

Recovering Host Overloading in Cloud using Virtual Machines

N.Suresh Kumar
IInd Year - M.E CSE,
Srinivasan Engineering College,
Perambalur, Tamilandu India

B. Karthikeyan, Ph.D
Professor/ECE,
Srinivasan Engineering College,
Perambalur, Tamilandu India

ABSTRACT

Virtualization provides an efficient solution to the objectives of the cloud computing paradigm by facilitating creation of Virtual Machines (VMs) over the underlying physical servers, leading to improved resource utilization and intellection. Virtualization proposes to create a virtual version of a device or a resource likely to virtualize a server, a storage space, operating system or even network where the mechanism divides the resource into one or more execution environments. To analyze the behavior of a Cloud data center by means of various aspects like availability, utilization, responsiveness and waiting time. The factors that a cloud provider must take into account are elasticity, scalability, live migration of VMs and performance isolation. Live migration of VMs, the process of dynamically transferring a virtual machine across different servers, has proved to represent a new opportunity to enable responsive and dynamic resource management in modern data centers. Migrations of VMs is performed on the basis of overloading that occurs in physical servers in cloud data centers, which leads to relatively less performance degradation. The SLA violation is also less compared to other techniques and this means the cloud provider will incur less cost from VM migrations. Further, MHOD algorithm performs host selection and VM reallocation quicker than the existing algorithms. Maintenance of physical servers can be efficiently achieved through our algorithm as it leads to efficient consolidation of VMs.

KEYWORDS

Virtual Machine, Cloud Data Centers, Markov Host Overload Detection

1. INTRODUCTION

Cloud Computing involves aggregation of computing, storage and network resources into a single entity called “cloud” into which location-independent computing is performed. Cloud computing is an evolution of the virtualization, Service-Oriented Architecture (SOA) and Utility Computing. Virtualization in computing is the creation of a virtual version of an Operating system, hardware platform, a storage device or network resources. For example, a single physical server or machine can be sliced into various virtual machines, each embodying various resources like memory, disk, CPU cores, etc. Virtualization can be viewed as part of an overall IT enterprise that includes autonomic compute, a development in which the IT environment will be able to manage itself based on utility computing. The aim of virtualization is to centralize administrative tasks while improving Cloud Computing and virtualization are synonymous. Cloud computing is based upon vitalizing and allocating storage and network services in a shared multi-tenant environment. Virtualization is a key enabler for cloud computing. At the meanwhile, cloud

computing is also a powerful force by pulling virtualization into the enterprise.

2. RELATED WORK

Anton Beloglazov and Rajkumar Buyya [1] has stated in his findings to two lines of research, namely, to that of application placement and to that of load balancing in processor networks. A scheme for decentralized utility maximization considering a single type of resource. There, the authors present an optimal solution under the assumption that the demand of an application can be split over several machines. Their notion of utility is different. Their solution has limited applicability in context, multiple resources that need to be allocated on the same machine and the demand for memory cannot be split between machines.

Bobroff *et al* [2] states to perform trace analysis on commercial web servers and outline a method to identify the servers that are good candidates for dynamic placement. However, none of these studies provide a characterization of the inter-relationship between various workloads, as required for static consolidation.

The approach to minimize migrations over consecutive control interval. Some researchers propose to limit the capacity requirement of an application workload to a percentile of its demand [3]. This does not take into account the impact of sustained performance degradation over time on user experience as our required capacity definition does.

Dynamic power management (DPM) is a flexible and general design methodology aiming at controlling performance and power levels of digital circuits and systems, by exploiting the redundancy of their components. A scheme is provided with a power manager that monitors the overall system and component states and controls the state transitions. The control procedure is called power management policy[4],[5],[6].

3. SYSTEM DESIGN

Role of Virtualization

The new technology underlying this is the system virtual machine that allows multiple instances of an operating system and associated applications to run on single physical machine. In network, Delivering on-demand, is termed Infrastructure as a Service (IaaS).

The IaaS offers a great impact on the today's market, which allow quickly provisioning and deployment of applications and their underlying operating systems onto an infrastructure that expands and contracts as needed to handle the load. The flexibility defines the resources to come in handy. Thus they can be better matched to the demand on the applications.

Storage Virtualization Benefits

Storage virtualization leads to a better way to combine all those drives into one central manageable resource, will be beneficial for users via money and time savings. Optimizing performance and improving storage utilization, understanding or dynamic provisioning, creating a dynamic provisioning pool, space saving and wide striping for performance improvements virtualization can improve several areas of managing a SAN (Storage Area Network), including:

- Provides cost-effective ways to eliminate single points of failure in storage.
- Effective Quality of Service (QoS) by managing and improving performance in real time.
- Failure recovery and data archiving.
- Improving utilization across the data stored in the data center
- Better scalability and flexibility on the network.

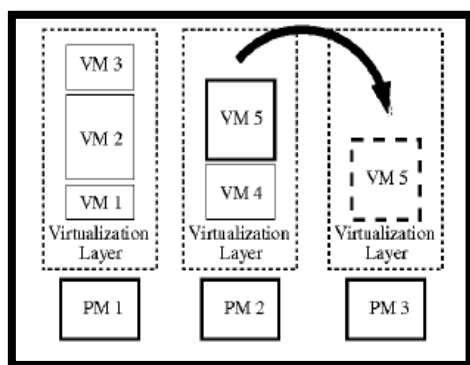


Figure 1: Virtualized servers with live migration capability

Host Overload Detection

Each host periodically executes an overload detection algorithm to de-consolidate VMs when needed in order to avoid performance degradation and SLA violation. This section describes several heuristics proposed for the host overload detection problem.

A novel Markov chain model that allows a derivation of a randomized control policy that optimally solves the problem of maximizing the mean time between VM migrations under an explicitly specified Quality of Service (QoS) goal for any known stationary workload and a given state configuration in the online setting.

1. Deciding if a host is considered to be overloaded, so that all VMs should be migrated, and the host should be switched to a low-power consumption mode.
2. Deciding if a host is considered to be overloaded, in that case, VMs should be migrated from it to other active or reactivated hosts to avoid violating the QoS requirements.
3. Selecting VMs to migrate from an overloaded host.
4. Placing VMs selected for migration on other active or reactivated hosts.

Algorithm 1: The MHOD Algorithm

Input: A CPU utilization history

Output: Make decision to migrate on VM

- 1: **if** the CPU utilization history size > TI **then**
- 2: Convert the last CPU utilization value to a state
- 3: Invoke the MultiSize Sliding Window estimation to obtain the estimates of transition probabilities
- 4: Invoke the MHOD-OPT algorithm
- 5: **return** the decision returned by MHOD-OPT
- 6: **end if**

7: return false

The models and algorithms proposed in this paper are suitable for both single core and multi-core CPU architectures. The capacity of a single core CPU is modeled in terms of its clock frequency F .

A VM's CPU utilization (u_i) is relative to the VM's CPU frequency (f_i) and is transformed into a fraction of the host's CPU utilization (U). These fractions are summed up over the N VMs allocated to the host to obtain the host's CPU utilization by the following equation (1).

$$U = F \sum_i^N f_i u_i \dots (1)$$

4. PROPOSED WORK

A technology that is able to improve the utilization of server resources, and thus, reduce power consumption, is virtualization of computing resources. Virtualization introduces an abstraction layer between an OS and hardware. Physical resources can be split into a number of logical slices called Virtual Machines (VMs). Each VM can accommodate an individual OS creating for the user a view of a dedicated physical resource and ensuring the performance and failure isolation between VMs sharing a single physical machine. The problems of host overload detection as a part of dynamic VM consolidation. Determining when it is best to reallocate VMs from an overloaded host is an aspect of dynamic VM consolidation that directly influences the resource utilization and Quality of Service (QoS) delivered by the organization is shown in Figure 2.

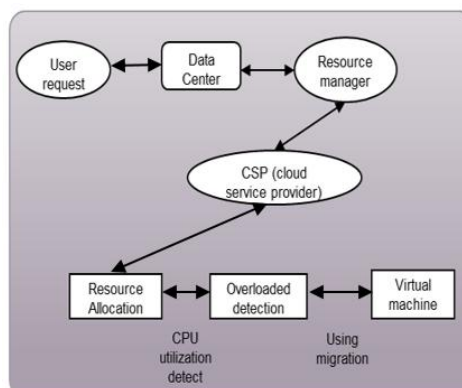


Figure 2: Virtualized servers with live migration capability

The authority on the QoS is explained by the fact that server overloads cause resource shortages and performance degradation of applications. In present solutions to the problem of host overload detection are generally heuristic-based, or rely on statistical analysis of past data. The limits of these approaches are that they lead to sub-optimal results and do not allow explicit specification of a QoS goal.

A novel approach that for any known stationary workload and a given state configuration optimally solves the problem of host overload detection by maximizing the mean inter-migration time under the specified QoS goal based on a Markov chain model. The algorithm is heuristically adapted to handle unknown non-stationary workloads using the MultiSize Sliding Window workload estimation technique.

Key contributions

1. An analytical model showing that to improve the quality of VM consolidation, it is necessary to maximize the mean time between VM migrations initiated by the host overload detection algorithm.
2. An optimal offline algorithm for host overload detection, and proof of its optimality.
3. A novel Markov chain model that allows a derivation of a randomized control policy that optimally solves the problem of maximizing the mean time between VM migrations under an explicitly specified QoS goal for any known stationary workload and a given state configuration in the online setting.
4. A heuristically adapted algorithm for handling unknown non-stationary workloads using the Multisize Sliding Window workload estimation approach, which leads to comparable to the best benchmark algorithm performance in terms of the inter-migration time, while providing the advantage of precise specification of a QoS goal.

5. CONCLUSION

Virtualization, in computing is the creation of a virtual i.e., rather than actual version of a storage device or network resources. By using some interfaces we can access the data in cloud. This paper gives about the cloud data management interface by using storage virtualization mechanism. The open cloud computing interface is an emerging standard for interoperable interface management in the cloud. To implement the MHOD algorithm as an extension of the VM manager within the OpenStack Cloud platform to evaluate the algorithm in a real system as a part of energy-efficient dynamic VM consolidation.

6. REFERENCES

- [1] Anton Beloglazov and Rajkumar Buyya, "Managing Overloaded Hosts for Dynamic Consolidation of Virtual

Machines in Cloud Data Centers Under Quality of Service Constraints", July 2013.

- [2]. N. Bobroff, A. Kochut, and K. Beaty, "Dynamic placement of virtual machines for managing SLA violations," in Proc. of the 10th IFIP/IEEE Intl. Symp. on Integrated Network Management (IM), 2007, pp. 119–128
- [3]. D. Gmach, J. Rolia, L. Cherkasova, G. Belrose, T. Turicchi, and A. Kemper, "An integrated approach to resource pool management: Policies, efficiency and quality metrics," in Proc. of the 38th IEEE Intl. Conf. on Dependable Systems and Networks (DSN), 2008, pp. 326–335.
- [4]. Christian Strack "Performance and Power Management for Cloud Infrastructures" Philipps-University Marburg, Germany -Department of Mathematics and Computer Science, Distributed Systems Group.
- [5] J. Koomey, Growth in data center electricity use 2005 to 2010. Oakland, CA: Analytics Press, 2011.
- [6] Gartner, Inc., Gartner estimates ICT industry accounts for 2 percent of global CO2 emissions. Gartner Press Release (April 2007).
- [7] R. Nathuji and K. Schwan, "Virtualpower: Coordinated power management in virtualized enterprise systems," ACM SIGOPS Operating Systems Review, vol. 41, no. 6, pp. 265–278, 2007.
- [8] A. Verma, P. Ahuja, and A. Neogi, "pMapper: Power and migration cost aware application placement in virtualized systems," in Proc. of the 9th ACM/IFIP/USENIX Intl. Conf. on Middleware, 2008, pp. 243–264.