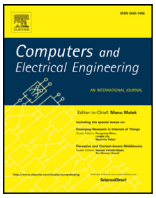
Computers and Electrical Engineering 000 (2018) 1–14

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng



# A combined forecast-based virtual machine migration in cloud data centers[☆]

Getzi Jeba Leelipushpam Paulraj[a,*], Sharmila Anand John Francis[b], J. Dinesh Peter[a], Immanuel Johnraja Jebadurai[a]

[a] Department of Computer Sciences Technology, Karunya University, Coimbatore, India
[b] Department of Computer Science, King Khalid University, Abha, Saudi Arabia

ARTICLE INFO

*Article history:*
Received 1 June 2017
Revised 9 January 2018
Accepted 10 January 2018
Available online xxx

*Keywords:*
Cloud computing
Combined forecasting
Artificial neural networks
Live virtual machine migration
Virtual machine placement

ABSTRACT

Live virtual machine (VM) migration improves the performance of cloud data center in terms of energy efficiency, fault tolerance, and availability. The workload handled by cloud data center is dynamic in nature. This increases the resource requirement of either the migrated virtual machine or collocated virtual machine at any time leading to further migration. Inappropriately handled live VM migration imposes severe application performance degradation. In this paper, a combined forecasting technique to predict the resource requirement of any virtual machine is proposed. Based on the current and predicted resource utilization, live migration is performed by Combined Forecast Load-Aware technique. Experiments were carried out to evaluate the performance of the proposed technique on live VM migration. The outcomes indicate that the proposed approach has minimized the number of migrations, energy usage, and the message overhead when compared with the existing state-of-art technique.

## 1. Introduction

Cloud computing provides Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS) on demand. The IaaS provides computing services through virtual machine instances in the cloud environment to bring down installation and maintenance cost for the customer. Virtualization aids cloud servers to service their customers efficiently. Live VM migration is one of the promising virtualization technologies. During live VM migration, the virtual machine from source server is moved to the destination server while the process continues running on the source server [1]. This phenomenon of migration is transparent to the customer enjoying the application over the cloud.

Server consolidation, load balancing, and fault tolerance are the key factors to improve the performance in the cloud [2]. Server consolidation minimizes the energy utilization by reducing the number of servers running in the cloud. Virtual machine instances of underutilized servers are migrated to the best-fit servers to minimize energy consumption in the cloud [3]. Load balancing distributes load across the servers. The virtual machine instances of heavily loaded servers are migrated to lightly loaded servers [4]. VM instances of servers experiencing downtime due to maintenance and failure are migrated.

The problem of Inter-VM interference is caused by the migrated virtual machine when it demands more resources due to sudden workload spikes. This instigates resource starvation and migrates the collocated VM to another server. Hence, the

---

[☆] Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. L. Bittencourt.

[*] Corresponding author.
*E-mail address:* getzi@karunya.edu (G.J.L. Paulraj).

https://doi.org/10.1016/j.compeleceng.2018.01.012

Please cite this article as: G.J.L. Paulraj et al., A combined forecast-based virtual machine migration in cloud data centers, Computers and Electrical Engineering (2018), https://doi.org/10.1016/j.compeleceng.2018.01.012

performance of the cloud is affected due to the increase in the number of migrations, energy consumption and bandwidth utilization [1]. Hence, a combined forecasting technique has been proposed to predict the behavior of workloads handled by virtual machines. Based on the prediction, the migration is performed using load aware migration technique.

The remainder of this paper is structured as follows. Section 2 reviews various related works in the literature. Section 3 explains the VM migration problem. Section 4 describes the proposed technique. Section 5 presents the simulation and performance evaluation. Finally, the conclusion is presented in Section 6.

## 2. Related works

A detailed study of the various approaches available in the literature for VM migration has been done. Many researchers have attempted optimization based VM placement techniques for migration. In [5], a Particle Swarm Optimization (PSO) based approach has been proposed. The power consumed by the server and the cost of migration are the two parameters taken into account for migration and consolidation of servers. A multi-objective based VM placement solution has been proposed in [6]. The resource wastage and power usage are optimized by considering factors such as CPU, RAM and bandwidth utilization. In [7], a multi-objective optimization based solution is given to avoid conflict among resource usages. This technique avoids hot spots and reduces energy consumption. These optimization techniques consider only the current resource utilization. Nonetheless, other criterion such as sudden spikes in cloud workload is not taken into account. Hence, these methods are not desirable for the real-time implementation.

In [8], the Entropy resource manager takes the responsibility of VM allocation and migration based on constraint programming. In [9], the author devises demand, communication and availability constraints. The demand planner places the virtual machine based on the resource usage and the structural planner places VM to the nearer server. In these techniques, additional modules have been included for VM placement. These modules increase the overhead and communication cost considerably. Higher overhead and communication cost are not desirable in the practical cloud environment. In [10], data center constraints are considered, and an effort is made to optimize the communication and network traffic. This technique works only based on the network resource parameters and it has not included the critical parameters such as CPU, Memory and Disk utilization.

In [11], event invoker receives information about the virtual machines running in the cloud. An assignment algorithm runs in the concentrate manager to command the event invoker about the VM Placement. However, this technique is based on energy parameter of the servers. It does not bring into account the resource constraints. In [12], the maximum number of VMs is allotted to a server. The physical machines are sorted in decreasing order and the virtual machines are packed using cloud broker. This technique concentrates on the initial placement of virtual machines. Simply, this methodology has not offered any solution for the VM migration. These techniques lack clarity on efficient migration techniques.

In [13], the algorithm considers hop reduction, energy saving, and load balancing. VMs are partitioned and placed on servers considering only their resource demands. Then open flow controller balances the load. In [14], the author analyzes the historical data and plans the future behavior of the VM. The VMs are allocated and reallocated based on the best-fit decreasing optimization technique. In [15], minimal cut hierarchical clustering is used with the best-fit decreasing algorithm to place the VM. These techniques also fail to include the behavior of sudden rise in the workload for efficient VM migration. In [16], game theory based VM placement scheme is proposed. However, this scheme is based on the traffic and the energy parameters. But it is recommended to include resource parameters for optimal VM migration technique.

## 3. Virtual machine migration problem

Let $S=\{S_1, S_2 ... S_n\}$ represent the set of physical servers in the data center and $VM=\{VM1, VM2, ....VMi\}$ represent the set of virtual machines hosted in the physical server. Let $VM_i^n$ represent the ith virtual machine in the nth server. Consider the case that migration is initiated in $VM_1^1$. The current CPU, Memory and Disk requirement of $VM_1^1$ are represented by $\Omega_{\text{cpu}}$, $\Omega_{\text{memory}}$, and $\Omega_{\text{disk}}$ respectively. CPU requirement is presented in terms of the number of cores required to complete a job. Similarly, memory requirement and Disk requirement refer to the space required in terms of bytes to complete a job within the stipulated time. The migration controller takes into account the resource requirement of $VM_1^1$ and migrates it to target server $S_T$. After migration, the CPU, Memory and Disk availability of $S_T$ becomes $\omega_{\text{cpu}}$, $\omega_{\text{memory}}$, and $\omega_{\text{disk}}$ respectively. The Energy consumption ($E_{\text{mig}}$) in joules during this migration is given by the Eq. (1) [17].

$$E_{mig} = 0.512\beta_{mig} + 20.165 \tag{1}$$

$\beta_{\text{mig}}$ represents the network traffic during migration process.

The load on the physical servers in the data center is always dynamic [18]. For instance, Fig. 1 shows the percentage of CPU utilization over different time instances for a PlanetLab server [19]. These are utilization values collected from PlanetLab server in every five minutes. It is evident that the CPU utilization is not constant over time. There are sudden rises in the CPU utilization. When there is a sudden rise in the load, the resource requirement of $VM_1^1$ is increased by $\Delta_{\text{cpu}}$, $\Delta_{\text{memory}}$, and $\Delta_{\text{disk}}$. Hence the resource requirement of $VM_1^1$ is escalated to $\Omega_{\text{cpu}}+\Delta_{\text{cpu}}$, $\Omega_{\text{memory}}+\Delta_{\text{memory}}$, and $\Omega_{\text{disk}}+\Delta_{\text{disk}}$. Server $S_T$ experiences resource scarcity once $\Omega_{\text{cpu}}+\Delta_{\text{cpu}}$, $\Omega_{\text{memory}}+\Delta_{\text{memory}}$, and $\Omega_{\text{disk}}+\Delta_{\text{disk}}$ values exceed $\omega_{\text{cpu}}$, $\omega_{\text{memory}}$, and $\omega_{\text{disk}}$ respectively. This problem occurs due to the improper migration of the virtual machine.

Please cite this article as: G.J.L. Paulraj et al., A combined forecast-based virtual machine migration in cloud data centers, Computers and Electrical Engineering (2018), https://doi.org/10.1016/j.compeleceng.2018.01.012