ISSN:0975-9646

Noumankhan Pathan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1833-1835

Virtual Machine Placement in Cloud

Noumankhan Pathan
Information Technology Department
SGGS IE&T, Nanded, India

Balaji Shetty
Information Technology Department
SGGS IE&T, Nanded, India

***Abstract*—Unlike in traditional datacenters in which applications are tied to specific server, virtual machines are deployed on a single physical server in cloud computing, which results in consolidation of multiple applications on to a minimum number of physical servers. This enable in an increase in server utilization and energy saving. In this paper we have discussed approaches related to mapping of virtual machines to physical machine in cloud computing.**

***Keywords*— Cloud Computing, Virtual Machine, VM Placement, Bin Packing, Genetic Algorithm, VM Migration.**

I. INTRODUCTION

Virtualization allows consolidation of services onto lesser number of physical servers than originally required [1]. The applications are used to tie down to specific physical servers that are often overprovisioned in order to serve the complex requirements of resources of enterprise services and it is hoped that it will handle unexpected surges in resource demands. As a result, the utilization of resources on any server is typically very low. This invites high operational costs and heavy investments, also over power utilization and wastage of space and a significant management overhead.

Cloud computing offers a means to decouple the application activities from the physical resources required. Cloud computing is one such paradigm with which a shared pool of resources (networks, servers, storage, applications, and services) can be accessed conveniently and on-demand, and that can be rapidly provisioned or released with minimal management effort or service provider interaction[2]. This results in consolidation of multiple applications onto minimum number of physical servers enabling in an increase in server utilization. Such decoupling of resources is achieved by the concept of a 'virtual machine' which encompasses of an application with a specific set of functionalities. Physical resources are made available to the virtual machine by a guest operating system running on each physical machine. The virtual machine runs over this guest operating system which also provides facilities for creation, destruction and migration of virtual machines. Xen [3], KVM [4] are some of the products that are widely used for management of such virtual machines.

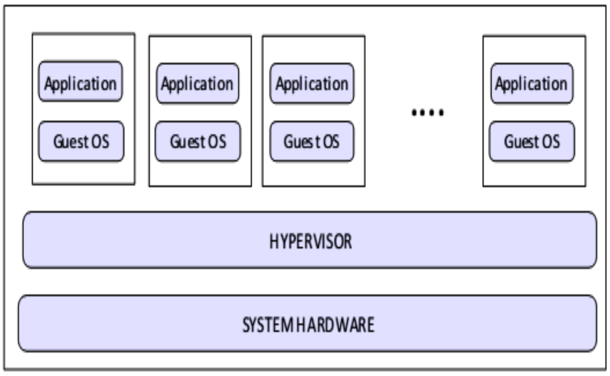

FIGURE 1: MULTIPLE VMS ON A PHYSICAL SERVER

Due to the ability to run multiple virtual machines onto the same physical servers while also being able to migrate them seamlessly across different servers, new challenges have come into the light. These challenges include balancing load amongst all physical machines, figuring out which VMs to place on which PMs and handling sudden surges in resource demands and cost [5].

Moving an application running in a physical non-virtualized environment to a virtualized one involves three main stages:

1) *Application Profiling*: during which the application is profiled in accordance with its physical environment to get its resource utilization,

2) *Generation of Virtual Machine Configurations*: during which the amount of resource utilizations are used to generate configurations for virtual machines after consideration of factors like virtualization overheads, etc. and

3) *Virtual Machine Placement*: which takes these generated virtual machine configurations as input and try to figure out the best possible mapping of the VMs onto the physical servers. In this paper we will be discussing this stage.

In the environment where there is small number of virtual and physical machines, it might be possible for an operator to manage the placement of virtual machines. However, as the number of VMs and PMs increases, situation becomes uncontrollable and automation becomes necessary. In the reminder of this paper we will be discussing different approaches of mapping virtual machine to physical machine.

II. BIN PACKING

The bin packing problem is a combinatorial NP-hard problem. In bin packing problems, the objects of different volumes must be packed into a finite number of bins of capacity M in a way that minimizes the number of bins used. Many variations of this problem are present, such as 2D packing, linear packing, packing by weight, packing by cost, and so on. The applications of these problems include, filling up containers, loading trucks with weight capacity, and creating file backup in removable media. Most of the reliable bin packing algorithms use heuristics to accomplish results. This provides a solution, which, though very good in most cases, may not be the optimal solution. For example, the first fit algorithm provides a fast but often non-optimal solution, involving placing each item into the first bin in which it will fit. It requires (n log n) time, where n is the number of elements to be packed.

The VM placement problem can be designed as a bin packing problem by considering Physical machines as bins and the VM's to be placed can be considered as objects to be filled in the bin[6, 7, 8].

III. GENETIC ALGORITHM

Genetic algorithm is based on biological concept of generation of the population, and survival of the fittest. Genetic algorithms are inspired by Darwin's theory of Evolution. The field of Genetic algorithm is a rapid growing area of Artificial Intelligence.

The concept of Generation of Population [9, 10] for Genetic algorithms is explained as follows:

A. *Initial Population*

Initial population is the set of all the individuals that are used in the Genetic Algorithm to find out the optimal solution. Each solution in the population is referred as an individual. And every individual is referred as a chromosome for making it appropriate for the genetic operations. From the initial population the individuals are selected and some actions are applied on those to form the next generation. The mating chromosomes are selected based on some specific standards.

B. *Fitness Function*

The output of any individual depends on the fitness value. It is the measure of the superiority of an individual in the population. The fitness value shows the performance of an individual in the population. If the fitness value is large, then the performance of an individual is better. Depending on the fitness or function value, individuals will survive or die. Hence, the fitness function is the important factor in the Genetic Algorithm.

C. *Selection*

Selection mechanism is used to select a best solution for the next generation based on the survival of the Darwin law. This operation is the guiding network for the Genetic Algorithm based on the performance. There are various selection approaches to select the best chromosomes e.g. roulette wheel, Boltzmann strategy, tournament selection, selection based on rank, etc.

D. *Crossover*

Crossover/hybridizing operation can be realized by selecting two parent individuals and then creating a new individual tree by alternating and reforming the portions of those parents. Hybridization operation is a guiding process in the genetic algorithm and it enhances the searching mechanism.

E. *Mutation*

After crossover Mutation takes place. It is the genetic operator that introduces genetic diversity in the population. Mutation takes place whenever the population likely to become homogeneous due to frequent use of reproduction and crossover operator. It occurs during evolution according to a user-defined mutation probability, usually set to fairly low. Mutation changes one or more gene values in chromosome from its starting state. This can produce the completely new gene values being mixed with the gene pool. With this new gene values, the genetic algorithm may be able to produce the better solution than was previously.

The VM placement problem can be designed as a genetic programming problem by representing the solution domain as the physical machines with a resource provisioning capacity. The fitness function can be defined over the number of bins in the solution. The aim would be to deliver a solution that is nearly optimal in terms of the number of bins used and the efficiency of packing of the bins [11].

IV. VM PLACEMENT VIA MIGRATION

In cloud computing, in order to efficiently utilize data-center resources, consolidation of existing systems are done through virtualization solutions i.e. placing multiple virtual machines onto a single physical server. But extensive utilization of a single physical server may result in the performance degradation of the application (virtual machine). Solution to this problem is live VM migration.

Live migration is the movement of a virtual machine from one physical host to another while continuously powered-up. When properly carried out, this process takes place without any noticeable effect from the point of view of the end user. Live migration allows an administrator to take a virtual machine offline for maintenance or upgrading without subjecting the system's users to downtime. One of the most significant advantages of live migration is the fact that it facilitates proactive maintenance. If an imminent failure is suspected, the potential problem can be resolved before disruption of service occurs. Live migration can also be used for load balancing, in which work is shared among computers in order to optimize the utilization of available CPU resources [12].

The VM migration usually includes three main stages [13].

1. Deciding when redistribution is necessary.
2. Deciding which VM to be migrated.
3. Deciding where to place selected VM. i.e mapping of selected VM to a new Physical server.

V. Classification of VM Placement Approaches

Virtual machine placement is the process of mapping virtual machines to physical machines. In other words, virtual machine placement is the process of selecting the most suitable host for the virtual machine. The process involves categorizing the virtual machines hardware and resources requirements and the anticipated usage of resources and the placement goal. The placement goal can either be maximizing the usage of available resources or it can be saving of power by being able to shut down some servers. The autonomic virtual machine placement algorithms are designed keeping in mind the above goals.

VM placement algorithms can be broadly classified into two categories on the basis of their placement goal.

A. *Power Based Approach*

The main aim of these approaches is to map virtual machines to physical machines in such a way, so that the servers can be utilized to their maximum efficiency, and the other servers can be either hibernated or shut down depending on load conditions. This is also called as server consolidation since we seek to consolidate VMs on as few physical machines as possible.

B. *Application QOS Based Approach*

These algorithms manage the mapping of virtual machines onto physical hosts with the aim of maximizing the quality of service (QoS) delivered. By continuously monitoring virtual machine activity and employing advanced policies for dynamic workload placement [13], such algorithms can lead to better utilization of resources and less frequent overload situations eventually leading to savings in investments.

VI. Conclusion

In cloud computing environment virtual machines are deployed on physical servers so as to provide resource consolidation. Virtual machine placement approaches are very important as they results in providing desired QoS for a service provided by a cloud service provider, and also in managing energy consumption by servers. Therefore it is very important for a cloud service provider to choose a VM placement strategy that will result in better QoS and energy saving.

References

[1] Borja Sotomayor, Rubén S. Montero and Ignacio M. Llorente, Ian Foster, “Virtual Infrastructure Management in Private and Hybrid Clouds.”

[2] P. Mell and T. Grance, “Definition of Cloud Computing,”technical report, NIST, 2009.

[3] R. Wojtczuk, “Subverting the Xen Hypervisor,”Black Hat USA, 2008.

[4] I. Habib, “Virtualization with kVM,” Linux Journal, article 8, no. 166, 2008.

[5] Wu Ming, Yang Jian, Ran Yongyi, “Dynamic Instance Provisioning Strategy in an Iaas Cloud,” Proceeding of the 32nd Chinese Control Conference, July 26-28, 2013, Xi’an, China.

[6] Norman Bobroff, Andrzej Kochut, Kirk Beaty, “Dynamic Placement of Virtual Machines for Managing SLA Violations.”

[7] K. Mills, J. Filliben and C. Dabrowski, “Comparing VM-Placement Algorithms for On-Demand Clouds,” 2011 Third IEEE International Conference on Coud Computing Technology and Science.

[8] Pinal Salot, “A Survey of Various Scheduling Algorithm in Cloud Computing environment,” ISSN: 2319 – 1163, 131 – 135.

[9] Tarun goyal & Aakanksha agrawal, “Host scheduling algorithm using genetic algorithm in cloud computing environment,” International Journal of Research in Engineering & Technology (IJRET) Vol. 1, Issue 1, June 2013, 7-12.

[10] Rajkumar Buyya, Rajiv Ranjan, Rodrigo N. Calheiros,“Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges andOpportunities”.

[11] Hidemoto Nakada, Takahiro Hirofuchi, Hirotaka Ogawa, and Satoshi Itoh. “Toward virtual machine packing optimization based on genetic algorithm.” In IWANN '09: Proceedings of the 10th International Work-Conference on Artificial Neural Networks, pages 651-654, Berlin, Heidelberg, 2009. Springer-Verlag

[12] Margaret Rouse, “live migration,” TechTarget, Published: 31 Oct 2006

[13] Mauro Andreolini, Sara Casolari, Michele Colajanni, and Michele Messori, “Dynamic load management of virtual machines in a cloud architectures.”