Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 78 (2016) 491 – 498

Procedia Computer Science

International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA

A Survey of Virtual Machine Placement Techniques in a Cloud Data Center

Zoha Usmani[a], Shailendra Singh[b], Senior Member IEEE, a*

[a]*PG Scholar,Departmnet of CEA, NITTTR, Bhopal 462002, India*
[c]*Professor& Head, Department of CEA, NITTTR, Bhopal 462002,India*

Abstract

Energy consumption of massive-scale cloud data centers is increasing unacceptably. There is a need to improve the energy efficiency of such data centers using Server Consolidation which aims at minimizing the number of Active Physical Machines (APMs) in a data center. Effective VM placement and migration techniques act as a key to optimum consolidation. Many of the recently proposed techniques realize dynamic consolidation while optimizing the VM placement. This paper presents a comprehensive study of the state-of-the-art VM placement and consolidation techniques used in green cloud which focus on improving the energy efficiency. A detailed comparison is presented, revealing pitfalls and suggesting improvisation methods along this direction.

Peer-review under responsibility of organizing committee of the ICISP2015

Keywords: Cloud Computing; Data Center; Green Computing; Server Consolidation; Virtual Machine Placement

1. Introduction

The highly flexible IaaS (Infrastructure as a Service) model of cloud computing has been rapidly adopted by many enterprises as an alternative to raise their Total Cost of Ownership (TCO), providing optimal utilization of resources and money. Today a lot of companies like Amazon with its EC2 (Elastic Compute Cloud) are switching towards greening their data centers, i.e. trying to use least numbers of actively running servers. This calls for

* Corresponding author. Tel.: +91-997-705-3137.
*E-mail address:*zohausmani05@gmail.com

Peer-review under responsibility of organizing committee of the ICISP2015
doi:10.1016/j.procs.2016.02.093

appropriate solutions like Virtualization which acts as a backbone behind the cloud computing technology. It enables sharing of computer hardware by partitioning the computational resources. A small software program, i.e. the Hypervisor or Virtual Machine Monitor (VMM) manages and controls all VM related operations. Live migration of virtual machines facilitates load balancing and consolidation. In a datacenter, often many services only need a small portion of the total available resources. This can lead to a scenario in which several virtualized servers operate and consume an increased amount of space and resources than expected and that it cannot be justified by their workload. This problem is referred to as *server sprawl*.

To prevent such wastage of resources, multiple virtual machines are packed on fewest possible physical servers and rest of the extraneous servers are turned down to sleep mode (low-power state). This approach can avoid server sprawl and is termed as *Server Consolidation*. The reduction in the number of servers has a noticeable benefit on data centers by improving system availability, reducing infrastructure complexity and of course saving energy and money. VM consolidation can achieve its goal of increasing the amount of suspended servers, but in an IAAS environment, there may arise some problems which might influence the energy efficiency of cloud.

- The first problem is the trade-off between the performance, energy consumption & the resource utilization of running VMs. Since they need to compete for the resources provided by their corresponding PM and share them too. The resources being PM's—*CPU, main memory* and *I/O capabilities*, the server suspensions and performance degradations can increase the execution time which in turns decreases the energy savings.
- Another concern is the overhead caused by continuous live migration of VMs during consolidation, due to the continuously changing resource demands. Whenever a VM is migrated, its CPU state, main memory, storage & network connections are also taken care of.
- The next concern lies in prediction of energy consumed by the server. This is quite complicated and requires energy-performance profiling techniques, but these are unacceptable due to their respective overheads.
- In particular, cloud data center resources (e.g. CPU, memory, network bandwidth and storage) need to be allocated with equal focus on reduction of energy usage as on satisfaction of Quality of Service (QoS) requirements specified by users via Service Level Agreements (SLAs).
- Final concern is that bringing about an efficient server consolidation is quite a complex NP-Hard problem.

Keeping these challenging aspects in mind many effective approaches of server consolidation have been introduced till date. These state-of-the-art techniques address some important issues like physical resource heterogeneity or dynamic nature of virtual machines and workloads. In this paper, we discuss the VM scheduling techniques which aim at server consolidation mainly focusing on the discussion of VM placement algorithms used by these scheduling approaches. The virtual machine scheduling mechanism is used to consolidate a data center, i.e. to reduce the number of active physical machines. As discussed earlier, dynamic server consolidation can improve the energy efficiency by optimum utilization of available resources.

1.1. Server Consolidation Steps

To solve the complex problem of dynamic server consolidation and to provide decentralization, it has been divided into four main events or steps [19] discussed as follows:

- **Host Overload Detection**: The scheduling technique must set a threshold limit in order to decide when a certain host/server is over-utilized. This limit can be termed as *'Hot Threshold'* and when this limit is crossed, some of this host's VMs need to be migrated to other hosts, [27].
- **Host Under-load Detection**: If a certain server is under-utilized, i.e. it has reached below the *'Cold Threshold'* (a scenario just opposite to host overload) the aim of server consolidation is to identify that server and migrate all of its virtual machines to other active hosts, [13]. Thus the under-utilized server is freed up & it can be switched to sleep/idle mode to save power.
- **VM Selection and Migration**: Appropriate candidates (VMs) are selected either from overloaded or under-loaded host for migration.
- **VM Placement**: The VM(s) selected in previous step is then placed on some other physical machine according to a suitable VM to PM mapping criteria, [29].

With the help of live migration of VMs, Server Consolidation aims at achieving—least possible number of Active physical machines, packing these Active PMs with VMs as tightly as possible to increase energy efficiency