Energy Efficient Resource Allocation During Initial Mapping of Virtual Machines to Servers in Cloud Datacenters

Nimisha Patel, Rai University, Ahmedabad, India & Sankalchand Patel College of Engineering, Visnagar, Gujarat, India
Hiren Patel, LDRP Institute of Technology and Research, Gandhinagar, Gujarat, India

ABSTRACT

Energy consumption has been identified as one of the key research challenges during recent time in Cloud computing. Proper placement of Virtual Machines (VMs) on servers may address the issue. The process of placing VMs on servers can be divided into two phases viz. (a) Mapping of VMs on servers during the phase and (b) subsequent VM selection, migration and placement during consolidation phase. If the initial mapping is not efficient, subsequent operations may lead to unnecessary VM migrations, which in turn, may result into increase in migration cost and increase in SLA violations. In this research, the authors aim to improve the resource utilization to address these issues by keeping (i) the number of live server as minimal as possible for achieving energy efficiency, and (ii) the live server, as busy as possible by efficiently utilizing them. The authors conducted series of experiments with existing default technique and various other approaches. The results of our experiments make us conclude that there is a scope of improvement in the default mapping technique currently being used in CloudSim.

KEYWORDS
Cloud Computing, Energy Efficiency, Initial Mapping, Virtual Machine, VM Consolidation

INTRODUCTION

The Cloud computing model (Buyya et al., 2009) has quickly exerted a pull on much of the user’s focus in recent years. NIST defines Cloud as a “model for enabling ubiquitous, 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.” (Mell & Grace, 2011) NIST further lists five essential characteristics of Cloud computing viz. (i) on-demand self-service, (ii) broad network access, (iii) resource pooling, (iv) rapid elasticity or expansion, and (v) measured service. It also lists three service models viz. (a) Software as a Service (SaaS), (b) Platform as a Service (PaaS) and (c) Infrastructure as a Service (IaaS), and four deployment models viz. (1) private, (2) community, (3) public and (4) hybrid. These Cloud services and deployment models assist programmers having ground-breaking thoughts but lack huge capital investment in computing infrastructures to deploy their products in the real market.

DOI: 10.4018/IJDST.2018010103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Cloud works on the top virtualization technology (Fox et al., 2009). Virtualization creates virtual resources on the top of physical machines. These virtual resources may include computing resource, operating platforms, storage devices, main memory, internet bandwidth etc. Virtual machine (VM) is an emulated machine which provides the utility of offering resources in form of platform, storage, compute or network. When a task is submitted to Cloud for computation or for another purpose, the same is served through one or more virtual machines created at Cloud service providers’ (CSP) premise. Hence, any job submitted to Cloud is run under one or more VMs. Multiple logical VMs run under a common server generally known as the host. There are many hosts under a datacenter and one CSP may have several datacenters.

The Cloud datacenters usually comprise of a great number of well-configured and interconnected computing resources (Luo et al., 2014) which consume a significant amount of electricity for their functioning. Increase usage of Cloud computing has led to augmentation in electrical energy consumption by the huge amount of servers in a large number of data centers. The survey shows that the average energy consumption of a datacenter is comparable to that consumed by 25,000 domestic usages (Kaplan et al., 2008). This has attracted consideration of research community in recent years. Out of many different mechanisms to address the issue, efficient initial VM placement has been recognized as few of the popular solutions. During initial VM placement/mapping, we try to minimize active servers in a datacenter without compromising the performance of tasks and user requirement. Sleep/Wake up has been identified as one of the top classifications by Brienza et al. (2016) in which some of the servers are switched off when not in use to save energy and are awakened whenever necessary. It has been seen that even idle servers consume about 70% of peak power (Fan et al., 2007). In a nutshell, proper distribution of existing tasks among available servers may result into minimizing the active servers without compromising SLA with Cloud users.

Hosts running in a datacenter are classified into three categories based on their usage viz. (i) overloaded hosts (ii) underloaded hosts and (iii) normal hosts. This classification is based on host’s utilization, for instance, hosts with utilization more than a certain value (commonly known as upper threshold) may be considered as overloaded hosts and similarly, hosts with utilization less than a certain value (commonly known as lower threshold) may be considered as underloaded hosts. All other hosts except these two categories are considered as normal hosts. According to (Barroso & Holzle, 2007), under the normal scenario, hosts in a datacenter operate only at 10%–50% of their peak capacity, and these underloaded hosts become a reason for the waste of electricity. Hence, it is required to reduce the energy consumption by enhancing hosts’ resource utilization in Cloud datacenters.

Various researchers (Mastelic et al., 2014; Beloglazov & Buyya, 2010; Beloglazov & Buyya, 2012; Beloglazov et al., 2012; Wu et al., 2014; Yang et al., 2014; Huang et al., 2014) claiming to address the issue of enhancing energy-efficiency, work on the concept of minimizing active servers and switching the idle servers into power saving mode. This process is widely known as server consolidation where the workload is transferred from under-utilized and over-utilized hosts to normal hosts. While doing so, the users’ obligation of offering Quality of Service (QoS) in terms of SLA, needs to be sincerely taken into consideration. Hence, it is utmost significant to preserve energy-performance tradeoff during this consolidation progression. While most of the researchers have overlooked the initial mapping, we believe that efficient allocation of workload to the server during the initial phase would result in a momentous enhancement in successive consolidation for a certain time period, which is our main source of motivation for this research.

To understand the gap in existing literature, we have gone through the initial mapping technique available in the CloudSim toolkit (Calheiros et al., 2011). To the best of our understanding, we could learn that the existing initial VM placement technique has drawbacks. First, the technique allocates all the VMs to available hosts, roughly uniformly. It does not take a load of the respective host into consideration while placing a VM on the host during the initial phase. This results into a lot of migrations in a subsequent phase of consolidation and consequently degradation in performance of overall system. Second, on every host, there can
TVGuarder: A Trace-Enable Virtualization Protection Framework against Insider Threats for IaaS Environments
Li Lin, Shuang Li, Bo Li, Jing Zhan and Yong Zhao (2016). *International Journal of Grid and High Performance Computing* (pp. 1-20).
www.igi-global.com/article/tvguarder/172502?camid=4v1a