Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 125 (2018) 683–690

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Smart Computing & Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

A Probabilistic Model for Finding an Optimal Host Framework and Load Distribution in Cloud Environment

Sakshi Chhabra*,[1], Ashutosh Kumar Singh[2]

Department of Computer Applications, National Institute of Technology, Kurukshetra-136119, Haryana, India

Abstract

Cloud Computing is being widely accepted in this computing world and comprises of many hardware and software resources. It becomes critical to meet their client's needs on time if the cloud data centers are so overloaded. In this regard, if the utilization of resources becomes an intelligent way so that we focused on the selection problem of the physical hosts for deploying the tasks. So, OPH-LB (Optimal Physical Host with effective Load Balancing) framework is proposed to model the service of client's requests in this IaaS architecture with heterogeneous virtual machines. It proposes the thought of accomplishing load balancing in this dynamic environment. Firstly, this OPH-LB approach is filtering the qualified hosts among all which accomplishes the requirement of deploying tasks. Then out of those qualified sets, we apply probablistic model which helps to find the most optimal host in terms of its computing capability and its performance function. Further, the performance is analyzed using Cloudsim simulation tool and compared with existing approaches. The results demonstrate that our model has improved the throughput, reduced the failure rate and optimized the attainment of cloud data centers.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications.

Keywords: Cloud computing, load balancing, task deployment, probability, performance, optimality

1. Introduction

Cloud computing is mainly admired in the IT sector and academia both. With the support of virtualization, it is well recognized as *"utility based systems"* [1] [2]. Because users only pay for that resources which is utilized by them. The cloud service provider gives the multiple virtual machines (*VMs*) on lease to users and they create a highly secure platform for all the computations and can save their infra-cost [3] [4].

[1]sakshichhabra555@gmail.com
[2]ashutosh@nitkkr.ac.in

1877-0509
Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications
10.1016/j.procs.2017.12.088

But the main issue is how to reduce the task execution time, utilize the resources efficiently and schedule the incoming requests effectively. Many algorithms have implemented till now like Round Robin, Active VM Monitoring but for the high communication delays in data centers are not addressed clearly [6]-[8]. This leads not to proper utilization of resources and participate in executing the task requests. Sometimes, it happens that cloud data centers accumulate the tasks continually and this will make the whole cloud imbalanced. So, load balancing [15] helps in a major part to distribute the resources efficiently so that it manages the workload traffic or allocation of resources among multiple networks. We review some related papers to our work: The method introduced in [9] enables to pick out the optimal host for deploying the requested tasks to achieve the immediate load balancing. They have combined the procedure with bayes method and cluster to obtain the optimal clustering set of physical hosts. An approach [13] proposed by A. Paulin Florence et. al. a load scheduling algorithm for the cloud networks. This paper enhances the firefly algorithm in which it deals with set of requests and servers in the simulated cloud network. Another approach for the perspective of consumer-centric by Qi Liu et. al.[14]specifies how to improve the execution time of incoming jobs in the heterogeneous cloud environment. It is based on real-time and used MapReduce for optimizing the execution time with prediction model. Xiang Deng et. al.[17] developed the online power management system which helps to reduce the carbon footprints and the operating cost of the data centers. These authors recommended for the use of green datacenters powered by renewable energy. They suggested an online algorithm called *EcoPower* which performs better load balancing and eco-aware power management simultaneously. They have also applied Lyapunov optimization algorithm for design and control this online algorithm. An optimal power allocation technique for mutiple processors [11] introduced by Junwei Cao et. al. aims to develop power and performance across cloud data centers. Their technique efficiency is based on some optimization problems like they are fixing the one factor and minimized the other one. It is well understood that power reduction and performance optimization both are the essential factors for cloud providers to efficiently utilize all the available resources. For networked clouds, launched a new method by [5] for finding the optimal virtual machines for deploying the tasks. They have done mapping with Mixed Integer Programming (MIP) problem because their objectives are related to cost efficiency and QoS-aware virtual resources. An Optimization of virtual machine placement for energy efficient in cloud [10] have investigated the VM placements in IaaS clouds for improving the utilization of data centers and reduces the energy consumption. Their results shows that their energy consumption reduced to 23.01 kWh with percentage of 0.00029 and minimizes the 770 number of VM migrations. J. Zhao et. al. [12] proposed a method for energy-saving through placement selection of live virtual machine migrations. Their performance is also based on its probability theory and Mathematical Statistics to get the better experimental results. It reduces the total energy consumption and preserves the presentation of VM running and its migration and their results become more effective and reliable. The problem that we have considered in this paper can be outlined as follows:

1.1. Our Contribution

The work aims to achieve long-term load balancing in cloud data centers which distributes the load uniformly at the servers by providing efficient performance. Using probabilistic estimations, OPH-LB can find most optimal physical host for the task deployment effectively and achieves relatively accurate estimation with less communication overheads of upcoming load. Out of those qualified physical hosts, this OPH-DL chooses the most optimal node for processing the tasks so that waiting time decreases and usage of computing resources becomes efficient. It produced the random dynamic request streams to simulate a real cloud environment through the task deployment strategy so that we can achieve a better load balancing and resourceful performance of external services which can be provided to the clients. The rest of this paper shows the proposed problem and system architecture is designed. Then simulation results and compared with previous approaches is highlighted in the next section.

2. Proposed Work

This IaaS architecture in the cloud provides the lightness of virtual hardware or infrastructure services by the third-party providers on behalf of its clients. When clients submit their upcoming requests to the