
Rajeev Kumar Gupta, Maulana Azad National Institute of Technology (MANIT), Bhopal, India
Rajesh Kumar Pateriya, Maulana Azad National Institute of Technology (MANIT), Bhopal, India

ABSTRACT

One of the major challenges for the cloud provider is the efficient utilization of the physical resources. To achieve this, this paper proposed a Balance Resource Utilization (BRU) approach that not only minimizes the resource leakage but also increases the resource utilization and optimize the system performance. The proposed approach consider two resources i.e., CPU and memory, as decision metrics for load balancing and use three thresholds named lower threshold, upper threshold and warning threshold to define underloaded, overloaded and warning situations, respectively. The main concept of this approach is to place VM to the PM, where resource requirement of the VM and resource utilization of the PM are complements to each other. To evade unnecessary migrations due to the temporary peak load AR time series prediction model is used. The authors’ approach treats load balancing problem from the practical perspective and implemented in OpenStack cloud with KVM hypervisor. Moreover, proposed approach resolve the issue of VM migration in the heterogeneous environment.

KEYWORDS

Auto Regression (AR) Model, CPU Load, CPU Utilization, Energy Efficient, Response Time, Virtual Machine, Warning Threshold

INTRODUCTION

Cloud computing is emerging as a new paradigm of large-scale distributed computing (Armbrust et al., 2010). It provides on-demand computing resources i.e., CPU, memory, network etc., as a service to the client through internet (Carretero and Garcia Blas, 2014; Slabevea and Wozniak, 2010; Wind et al., 2011). Virtualization is the technology that makes cloud computing possible (Haro et al., 2012; Younge et al., 2011). It allows sharing of the physical resources. It is implemented through the hypervisor also known as virtual machine monitor which reside between hardware and OS. When a user request for the resources, hypervisor engenders VM according to the user requirements, bind service to the VM and then assign to the user. Although virtualization increases the resource utilization, but it brings the problem of resource leakage and load balancing which needs to be tackled. Virtualization interfuses with the migration enable the cloud provider to balance the load on
the server and consolidate their computing needs to a minimum number of servers. VM migration is a process which allows seamless movement of VM from one physical machine (PM) to another (Falco et al., 2015; Cerroni and Esposito, 2016; Wasim et al., 2015). Migration process consumes some resources for processing the data. Hence, successive migrations degrade the performance of the PM and must be evaded.

After reviewing the theory of cloud computing it is found that electricity bill highly affects the provider revenue which mainly depends on the number of active servers and CPU utilization (Beloglazov et al., 2012). A Recent study (Mishra and Sahoo, 2011; Xu et al., 2012), says that non-uniform resource utilization is one of the main causes for increasing the number of active servers, which will result in higher power consumption. Since cloud resources are multidimensional i.e., CPU, memory, network, bandwidth etc., so there is a situation where resource in one dimension is over utilized while it is underutilized in another dimension. This situation is known as resource leakage which is a phenomenon of resource wastage. Hence, resource leakage, number of migrations and energy consumption are the three critical issues in cloud and must be resolved for optimizing the cloud services.

Cloud is a business model where provider wants to generate higher revenue with minimum investment and high user gratification. Three metrics resource utilization, energy consumption and service level agreement (SLA) violation contributes major role in generating the high revenue to the service provider. These three metrics can be tackled effectively by utilizing the physical resources uniformly. So prime goal of the cloud provider is to reduce the resource leakage which results in optimizing the resource utilization. In order to achieve this goal an efficient load balancing approach is required that dynamically place the VM on the suitable host.

This paper, proposed a dynamic load balancing approach named balance resource utilization (BRU) which will use physical resources uniformly and reduce power consumed by the data center. The main idea of proposed approach is to place VM to the PM, where the resource requirements of the VM and resource utilization of the PM are complements to each other. This approach uses three thresholds to define the under loaded, overloaded and warning situation. Auto regression (AR) prediction model is used to evade the nonessential migration due to the interim peak load. The major contribution of this paper can be summarized as:

- A comparative study of various existing load balancing approaches with their anomalies;
- Find the suitable AR prediction model for forecasting the load value;
- Design an efficient dynamic load balancing approach (BRU) that gives the solution for three questions i.e. when migration is triggered, which VM is migrated and where it is placed;
- Deploy private cloud named OpenStack with kernel virtual machine (KVM) hypervisor to evaluate the performance of our approach;
- Comparative analysis of BRU approach with other existing approaches.

RELATED WORK

(Falco et al., 2015), present a solution to balance the physical server and to minimize the total migration time. Instead of migrating entire VM, this approach migrates task from the overloaded server to another. This approach uses Extremal Optimization (EO) strategy to select candidate task for the migration and for placing the selected task to the appropriate server. Extremal Optimization is a nature-inspired optimization technique which has small computational complexity and memory
Social and Usage-Process Motivations for Consumer Internet Access
[www.igi-global.com/article/social-usage-process-motivations-consumer/3842?camid=4v1a](www.igi-global.com/article/social-usage-process-motivations-consumer/3842?camid=4v1a)

PDAs as Mobile-Based Health Information Deployment Platforms for Ambulatory Care: Clinician-Centric End-User Considerations
[www.igi-global.com/chapter/pdas-mobile-based-health-information/18238?camid=4v1a](www.igi-global.com/chapter/pdas-mobile-based-health-information/18238?camid=4v1a)