7]. Then the placement involves two main steps which are provisioning of resources for the virtual machines according to the capacity requirements of corresponding VM sizing and actual placement of VMs onto hosts. The placement approach should consider multiple resources such as CPU, memory, disk storages and network bandwidth to reduce the energy consumption at data centers and also maintain the energy performance tradeoff. The goal of virtual machine placement is to do VM consolidation for saving a power or load balancing in terms of QoS to the applications running in VMs [7].

CPU utilization refers to a computer's usage of processing resources, or the rate of work done by a CPU. CPU utilization depends on the amount of task run on it. Certain tasks require more CPU time, while some require less because of non-CPU resource requirements [8]. The energy consumption of host depends on the utilization of a CPU, memory, disk and network card. The resource utilization of a host is usually represented by its CPU utilization [9].

Host prediction is a center concern for efficient resource utilization in a dynamic Cloud computing environment. Effective host prediction will help us to prevent the system suffering from heavy workload. The key to accurate host prediction in Cloud computing is appropriate modeling of the relationship between historic data and future values [10].

III. RELATED WORK

In Cloud, techniques such as dynamic VM consolidation aim to scale down energy consumption and to improve the resource utilization. Secron[3] considers a threshold amount to any CPU's utilization that leads to degradation in performance. So, they try to keep total usage of host below threshold value. They predict future demand based on historical data of each server in Cloud datacenters using linear regression.

VM consolidation is done by dealing with four following scenarios [11][12].

1. To find the underloaded hosts to place them in sleeping mode by migrating all VMs to other active hosts (**Under load Detection**).
2. To detect which host is overloaded and migrate some VMs from identified overloaded host to other host to preserve QoS (**Overload Detection**).
3. To select which VM should be migrated (**VM Selection**) from overloaded host.
4. Placing selected VMs on other active hosts (**VM Placement**).

Monil et al. [13] design an overloaded detection method which can give us to better results. It provides an overload detection function based on statically declared threshold value and accordingly, SLA violation is measured. After detection of overloaded host, some VMs are selected based on basic VM selection methods and migrate that selected VMs according VM selection strategies in such a manner that minimum host remained active and reduction of migration control.Hieu et al. [14] present a virtual machine consolidation with usage prediction (VMCUP) for improving energy efficiency of Cloud. Virtual machine consolidation process estimates a short time future CPU utilization based on historical data with the combination of current and predicted CPU metrics. We can characterization of overloaded and under loaded host so we reduce the power consumption and load after consolidation. Farahnakin et al. [5] propose a CPU usage prediction method based on linear regression. It shorts the CPU utilization for short time with usage of history of each host. When host becomes overloaded some VMs are migrated to other host to avoid SLA violation. Monil et al. [15] propose a technique where utilization of datacenter are monitored and underutilized datacenter are put in sleep mode and migrate VMs on capable datacenters. Authors propose a consolidation algorithm with the amalgam of heuristics approaches and migration control for improving VM migration with the use of a maximum correlation with migration control.

IV. OUR PROPOSAL

In this paper, we propose a VM consolidation approach which is known as modified utilization prediction aware best fit decreasing (MUP-BFD) that increase the VM placement according to the current and future resources requirements. The MUP-BFD algorithm migrates VMs from most-loaded hosts to moderately-loaded hosts in order to reduce the energy consumption of data centers. Algorithm 1 depicts our proposal. For the clarity, the concepts used in the proposed algorithm and their notations are tabulated at Table I.

L_h	load of the host *h*
L_v	Load of the VM *v*
$C_{CPU(h)}$	total CPU capacity of the host *h*
$C_{CPU(v)}$	total CPU capacity of the VM *V*
$C_{MEM(h)}$	total memory capacity of the host *h*
$C_{MEM(v)}$	total memory capacity of the VM *v*
$U_{CPU(h)}$	used CPU capacity of the host *h*
$U_{CPU(v)}$	used CPU capacity of the VM *v*
$U_{MEM(h)}$	used memory capacity of the host *h*
$U_{MEM(v)}$	used memory capacity of the VM *v*
$R_{CPU(h)}$	CPU load of the host *h*
$R_{CPU(v)}$	CPU load of the VM *v*
$R_{MEM(h)}$	memory load of the host *h*
$R_{MEM(v)}$	memory load of the VM *v*
$PU_{CPU(h)}$	predicted used CPU capacity of the host *h*
$PU_{CPU(v)}$	predicted used CPU capacity of the VM *v*
T_{CPU}	utilization threshold value
Hover	set of overloaded hosts
Hˆover	set of predicted overloaded hosts
M	the migration plan
hs	the source host for migrated VM allocation
hd	the destination host for migrated VM allocation
v	the migrated VM
Vm	set of selected VMs on a host for migration

Table 1 Summary Of Concepts & Their Notations

In our algorithm, we start with sorting all hosts based on decreasing order of utilization. We have a set of overutilized hosts and predicted overutilized hosts. For each host from this list, we sort all the VMs on the host based decreasing order of utilization. For each VM, we try to find a target host in such as way that the target host goes not get overutilized and also it should not get overutilized in future (predicted utilization) based on response time. For all such suitable hosts, we further look for host with minimum increase in power consumption, which is set as target host for the selected VM from a source host. If we happen not to find any suitable host, we switch on a dormant host and place the VM on it.

In second half of our algorithm, we try to vacate least loaded hosts. We start with the host with lowest utilization and try to place all VMs from it to another targeted host. If we successfully transfer all VM, then we switch off the source host and continue the process with next least utilized host until there is no further scope of VM placement.

Algorithm 1 Modified Utilization Prediction Aware best fit Decreasing Algorithm(MUP-BFD)

```
--------OVER-UTILIZED HOST MANAGEMENT--------
M = Ø
H ← sort all hosts in descending load Lh
for h_s ∈ [H_over, Hˆ_over] do
Vm ← sort VMs on host h_s in descending load L_v
    for v ∈ Vm do
    minPower ← MAX
        if ((U_CPU(h_s) ≥ C_CPU(h_s))|| (PU_CPU(h_s) ≥ C_CPU(h_s)) then
          for h_d ∈ H − [H_over, Hˆ_over] do
                if (U_CPU(h_d)+U_CPU(v) ≤ T×C_CPU(h_d)) & (PU_CPU(h_d) +
PU_CPU(v) ≤ T ×C_CPU(h_d))          then
                    if(AfterPower(h_d)-BeforePower(h_d)<minPower) then
                    minPower = AfterPower(h_d) - BeforePower(h_d)
                    DestHost=h_d
                    endif
                endif
          endfor
                if(minPower != MAX) then
                            M = M ∪ {(h_s, v, DestHost)}
                            Update U_CPU(h_s) and U_CPU(h_d)
                            break;
                endif
                else
                            break;
          end if
    end for
if ((U_CPU(h_s) ≥ C_CPU(h_s))|| (PU_CPU(h_s) ≥ C_CPU(h_s)) then
Switch on a dormant host
end if
end for

--------UNDER-UTILIZED HOST MANAGEMENT--------
H ← sort all hosts in descending load L_h
h_s ← last host in H
Vm ← sort VMs on host h_s in descending load L_v
Count ← Total VMs on h_s
        for v ∈ Vm do
        minPower=MAX
          for h_d ∈ H − h_s do
              if(U_CPU(h_d)+U_CPU(v)≤T×C_CPU(h_d)) &(PU_CPU(h_d)+PU_CPU(v) ≤
T ×C_CPU(h_d)) then
                  if(AfterPower(h_d)-BeforePower(h_d)<minPower)then
                  minPower = AfterPower(h_d) - BeforePower(h_d)
                  DestHost=h_d
                    endif
                  end if
            end for
    if(minPower != MAX) then
                M = M ∪ {(h_s, v, h_d)}
                    Update U_CPU(h_s) and U_CPU(h_d)
                    Count--
                   break;
    endif
          end for
if Count = 0 then
M = Ø
Recover U_CPU(h_s) and U_CPU(h_d)
else
Switch h_s to the sleep mode
end if
```

 ISBN : 978-93-5267-370-4

Flowchart :

The Flowchart illustrated in figure 1 for Overloaded Host Management and figure 2 for Under loaded Host Management , summarized the steps of Algorithm.

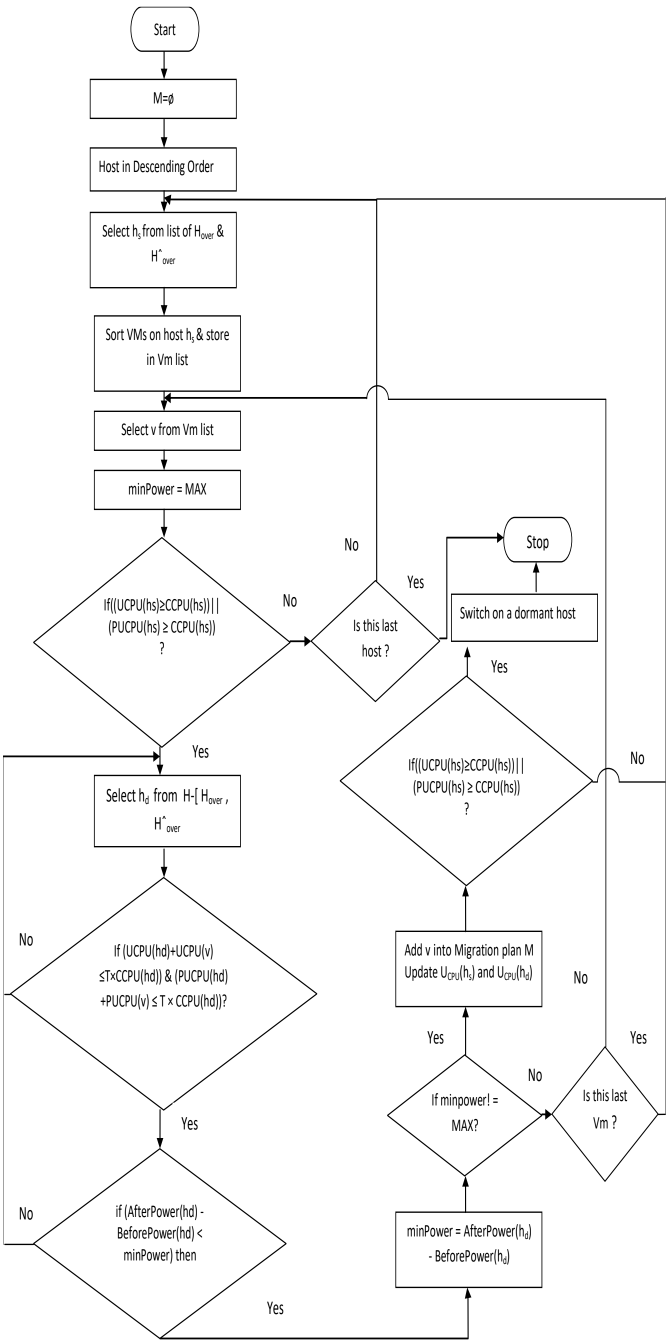

Fig 1 flowchart 1 for Overloaded Host Management

Fig 2 flowchart 2 for Under loaded Host Management

V. CONCLUSION

In this paper, after studying various VM consolidation schemes, we propose a dynamic VM consolidation approach which reduces unnecessary VM migrations and SLA violation using prediction model. The mechanism aims migrating VMs from overloaded host or the host which may tend to become overloaded in near future. Further, we recommend putting underloaded host in sleep mode to save energy. The proposed algorithms are only theoretically tested. In next phase of our work, we aim to implement our resource allocation algorithms on CloudSim [16] simulator and test them with real dataset provided by PlanetLab [17]. The outcome of experimentation will be compared with that of contemporary policies.

VI. REFERENCES

[1] Peter Mell Timothy, Grance :P(2011).The NIST Definition of Cloud Computing, Special Publication 800-145.

[2] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, “Live migration of virtual machines,” in *Proceedings of the 2Nd Conference on Symposium on Networked Systems Design & implementation*, ser. NSDI’05, vol. 2, 2005, pp. 273–286.

[3] A. Murtazaev and S. Oh, “Sercon: Server consolidation algorithm using live migration of virtual machines for green computing,” *IETE Technical Review*, vol. 28, no. 3, pp. 212–231, 2011.

[4] F. Farahnakian, T. Pahikkala, P. Liljeberg, J. Plosila and H. Tenhunen, "Utilization Prediction Aware VM Consolidation Approach for Green Cloud Computing," 2015 IEEE 8th International Conference on Cloud Computing, New York City, NY, 2015, pp. 381-388.

[5] F. Farahnakian, P. Liljeberg, and J. Plosila, “LiRCUP: Linear regression based CPU usage prediction algorithm for live migration of virtual machines in data centers,” in *Software Engineering and Advanced Applications (SEAA), 2013 39th EUROMICRO Conference on*, 2013, pp. 357–364.

[6] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.

[7] Girish Metkar, Sanjay Agrawal, Dr. Shailendra Singh “A Live Migration of Virtual Machine Based on the Dynamic Threshold at Cloud Data Centres” Volume 3, Issue 10, October 2013 ISSN: 2277 128X.

[8] Hesham Eldeeb, Hanan Elazhary , “Virtual Machine Consolidation Challenges: A Review” International Journal of Innovation and Applied Studies ISSN 2028-9324 Vol. 8 No. 4 Oct. 2014, pp. 1504-1516.

[9] A. Beloglazov, J. Abawajy, and R. Buyya, “Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing,” Future Generation Computer Systems (FGCS),vol. 28, no. 5, pp. 755–768, 2011.

[10] Chandini M.S , Mrs Pushpalatha R , Dr. Ramesh Boraiah , “A Brief study on Prediction of load in Cloud Environment International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, pp. 157 – 162, 2016.

[11] A. Beloglazov and R. Buyya, “Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers,” Concurrency and Computation: Practice and Experience (CCPE), vol. 24, no. 13, pp. 1397–1420, 2012.

[12] A. Beloglazov: “Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing”, PhD Thesis, University ofMebourne, 2013. Link: http://beloglazov.info/thesis.pdf.

[13] M. A. H. Monil and R. M. Rahman, "Implementation of modified overload detection technique with VM selection strategies based on heuristics and migration control," Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on, Las Vegas, NV, 2015, pp. 223-227.

[14] N. T. Hieu, M. Di Francesco and A. Ylä-Jääski, "Virtual Machine Consolidation with Usage Prediction for Energy-Efficient Cloud Data Centers," 2015 IEEE 8th International Conference on Cloud Computing, New York City, NY, 2015, pp. 750-757.

[15] M. A. Haque Monil, R. Qasim and R. M. Rahman, "Energy-aware VM consolidation approach using combination of heuristics and migration control," Digital Information Management (ICDIM), 2014 Ninth International Conference on, Phitsanulok, 2014, pp. 74-79.

[16] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R.Buyya, “CloudSim: A toolkit for modeling and simulation of Cloudcomputing environments and evaluation of resource provisioningalgorithms,” Software: Practice and Experience, vol. 41, no. 1, pp.23–50, 2011.

[17] K. S. Park and V. S. Pai, “CoMon: a mostly-scalable monitoring system for Planet- Lab,” ACM SIGOPS Operating Systems Review,vol. 40, no. 1, pp. 65–74, 2006.