Resource Allocation in Cloud Data Centre for Enhancing Energy Efficiency

Neha Solanki

Department of Computer Science and Engineering, Jai Narain Vyas University
Jodhpur, India

Abstract *- In cloud computing various cloud consumers demand variety of services as per their dynamically changing needs. So it is the job of cloud computing to avail all the demanded services to the cloud consumers. Hence the cloud computing environment required a powerful resource allocation and scheduling mechanism to make it reliable. In this paper two resource allocation strategies are discussed.*

Keywords — *First-Fit Decreasing algorithm, Best-Fit Decreasing algorithm, Genetic algorithm.*

I. Introduction

Several large companies, such as Amazon, Google, Yahoo!, Microsoft, IBM, and Sun have developed their own cloud platforms for consumers and enterprises to access the cloud resources through services. The cloud computing is the concept of provide the virtualized resources to the multiple user's at a time.

Virtualization is making one computer appear to be one or multiple virtual computer systems. The basis for virtualization is virtual machine monitor (VMM) or the hypervisor as software layer that provides the illusion of real hardware for multiple virtual machines (VMs). VMM supports the creation and execution of multiple VMs on one physical machine, and multiplexes a single physical resource onto multiple virtual resources. Each VM has the operating system (called the guest OS) and applications that run on the VM's own virtual resources such as virtual CPU, virtual memory, virtual disks and virtual network card [1] (Fig 1).

In data centers, VMs can be used as resource containers that provide more fine-grained resource allocation than physical machines. Data center operators then allocate VMs to their customers, not physical machines. Because of its simpler abstraction, VMs can also be migrated from a physical machine to another without suspending their execution.

II. Resource Allocation in Cloud

For resource allocation, three types of resources (RAM, CPU and Bandwidth) are taken into account. Recent studies [2, 3] have shown that, even only considering CPU and memory resources in resource allocation for cloud computing, the problem is still NP complete, and adding network bandwidth into the mix will further increase the complexity, so this method uses a constraint programming approach to

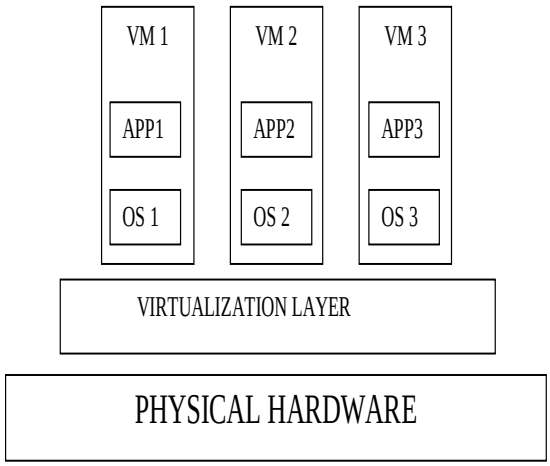

Fig 1: Virtualization in Cloud

formulate and solve the problem of the resource allocation in cloud computing. By solving constraint satisfaction problem using constraint programming solver, we can get the optimized number of physical machines (PMs) that host virtual machines (VMs) [4]. First/Best Fit Decreasing and Genetic Algorithm with Multiple Fitness resource allocation are discussed here.

A. First/Best-Fit Resource Allocation

The purpose of first-fit algorithm is to find as few PMs as possible to satisfy all the VM's requirements. We keep a dynamic sorted list in whole process of FFD running.

Algorithm for FFD(First Fit Decreasing)

Input: PMList and VMList with their underlying resource capacity (CPU, RAM, BW).
Output: Allocation of VMs

Step1. Clear the usedPMlist and copy all the PMs to unusedPMlist.

Step2. For each PM, calculate its X

$$x_j = r_j^{CPU} * p^{CPU} + r_j^{RAM} * p^{RAM} + r_j^{BW} * p^{BW}$$

Step3. For each VM, check the first machine of usedPMlist. If the PM can satisfy VM's requirements,

assign the VM to it, and reinsert the PM to usedPMlist.

Step4. If cannot, find one PM which can satisfy the VM's requirement in all the PMs on unusedPMlist and assign the VM to it. And remove the PM from unusedPMlist to usedPMlist.

Step5. If not existed such PM in unusedPMlist, the allocation fails and exits.

BFD algorithm is like to FFD; the difference between BFD and FFD is that in the step 3 BFD check all the PM of usedPMlist, while FFD only check the first PM. The algorithm is shown below:

Algorithm for BFD(Best Fit Decreasing)

Input: PMList and VMList with their underlying resource capacity (CPU, RAM, BW).
Output: Allocation of VMs.

Step1. Clear the usedPMlist and copy all the PMs to unusedPMlist.

Step2. For each PM, calculate its X

$$x_j = r_j^{CPU} * p^{CPU} + r_j^{RAM} * p^{RAM} + r_j^{BW} * p^{BW}$$

Step3. For each VM, check the all machine of usedPMlist. If the PM can satisfy VM's requirements, assign the VM to it, and reinsert the PM to usedPMlist.

Step4. If cannot, find one PM which can satisfy the VM's requirement in all the PMs on unusedPMlist and assign the VM to it. And remove the PM from unusedPMlist to usedPMlist.

Step5. If not existed such PM in unusedPMlist, the allocation fails and exits.

B. Genetic Algorithm with Multiple Fitness Resource Allocation

This cloud computing resource allocation policy is based on genetic algorithm with multiple fitness, which aims at solving the problem of resource optimization within a cloud [5]. Traditionally, only one fitness value is considered in the genetic algorithm [6]. While in this algorithm, three sub fitness values are considered.

The base of this algorithm is that the three load dimensions: CPU load, network throughput and disk I/O load of all the virtual machines carried on one specific physical machine. The flow diagram of the resource allocation policy is as Fig. 2. Firstly, the system fetches the load information and original placement information of all the virtual machines for the next stage. Then the HGA module is executed. Secondly, the result of the HGA module is compared with the original placement information to decide whether the result is satisfied enough. The criteria for judgment could be a threshold of the number of physical machines that will be saved or if the utilization is raised up to a certain level and so on. Due to different emphasis, the threshold could be different. Finally, another judgment is conducted, to judge whether the migration cost is higher than the benefits of the allocation. Only if the migration cost is low enough that the migration can be actually conducted.

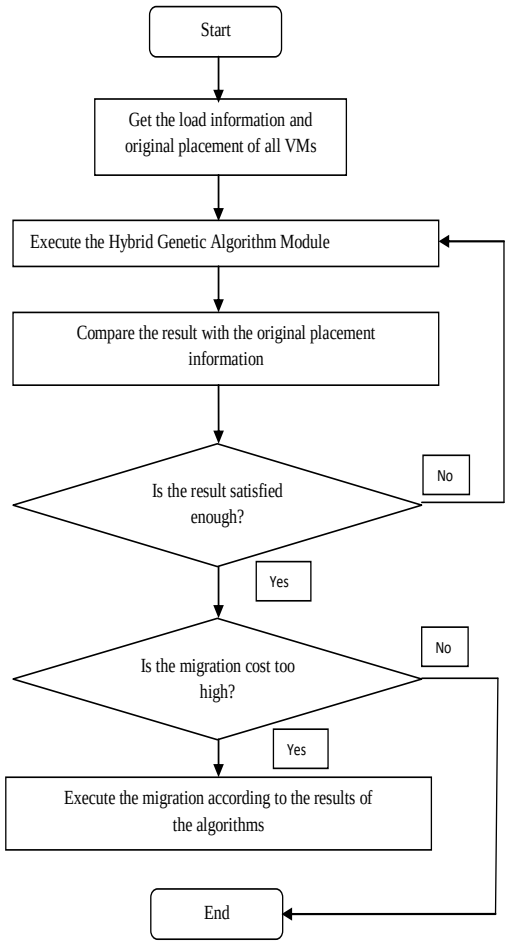

Figure 2: Flow Diagram of Resource Allocation through Genetic Algorithm

III. CONCLUSION

First/Best-Fit resource allocation and Genetic algorithm with multiple fitness resource allocation, these two strategies focus on load balancing (RAM, CPU and Bandwidth). They optimize the resource allocation in cloud data centers. The ultimate goal of resource allocation in cloud computing is to maximize the profit for cloud providers and to minimize the cost for cloud consumers.

REFERENCES

[1] Qiang Li; Qinfen Hao; Limin Xiao; Zhoujun Li; Adaptive Management of Virtualized Resources "Information Science and Engineering (ICISE), 2009 1st International Conference

[2] Hien Nguyen Van, Frédéric Dang Tran, Jean-Marc Menaud, Performance and Power Management for Cloud Infrastructures, pp.329-336, 2010 IEEE 3rd International Conference on Cloud Computing, 2010.

[3] F. Hermenier, X. Lorca, J.-M. Menaud, G. Muller and J. Lawall. Entropy: a Consolidation Manager for Cluster. In

proc. of the 2009 International Conferenceon Virtual Execution Environments (VEE'09), Mar.2009.

[4] N. Jussien, G. Rochart, and X. Lorca, the CHOCO constraint programming solver, in: Proceedings of CPAIOR'08 workshop on Open-Source Software for Integer and Contraint Programming (OSSICP'08), Paris, France, June 2008.

[5] Efficient Resource Allocation in Cloud Data Centers Through Genetic Algorithm. Ehsan Arianyan, Davood maleki, Alireza Yari, Iman Arianyan. 6'th International Symposium on Telecommunications (IST'2012).

[6] A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment. Jinhua Hu Jianhua Gu Guofei Sun Tianhai Zhao. DOI 10.1109/PAAP.2010.65 IEEE 3rd International Symposium on Parallel Architectures, Algorithms and Programming.