CHALLENGES TO ENERGY EFFICIENT VIRTUAL MACHINE ALLOCATION IN CLOUDS: A REVIEW

Monica Gahlawat[1], Dr Priyanka Sharma[2]

[1]*MCA Department, L.J Institute of computer Application, (India)*

[2]*IT Department, Raksha Shakti University, (India)*

ABSTRACT

Virtual Machine (VM) Consolidation contributed a lot in energy efficient computing in the cloud data centers. Virtualization played a great role in virtual machine consolidation by facilitating the live VMs to be migrated resulting in packing the maximum number of VMs in minimum possible physical hosts .Energy efficient computing is not only good for reducing the overall operational cost of the data centers but also provides environmental benefits. Various algorithms have been proposed for VM consolidation which utilimately leads to energy efficient computing. But only high VM consolidation ratio is not a key to energy efficient data centers. A tradeoff between the consolidation ratio and the QoS (Quality of Service) is required preventing the SLA (Service Level Agreement) violation. In this paper we have presented various challenges to energy efficient VM allocation in clouds.

Keywords: VM consolidation, Virtualization, Energy Efficient Computing

I INTRODUCTION

The traditional era of computing involves the use of software, hardware and storage to achieve the required computational service whereas cloud computing has isolated the services from resources (networks, storage, servers). The cloud is evolved as a service oriented paradigm by providing the services to users by utilizing the resources of provider[1]. Users are no longer required to purchase hardware, software or to manage storages. As data centers increase in size and computational capacity, numerous infrastructure issues become critical. There is a debate that cloud computing is green contributing to sustainability or a risk of climate change. Cloud computing if compared to traditional on-premise computing, is a cost effective, energy-efficient, scalable and on-demand computing. But as the concept of cloud computing evolved various cloud providers came in to existence like AMAZON, Google, SalesForce.com etc. working on different service models e.g. IaaS(Infrastructure as a Service), PaaS(Platform as a Service) and SaaS(Software as a Service). To provide on-demand, elastic service to the customers, the providers established the datacenters on different geographic locations. The cloud providers need to be up to date with the infrastructure all the time to handle the random, unpredictable client requests. To fulfill this requirement the servers should be active all the time consuming a lot of power resulting in lot of carbon (CO2) emission. Carbon emission leads to climate change and a health related risk to the society.

Cloud data centers use virtualization technology for provision computational resources in the form of virtual machines (VMs). Saving operating costs in terms of energy consumption (Watt-Hour) for a cloud system is highly motivated for any cloud resource owner.[2] but only packing more number of virtual machines in less number of servers is not enough for energy efficient data center. High VM consolidation ratio can leads to shortage of RAM, degradation in performance or in the worst case the failure of the physical host. So VM consolidation ratio should be bounded by the QoS constraints for better performance.

The biggest challenge to the energy efficient algorithms is the tradeoff between the minimizing energy consumption while satisfying Quality of Service (e.g. performance or resource availability on time for any reservation request) which leads to sustainable cloud computing in the near future. To explore both performance and energy efficiency, three crucial issues must be addressed. First, excessive power cycling of a server could reduce its reliability. Second, turning resources off in a dynamic environment is risky from a QoS prospective. Due to the variability of the workload and aggressive consolidation, some VMs may not obtain required resources under peak load, so failing to meet the desired QoS. Third, ensuring SLA brings challenges to accurate application performance management in virtualized environments. [3]. Virtual machine selection approach for VM migration is also greatly effects the VM consolidation quality. All virtual machines in the datacenter are not independent but normally for large applications execution, the applications are distributed between more than one virtual machines. In this case, the non-optimized virtual machine migration results in network communication which may involve network switches which leads to more energy consumption. So, the care should be taken for the VMs executing one application should be collocated such that the communication cost and energy consumption can be reduced.

II. VM CONSOLIDATION PROCESS AND CHALLENGES

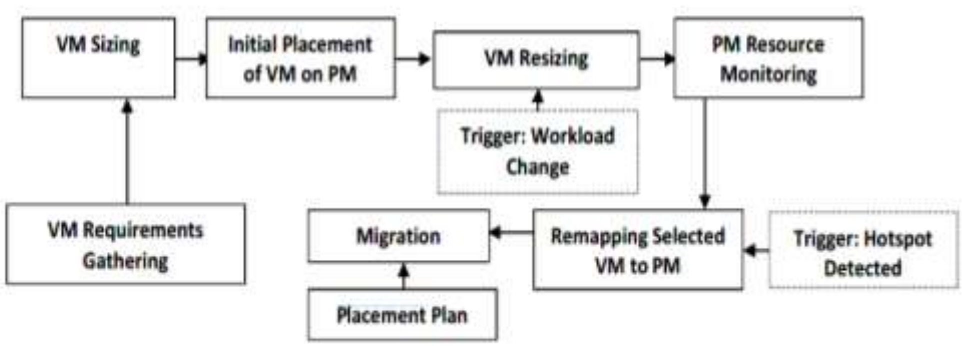

Fig 1: VM Allocation Process [4]

Figure 1 explains the VM Allocation process. VM allocation is performed in 2 phases, in the first phase according to the application size the VM is selected and the energy efficient initial placement algorithm executes and the physical machine is selected based on the algorithm. The second phase is the optimization of the VM allocation process called VM consolidation. The VM consolidation is based on the workload on the physical host. Most of the algorithms rely on the CPU utilization for the workload calculation. Based on the workload the mapping of virtual machine to physical machine is changed. The mapping is changed based on the optimized placement plan. The placement plan should be such that the energy consumption of the data center can be reduced. Virtual machine allocation and virtual machine migration are the backbone of the VM consolidation. The VM consolidation algorithm is divided in 4 sub-algorithms:

1) Host Overload Detection
2) Host underload Detection
3) VM Selection
4) VM Migration

There are various challenges to the energy efficient placement plan those needs to be addressed for all the above sub-algorithms. The challenges to the energy efficient VM allocation are as follows:

2.1 Energy Efficient Consolidation Vs Quality of Service

The biggest challenge to the energy efficient virtual machine allocation is allocating the virtual machines in such a way that the energy consumption in the datacenters is administered by the quality of the service normally defined in the form of SLA. It is therefore essential to carry out a study of Cloud services and their workloads in order to identify common behaviors, patterns, and explore load forecasting approaches that can potentially lead to more efficient resource provisioning and consequent energy efficiency. Host overload detection and underload detection can be predicted by their historical behavior. The statistical methods play a great role in forecasting the host overload and underload behavior. The IaaS cloud shows non-stationary workload behavior so the researchers can use the statistical methods based on the non-stationary behavior for the predictions. [3].

2.2 Virtual Communication Topologies

Large scientific applications often are executed in more than one virtual machine at a time and the virtual machines executing the same application often communicates with each other establishing a group and the communication topology. Due to the VM migration and the non-optimized virtual machine allocation the VMs in the same group may be allocated on different data centers situated in geographically dispersed area. The communication between the VMs requires the network devices like routers, network switches consuming a lot of power. Thus, VM migration should also consider the communication topology while deciding a particular machine to be migrated.[3]

2.3 Excessive Power Cycling Vs Reliability

Power cycling is the act of turning a piece of equipment, usually a computer, off and then on again. Reasons for power cycling include having an electronic device reinitialize its set of configuration parameters or recover from an unresponsive state of its mission critical functionality, such as in a crash or hang situation. Power cycling can also be used to reset network activity inside a modem. Frequently power cycling a computer can cause thermal stress and can reduce the overall reliability of the server. Second, turning resources off in a dynamic environment is risky from a QoS prospective. Due to the variability of the workload and aggressive consolidation, some VMs may not obtain required resources under peak load, so failing to meet the desired QoS.

2.4 Workload Variablity Vs Resource Availability

The workload in the IaaS is non-stationary means that the workload fluctuates depending on the user demand. Due to the variability of the workload it is very difficult to predict the utilization behavior of the host. The resource allocation in such a dynamic environment is a challenge. Though statistical methods can be applied to predict the behavior but we can't rely on statistical methods fully because of uncertain and dynamic behavior of the user demand. The aggressive consolidation in such an environment results in unavailability of the resources under peak load,, so failing to meet the desired QoS. While defining the VM consolidation plan the resource availability should also be keep in mind so that SLA is not violated.

2.5 High VM Consolidation Ratio Vs Performance

To avoid the overutilization and underutilization of cloud resources, virtual machine consolidation needs to be performed to minimize the energy consumption. If a physical machine fails due to some hardware failure or software failure before the completion of tasks then more number of virtual machines than the consolidated virtual machines need to be created again and execution of the tasks have to start again. This will increase great overhead in terms of energy consumption and resource utilization. [6] So only high VM consolidation ratio (the consolidation ratio is the number of virtual servers that can run on each physical host machine) only is not a key to performance. There should be a trade-off between VM consolidation ratio and performance.

2.6VM Selection For Migration

The major benefit of VM migration ais to avoid hotspots; however, this is not straightforward. Currently, detecting workload hotspots and initiating a migration lacks the agility to respond to sudden workload changes. Moreover, the in memory state should be transferred consistently and efficiently, with integrated consideration of resources for applications and physical servers.[8]

VM Selected for migration also affects the energy consumption of the data center. If the VM is a part of an application group and selected for the migration then again the communication topology will affect the energy consumption. The problems of convergence and local optimization have been challenging the research direction. On the other hand, we know that a data center doesn't have abilities in predicting the size and type of the next workloads. As a consequence, the optimal policy which the proposed algorithms have found out over a short period of time isn't necessarily the optimal solution over a long period of time. In a word, the global best which of some VM the proposed algorithms have found out in an algorithm cycle may be a local best in a long-term process. Moreover, since the capability with which the current random migration policy and optimal migration policy adapt to a dynamic cloud environment isn't excellent enough, they may cause many failure events of live VM migration in a real and dynamic cloud environment.

2.7 Host Selection for Migration

The selection of the physical host for allocation of the VM is also challenging. Various factors like network traffic between two virtual machines, energy efficiency, topology should be considered for selection of the physical machine However, due to the limitation of computing resource, it is impossible to locate all connected VMs into a partition. In order to reduce the whole data center traffic cost, VMs with heavy weight communication should have higher priority to be located close to each other, while ones with low weight communication have lower priority. However, to host two heavy weight communication VMs on a PM may increase the total power consumption and traffic cost instead of reduce them. If two VMs running network aware applications, which have low CPU usage, are hosted on a PM, the traffic cost for that connection is decreased. However, due to the low CPU usage, the PM utilization is low. Consequently, more PMs are turned on, thus increase the whole system traffic cost and energy consumption.

2.8 CPU Load Vs CPU Utilization

Virtual machine consolidation not only increases the utilization of the servers but also reduces the hardware requirement which ultimately reduces the energy consumption contributing to the reduction in overall operating cost of the data center. Most of the Virtual machine consolidation algorithms are based on predicting the CPU utilization of the servers and perform the Virtual machine migration approach to pack the maximum number of VMs on the minimum number of Servers. The CPU utilization of the servers (hosts) only is not enough for VM consolidation approach. The actual load of the server is determined by combining the CPU utilization with CPU load. The CPU load is defined by the number of application waiting to be executed on the queue which is a challenging task in such a dynamic environment like cloud.

REFERENCES

[1]. Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).

[2]. Quang-Hung, Nguyen, Nam Thoai, and Nguyen Thanh Son. "EPOBF: Energy Efficient Allocation of Virtual Machines in High Performance Computing Cloud."Transactions on Large-Scale Data-and Knowledge-Centered Systems XVI. Springer Berlin Heidelberg, 2014. 71-86.

[3]. Buyya, Rajkumar, Anton Beloglazov, and Jemal Abawajy. "Energy-efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges." arXiv preprint arXiv: 1006.0308(2010).

[4]. Amany Abdelsamea, Elsayed E. Hemayed, Hesham Eldeeb, and Hanan Elazhary, "Virtual Machine Consolidation Challenges: A Review," International Journal of Innovation and Applied Studies, vol. 8, no. 4, pp. 1504–1516, October 2014.

[5]. Belabed, Dallal, et al. "Impact of virtual bridging on virtual machine placement in data center networking." Teletraffic Congress (ITC), 2014 26th International. IEEE, 2014.

[6]. Sharma, Yogesh, Bahman Javadi, and Weisheng Si. "On the Reliability and Energy Efficiency in Cloud Computing." (2015).

[7]. Vu, Hieu Trong, and Soonwook Hwang. "A Traffic and Power-aware Algorithm for Virtual Machine Placement in Cloud Data Center." International Journal of Grid & Distributed Computing 7.1 (2014): 350-355.

[8]. Zhang, Qi, Lu Cheng, and Raouf Boutaba. "Cloud computing: state-of-the-art and research challenges." Journal of internet services and applications 1.1 (2010): 7-18.