Virtual Machine Algorithms

Renu Saini[1], Nishant Anand[2]

CBS Group of Institutions, Jhajjhar, Delhi (NCR)

Abstract- In cloud computing, there are many strategies used for virtual machine (VM) placement. Many such algorithms address distinct problems, such as initial placement, consolidation, or tradeoffs between honoring service-level agreements and constraining provider operating costs. Objectives for VM placement are to reduce the number of physical machines required, VM allocation time and to reduce resource and power wastage.

I. INTRODUCTION

There are a number of key technologies that make cloud computing possible. In this regard, power management in cloud data centers is becoming a crucial issue since it dominates the operational costs. Virtualization enables dynamic sharing of physical resources in cloud computing environments, allowing multiple applications to run in different performance-isolated platforms called virtual machines (VMs) in a single physical server. To the consumer, the cloud appears to be infinite, and the consumer can purchase as much or as little computing power as they need. From a provider's perspective, the key issue is to maximize profits by minimizing the operational costs. The emergence of cloud computing has made a tremendous impact on the information technology (IT) industry over the past few years, where large companies such as Amazon, Google, Salesforce, IBM, Microsoft, and Oracle have begun to establish new data centers for hosting cloud computing applications in various locations around the world to provide redundancy and ensure reliability in case of site failures. This technology also enables on-demand or utility computing—a just-in-time resource provisioning model in which computing resources such as CPU, memory, and disk space are made available to applications only as needed and not allocated statically based on the peak workload demand. Virtualization provides a promising approach through which hardware resources on one or more machines can be divided through partial or complete machine simulation, time-sharing, hardware and software partitioning into multiple execution environments, each of which can act as a complete system. The adoption and deployment of cloud computing platforms have many attractive benefits, such as reliability, quality of service and robustness. Moreover, power consumption in large-scale computer systems like clouds also raises many other serious issues including carbon dioxide and system reliability.
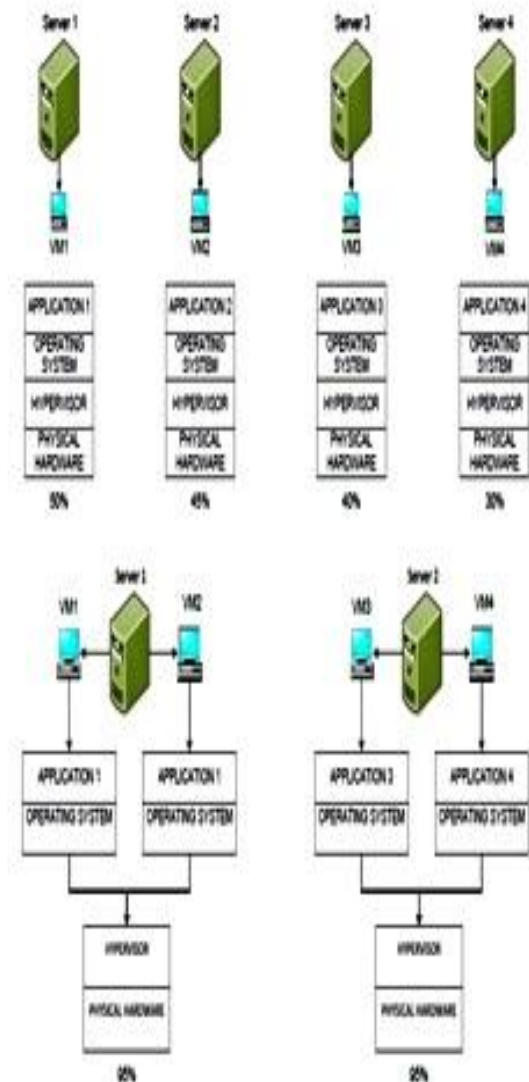
II. VM PLACEMENT ALGORITHMS

The following are the VM placement algorithms which are surveyed on.

A. Dynamic Server Allocation Problem (DSAP)

In cloud environment, VMs can be migrated to different servers over a time period for balancing the workload in datacenters.

By considering the history of workloads, DSAP is better when compared to assigning a VM to a particular server. So, the number of servers needed and allocation of VMs to servers in a time period are considered and VM placement is done in DSAP.

B. Static Server Allocation Problem (SSAP)

The main objective of SSAP is to reduce the number of servers used for VM placement because it will minimize the overall cost of servers. In the pre-processing step of SSAP, all virtual machines (VMs) are assigned to servers manually. The lower and upper bound capacity of VM specifies the minimum and maximum capacity that a VM can utilize respectively. Lower and upper bound for capacity are set for each VM. But SSAP cannot handle the situation when the workload of each VM varies over time.

C. Static Server Allocation Problem with variable workload (SSAPv)

SSAPv overcomes the disadvantage of SSAP by considering varying workloads of VM, when assigning VMs to servers. In SSAPv, a matrix is maintained for describing the maximum capacity of VM needed in a time t.

D. Multi-objective Ant Colony Optimization (ACO) Algorithm

In this multi-objective ACO algorithm, each ant constructs a solution for assigning VM to server. A multi-objective ACO algorithm to minimize SLA violation, resource wastage and power wastage in servers. By updating pheromone level, the optimal solution for VM placement was found. Finally, the pheromone update rule is applied. After the evaluation of constructed solution, if the constructed solution is in the list of best solutions, then its pheromone level is increased by increasing the pheromone evaporation rate ρ whose value lies between 0 and 1.

E. Novel Vector Based Approach for Static VM Placement

A novel vector based approach for static VM placement for minimizing the total number of servers required for VM placement. The set of VMs with their resource requirements and set of servers with their capacity are given as inputs for this algorithm. At last, the VM which has the least magnitude M on the current server is placed on the current server and this VM is marked as 'placed'. After placing each VM, the new RUV of the current server is calculated. The above steps are repeated until all VMs are placed in given set of servers.

If the server can host this VM, then the vector addition of resource imbalance vector (RIV) of that server and RIV of that VM is calculated and finally the magnitude M of this vector addition is also calculated. Each VM is taken and checked whether the server can host this VM. After that if there is no place for VM in current server and there is one or more VMs have to be placed, then a new server is introduced and it is checked whether the remaining VMs can be placed in this new server.

F. Novel Vector Based Approach for Dynamic VM Placement

The paper discusses a novel vector based approach for dynamic VM placement which is used when a new VM reaches. While selecting the potential servers, if the goal is to support load balancing then the servers with low resource utilization are selected and checked whether they support this new VM (checking of servers which support the new VM, starts from low to high resource utilization). If the goal is to provide server consolidation then the high utilization servers are selected for checking whether this server supports the new VM (checking of servers which support the new VM, starts from high to low resource utilization). After getting a list of potential servers, take each server and compute the vector addition of RIV of server and VM. Also, the magnitude M of the vector addition is calculated. The inputs for this algorithm are a set of servers with their current resource utilization and a new VM. The goal of this dynamic VM placement algorithm is either to provide load balancing or server consolidation. According to the goal (e.g., load balancing or server consolidation), a list of potential servers is selected for placing the new VM. After the vector addition and their magnitude M are calculated, the server which has the lowest magnitude M is marked as the server for the new VM.

G. VM Scheduler Algorithm

By this, the resource utilization is improved and VM allocation time is reduced. The input for the VM Scheduler algorithm is BST (Binary search tree) of VMs and the output is allocated servers. The goals of this algorithm are to reduce the time of allocation of VM to server and to optimize the resource utilization. Servers are also listed in a BST. If the value is 1 then that server perfectly fit the VM specification. In various algorithms, the VM manager sends VM specifications to VM scheduler where as in this algorithm a BST is created for VM specifications and sent to VM scheduler.

The server which best fit the VM specification is found by dividing required virtual machine specification (RVMS) by available host specification (AHS). If the value is greater than 1 then that server will not fit the VM and so that server is rejected and search for the best fit server continues. If the value lies between 0 and 1 then that server is taken as a possible candidate and search for the best fit server continues.
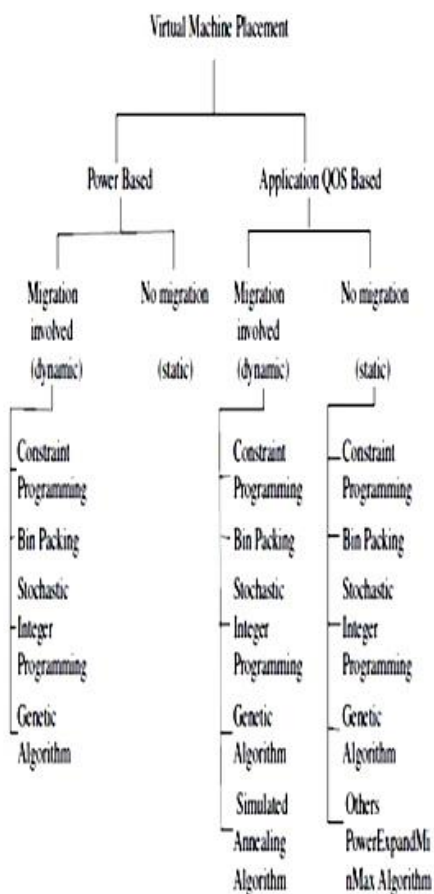

Figure 1.1: Classification of VM Placement Algorithms

III. CONCLUSIONS

Some VM placement algorithms with different objectives were surveyed. VM placement should be done by considering two or more goals at the same time which will improves the overall performance of servers in a datacenter. The VM placement is one of the research problems in cloud infrastructure.

REFERENCES

[1] M. Bichler, T. Setzer, B. Speitkamp, Capacity planning for virtualized servers, in: Workshop on Information Technologies and Systems (WITS), 2006.

[2] Sameer Kumar Mandal, Pabitra Mohan Khilar, Efficient Virtual Machine Placement for On-Demand Access to Infrastructure Resources in Cloud Computing, in: International Journal of Computer Applications, 2013, pp. 6–11.

[3] Fabien Hermenier, Xavier Lorca, Jean-Marc Menaud, Gilles Muller, and Julia L. Lawall. Entropy: a consolidation manager for clusters. In VEE, pages 41{50, 2009.

[4] Hien Nguyen Van, Frederic Dang Tran, and Jean-Marc Menaud. Autonomic virtual resource management for service hosting platforms. Software Engineering Challenges of Cloud Computing, ICSE Workshop on, 0:1{8, 2009.

[5] Fei MA, Feng LIU, Zhen LIU, Multi-objective Optimization for Initial Virtual Machine Placement in Cloud Data Center, in: Journal of Information & Computational Science, 2012, pp. 5029–5038.

[6] G. E. Box, J. S. Hunter, and W. G. Hunter, Statistics for Experimenters, 2nd ed., Wiley, 2005, 639 pages.

[7] Bu-Sung Lee Sivadon Chaisiri and Dusit Niyato. Optimal virtual machine placement across multiple cloud providers. In 2009, IEEE Asia-Paci_c Services Computing Conference, pages 103{110. APSCC, 2009.