Dynamic Virtual Machine Consolidation in Cloud

Padma Priya. V[1] R. Ranjana[2]
[1]P.G Scholar [2]Assistant Professor
[1,2]Department of Computer & Science Engineering
[1,2]Sri Sairam Engineering College

Abstract— Cloud computing has revolutionized the information technology industry by enabling elastic on demand provisioning of computing resources. Cloud data centers consume enormous amounts of electrical energy resulting in high operating costs and carbon dioxide emissions. The goal is to improve the utilization of computing resources and reduce energy consumption under workload independent quality of service constraints. One method to improve the utilization of data center resources, which has been shown to be efficient, is dynamic consolidation of Virtual Machines (VMs). Energy consumption is reduced by dynamically deactivating and reactivating physical nodes to meet the current resource demand. We propose an optimal offline algorithm for the host overload detection problem, as well as a Markov chain model that allows a derivation of an optimal randomized Control policy under an explicitly specified QoS goal for any known stationary Workload.. Here going to implement by using CloudSim Tool Kit by creating Data Center, hosts, Broker, Virtual Machine etc.

Key words: cloud computing, virtual machines, Qos, data center

I. INTRODUCTION

Cloud computing is computing paradigm, where a large number of systems are connected in private or public networks, to provide scalable infrastructure for application, file storage. Users can access the cloud services anytime and anywhere. Cloud services are referred to three categories: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). With growing these services, most of companies such as Google, Yahoo, Microsoft and IBM extend their datacenters.

Therefore, when large number of servers in datacenters leads to high energy consumption. In recent years, many efforts have been made to improve the energy efficiency of datacenter including processor, storage and energy management. Also, Virtualization is one of the important techniques to reduce energy consumption in datacenters. Here virtual machines (VMs) are allocated to minimum number of physical machines such as utilization of active physical machines is also maximized. In this work, consider MIPS of each VM instead of CPU load. For example, assume that a company has four servers (S1-S4) and each server has one CPU core with performance to 1000 Million Instructions per Second (MIPS). There are six VMs (VM1-VM6) with the performance of 250, 500, 1000, 750, 500 and 500 MIPS. Suppose all the servers are found to be initially idle and the first incoming users request needs 100, 200, 600, 300 and, 200 MIPS of each VM respectively to run. Here they are two possible solutions to assign VMs to cloud servers.

A. Solution 1:

A VM-based assignment is chosen for solution leading to assign VM1 and VM3 to S1, VM2 and VM5 to S2 and finally, VM6 to S3. Therefore, we use three servers with utilization of 40%, 40% and 60% respectively.

B. Solution 2:

A Request-Based assignment may lead to assign VM1 and VM3 to S1 and the remaining VMs to S2, and consequently, the S3 can be turned off.

In the first solution, Service Level Agreement (SLA) will not be violated but, utilization of servers is very low and the number of active servers is higher than the second solution which leads to the higher energy consumption.

Here I have contributed some steps to improve energy consumption in data centers 1) to improve the quality of VM consolidation, it is first necessary to maximize the mean time between VM migrations initiated by the host overload utilization of each server in such a way that some of the VMs which run on a server have to migrate to other physical machines. 2) An optimal offline algorithm for host overloads detection.

In this work, proposed a novel approach to the problem of host overload detection inspired on power management of electronic systems using Markov decision processes. By building a Markov chain model for the case of a known stationary workload and a given state configuration, and using a workload independent QoS metric derive a Non-Linear Programming (NLP) problem formulation. The solution of the derived NLP problem is the optimal control policy that maximizes the time between VM migrations under the specified QoS constraint in the online setting.

II. RELATED WORK

The primary focus of designers in the industry is based on the improvement of the system performance. This chapter discusses the causes and problems of high energy consumption, and presents energy-efficient design of computing systems which covers the hardware, operating system, virtualization, and data center. With the growth of computer components, the cooling problem becomes more crucial.

This chapter formally defines the single VM migration and dynamic VM consolidation problems. According to El-Yaniv, optimization problems in which the input is received in an online manner and in which the output must be produced online are called online problems. An online algorithm ALG is c-competitive if there is a constant a, such that for all finite sequences I:

ALG (h) _ b _OPT (h) + f

Where ALG (h) is the cost incurred by ALG for the input h; OPT (h) is the cost of an optimal offline algorithm for the input sequence h; and f is a constant. There is a single physical server, and VMs allocated to that host. In this problem the time is discrete and can be split into M time frames. The resource provider pays the cost of energy consumed by the physical server

The rest of the paper is organized as follows:

Section iii discusses the overview of our proposed system model. The calculations is presented in Section v. Section vii gives the conclusion and related work.

III. PROPOSED SYSTEM

Here by introducing two algorithms namely; optimal offline algorithm and CPU control algorithm. By using optimal offline algorithm the energy consumption can be reduced by changing active state of VM to sleep state when they are not in use. In the offline setting, the state of the system is known at any point in time. Another method is by using CPU control algorithm, the utilization of each host in cloud data center can be calculated separately and finally the total energy consumption of host is reduced. Also, the service level agreement (SLA) and bandwidth is taken as Quality of Service constraints.

IV. ALGORITHM

A. *Optimal Offline Algorithm:*

Input: A system state history
Input: N, the maximum allowed OTF
Output: A VM migration time

(1) while history is not empty do
(2) if OTF of history _ N then
(3) return the time of the last history state
(4) else
(5) drop the last state from history

Consider an offline algorithm that passé through the history of system states backwards starting from the last known state. The algorithm decrements the time and re-calculates the OTF $\frac{t_{0\,(t_g,u_t)}}{t_{a(t_g,u_t)}}$ at each iteration. The algorithm returns the time that corresponds to the current iteration if the constraint is satisfied.

V. MATHEMATICAL CALCULATION

A. *SLA Violation Metrics:*

Two metrics for measuring the level of SLA violations in an IaaS environment are proposed:

$$\mathrm{OTF} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \mathrm{T_{ti}}/\mathrm{T_{bi}}$$

$$\mathrm{PDM} = 1/\mathrm{M} \sum_{j=1}^{M} C_{ki}/C_{ri}$$

Where N is the number of hosts; T_{ti} is the total time during which the host i has experienced

the utilization of 100% leading to an SLA violation; T_{bi} is the total of the host I being in the active state (serving VMs); M is the number of VMs; C_{dj} is the estimate of the performance degradation of the VM j caused by migrations; C_{ri} is the total CPU capacity requested by the VM j during its lifetime. In this work, C_{dj} is estimated to be 10% of the CPU utilization in MIPS during all migrations of the VM j.

B. *Optimal Offline Algorithm:*

$$t_{a\ (t_g, u_t)} \to \mathrm{Ma}$$

$$\frac{t_{0\,(t_g,u_t)}}{t_{a(t_g,u_t)}} \le \mathrm{M}$$

where t_g is the time when a VM migration has been initiated; ut is the CPU utilization threshold defining the overload state of the host; $t_{0(t_m,u_t)}$ is the time, during which the host has been overloaded, which is a function of t_m andu_t ; t_a is the total time, during which the host has been active, which is also a function of t_m andu_t; and M is the limit on the maximum allowed OTF value, which is a QoS goal expressed in terms of OTF.

C. *CPU Utilization:*

The models and algorithms proposed in this chapter are suitable for both single core and multi-core CPU architectures. The capacity of a single core CPU is modelled in terms of its clock frequency F. A VM's CPU utilization u_i is relative to the VM's CPU frequency f_i and is transformed into a fraction of the host's CPU utilization U. These fractions are summed up over the N VMs allocated to the host to obtain the host's CPU utilization

$$\mathrm{U} = F \sum_{i}^{N} f_{iU_i}$$

A multi-core CPU with n cores each having a frequency f is modelled as a single core CPU with the n f frequency. This simplification is justified, as applications and VMs are not tied down to a specific core, but can by dynamically assigned to an arbitrary core by a time-shared scheduling algorithm.

VI. EXPERIMENTAL SETUP

By using CloudSim toolkit, components of cloud environment such as data center, virtual machines, and cloudlets are created. Here am going to conFig. 4 data centers and 12 virtual machines and these are get assigned to cloud broker. At first cloudlets are get assigned to the virtual machines. Data center work is to list all the status of workload that given. First step is to allocate the utilization of resources to all the virtual machines and to calculate the energy efficiency in each cloud data center.

Here, Fig1.1 shows the energy consumption of cloud data center and it is compared with Host Overload Detection (HUD), Host Under load Detection (HUD), Optimal Offline Algorithm (OOA).

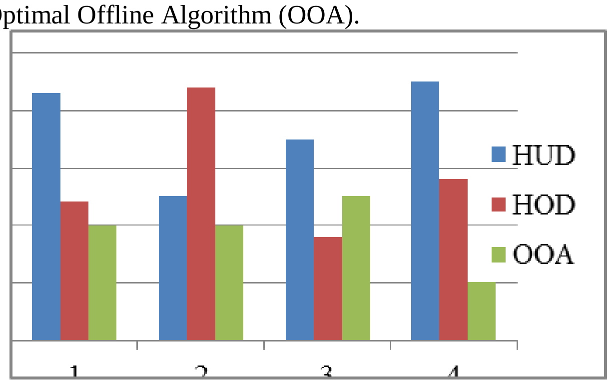

Fig. 1: Energy Efficiency in DC

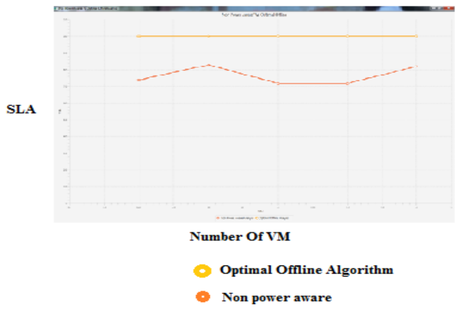

Fig. 2: SLA

Fig 1.2 shows the calculation of SLA between the cloud data center. It is compared between Non –power aware and Optimal Offline Algorithm. Fig 1.3 shows the bandwidth consumption among Virtual machines in host.

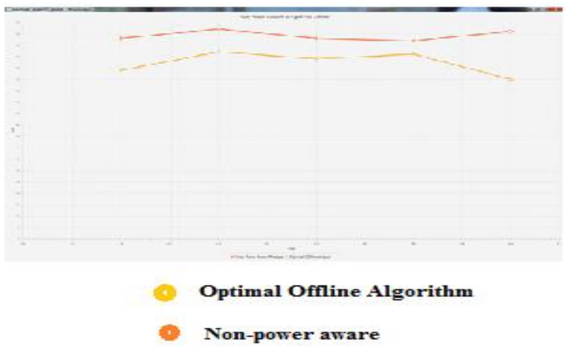

Fig. 3: Bandwidth Consumption

VII. Conclusion

In this paper, introduced a control algorithm to calculate the host CPU utilization for each virtual machine in data centers. In today computing world, the main drawback is data center consumes a huge amount of energy .so, by using offline algorithm going to reduce the energy consumption in data center and also to improve the utilization of resources.

As a future work, plan to implement a Markov chain model for calculating known workload which occur in cloud data center and also the live migration is done for the virtual machine which is overloaded. This is done by using cloudsim tool kit.

Reference

[1] Zhen Xiao, "Dynamic Resource Allocations Using Virtual Machines for Cloud Environment", IEEE Transaction ON PARALLEL AND DISTRIBUTED SYSTEMS, pp.1986-1999, JUNE 2013
[2] Haikun Liu, Hai Jin, Xiaofei Liao," Live virtual machine migration via Asynchronous Replication and state Synchronization", IEEE TRANSACTION ON PARALLEL AND DISTRIBUTED SYSTEM, DECEMBER 2012.
[3] Chen chuan, Jinan," A New live virtual machine migration strategy", 2012
[4] Xiaocheng Liu, Chen Wang, Bing Zhou, Junliang Chen "Priority-Based Consolidation of Parallel Workloads in the Cloud" IEEE Transaction on parallel and Distributed system, SEPTEMBER 2013
[5] G. Jung, M. A. Hiltunen, K. R. Joshi, R. D. Schlichting, and C. Pu, "Mistral: Dynamically managing power, performance, and adaptation cost in Cloud infrastructures, 2010.
[6] Hirofumi Ichihara, Kunio Hato, "On the Integrated Control of Virtual Machine Live Migration and Traffic Engineering for Cloud Computing "on IEEE Transaction, JUNE 2013
[7] LIHongyou, WANG Jiangyong, "Energy Aware Scheduling Using Workload –Aware Consolidation Technique in Cloud Data Centers," on china communication, DECEMBER 2013
[8] Sina Esfandiarpoor, Ali Pahlavan, "Virtual Machine Consolidation for Datacenter Energy Improvement," JUNE 2013
[9] X.Wang, Z.Liu, "An Energy –Aware VMs Placement algorithm in Cloud Computing Environment," in International Conference, 2012
[10] http://phys.org/news205737760.html
[11] http://searchservervirtualization.techtarget.com/definition/live-migration