Chapter 23
A Novel Meta-Heuristic Approach for Load Balancing in Cloud Computing

Subhadarshini Mohanty
Siksha 'O' Anusandhan University, Department of Computer Science and Engineering, Bhubaneswar, India

Mitrabinda Ray
Siksha 'O' Anusandhan University, Department of Computer Science and Engineering, Bhubaneswar, India

Prashanta Kumar Patra
College of Engineering and Technology, Department of Computer Science and Engineering, Bhubaneswar, India

Subasish Mohapatra
College of Engineering and Technology, Department of Computer Science and Engineering, Bhubaneswar, India

ABSTRACT

Cloud computing is gaining more popularity due to its advantages over conventional computing. It offers utility based services to subscribers on demand basis. Cloud hosts a variety of web applications and provides services on the pay-per-use basis. As the users are increasing in the cloud system, the load balancing has become a critical issue in cloud computing. Scheduling workloads in the cloud environment among various nodes are essential to achieving a better quality of service. Hence it is a prominent area of research as well as challenging to allocate the resources with changeable capacities and functionality. In this paper, a metaheuristic load balancing algorithm using Particle Swarm Optimization (MPSO) has been proposed by utilizing the benefits of particle swarm optimization (PSO) algorithm. Proposed approach aims to minimize the task overhead and maximize the resource utilization. Performance comparisons are made with Genetic Algorithm (GA) and other popular algorithms on different measures like makespan calculation and resource utilization. Different cloud configurations are considered with varying Virtual Machines (VMs) and Cloudlets to analyze the efficiency of proposed algorithm. The proposed approach performs better than existing schemes.

DOI: 10.4018/978-1-7998-5339-8.ch023

INTRODUCTION

Cloud computing is extensively adopted through a large range of users for providing a solution to massive length computational problems. Cloud environment includes heterogeneous computing resources along with some processors, bandwidth, and memory. By using the virtualized resources, cloud computing environment can create new application situations to provide distributed services within the form of infrastructure, platform, and software on demand basis. The advantage of cloud computing includes location independence, availability, reliability, and optimized cost (Rimal et al., 2009). To achieve the above goals, the task needs to be scheduled properly among various resources. Dynamic load balancing is achieved in cloud computing with the help of virtualization technology. The VM migration technique helps in remapping the digital machine and physical resources dynamically with flexible allocation and reallocation of resources (Clark et al., 2005; Jun et al., 2011). To provide a better opportunity like larger memory, high bandwidth, and computational power, the VM migration is performed by moving the live VMs on execution from one physical machine to another machine (Jin et al., 2011; Mann et al., 2015).

There are numerous current issues in cloud computing such as security, load balancing, performance monitoring, resource scheduling, scalability, identity management, optimal resource management, data transfer cost, and energy management. Among these issues, load balancing is one of the promising issues in cloud computing environment. It ensures that each computing aid is sent effectively and fairly. Load balancing affords a better quality of service by using optimizing the resource utilization and response time (Patel et al., 2013).

Cloud data centers are highly dynamic in nature. The request generates from different users (user bases) and the amount of resource required for executing the tasks changes dynamically. Hence their load behaviors are unpredictable. Although these situations are easy to trigger unbalanced loads in cloud data center but they may lead to performance degradation compromising in Quality of Service (QOS).

Load balancing improves the system overall performance by evenly distributing the workload among nodes (Randles et al., 2010). It ensures better user satisfaction and higher resource utilization. The objective of VM Scheduling with load balancing in cloud computing is to assign VMs to suitable hosts and balance the resource utilization within all of the hosts. Load balancing enables scalability, maximizing the performance and minimizing the response time by optimally utilizing the available resources and minimizing the resource consumption (Kansal et al., 2012).

Load balancing algorithms are implemented in load scheduler. A scheduling system may contain a centralized scheduler or distributed schedulers. The central load balancing algorithms in clouds are commonly supported by a centralized controller that balances VMs to hosts. The central management algorithms for load balancing are simpler in implementation point of view. As central algorithms need to obtain the global information (utilization, load, connections information etc.). In these algorithms, the schedulers are centralized to monitor for better management; the best-fit algorithm is the typical example, as well as algorithms introduced in articles (Song et al., 2015; Xu et al., 2012). In each execution process of the centralized algorithms, the statuses of all hosts are collected, analyzed, reordered to provide information for VM allocation. In heuristic algorithms, like greedy algorithms, the centralized scheduler allocates VMs to the hosts with the lowest load. In meta-heuristic algorithms, like genetic algorithms (Hu et al., 2010), the centralized scheduler controls crossover, mutation, interchange operations to achieve better VM-host mapping results according to fitness functions. The centralized load balancing algorithms rely on a single controller to monitor and balance loads for the whole system, which may be the system bottleneck but in other hand the distributed algorithm eliminates the bottleneck pressure