A Survey on Different VM Scheduling Approach in Cloud Computing Environment

Prafulla Chakankar
School of Information
Technology,
RGPV, BHOPAL

Nishchol Mishra
School of Information
Technology,
RGPV, BHOPAL

Sanjeev Sharma
School of Information
Technology,
RGPV, BHOPAL

ABSTRACT

Cloud is the one of the fastest emerging technology in the IT world. Its supply on-demand IT resources to the client on the rent basis. In a very short time demand for the computing resources is increasing dramatically. To accomplish this demand virtualization approach is used which allow the sharing of physical resources. PM is virtualized by using the hypervisor that creates the VM according to the user need and assign to the users. Each user has its own VM and number of VM is running in single physical machine. As the size of the datacenter is increases, it can serve the more and more users but it will also introduce some issues that have to resolve for the proper utilization of the cloud services. Resource scheduling or the proper distribution of the physical resources is one of the critical issues in the cloud. Proper resource distribution can not only maximize the throughput but also increase the resource utilization of the physical machine. Resource scheduling is a challenging task in the cloud because user requirement for the resource can change dynamically. Various techniques have been proposed during the last decade in the field of resource scheduling. This paper discussed some of the existing resource scheduling algorithm with their anomalies..

Keywords

Cloud computing, virtualization, migration, resource scheduling, load balancing.

1. INTRODUCTION

Cloud is the one of the fastest emerging technology in the field of computer science. It become so famous due to its attractive features like east to use, on-demand, meter services, universal access etc [1]. According to the NIST [2] cloud is pool of resources that provides convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. In a short it provide on-demand services to the user and user can access theses services anywhere in the world [3]. One of main advantage of the cloud services is that user need to pay only for that amount that he is actually use [4]. It support three type of services designated software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS), and four type of deployment model named private cloud, public cloud, protected cloud and hybrid cloud [5, 6] as shows in figure 1.
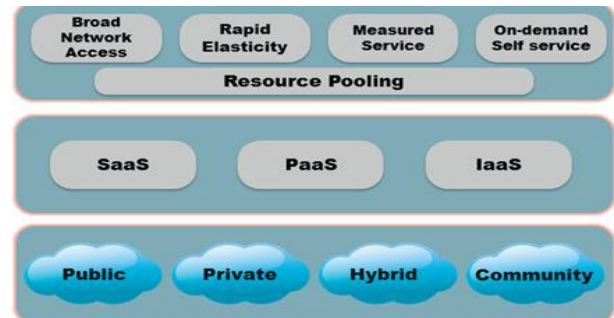

Figure 1: Cloud Computing Mode

Resource utilization is one of the key aspects of the cloud provider. For this purpose they use the virtualization technique which allows the sharing of the physical resources. Virtualization [7] allows the provider to divides the single physical resources into multiple according to the user need. Hence virtualization is the key technology in the cloud. It logically divides the physical resources into multiple and assigned the user in the form of virtual machine (VM) as shows in figure 2.

Figure 2: Virtualization

One of the key feature of the virtualization is the VM migration [8, 9], which allow to transfer the VM from one PM to another PM. This remapping enables the cloud provider the balance the load on the cloud. If the PM is unbalance, then VM can be transfer to the another PM to balance the PM. Resource utilization is increase due to the virtualization but it introduced some critical issues that have to resolve for the proper utilization of the cloud services. Resource scheduling or the proper distribution of the physical resources is one of the critical issues in the cloud. If the resources are not distributed properly then it will not only minimize the

throughput but also increase the number of migrations and number of running server or PM that increase the energy consumption. Since resources are given to the client in the form of VM, so resource scheduling means VM placement. Performance of the cloud services is mainly depends on the proper VM scheduling. Wrong scheduling of the VM can increase the number of the active servers. Hence, an effective VM placement is the prime requirement of the cloud provider.

VM placement can be static or dynamic. In the static VM placement approach fixed thresholds are use and the value of the threshold can't be change at the run time whereas in dynamic approach value of the threshold can be change at run time. Number of VM placement approach has proposed in the past few decades. This paper discussed some existing VM placement approach with their anomalies

2. LITERATURE SURVEY

M. Mishra et al. [10], proposed a vector based VM scheduling approach for the cloud environment. All parameter which required during the VM placement are represented by the vector. Resource required of the VM is represented by the RRV, resource utilization of the PM is represented by the RUV, total capacity of the PM is express by the TCV etc. This approach is mainly focus to the resource balancing and place the VM to the PM where resource utilization vector of the PM is complementary to the resource requirement vector of the VM. Load on the PM is depends on the number of VM inside the PM which is calculated by the following equation:

$$HL_k = \sum_{i=1}^{n}(w_i) * \frac{\sum_{j=1}^{m}(RU_{ij})}{(Hcap_{ik})}$$

Where, HL_k is the load on k^{th} host, n is the type of resources i.e. cpu, memory, io etc, m is the number of VM in k^{th} host, RU_{ij} is the ith type resources use by the jth VM and $Hcap_{ik}$ is the ith resource capacity of kth host. This approach may balance the resource but only theory about the method is given.

Y. Fang et al. [11], proposed layer based VM scheduling approach for the cloud. According to this approach VM scheduling approach consist of three layers named application layer, platform layer and infrastructure layer. Figure 3 shows the layer architecture use by this approach.

Application layer is design for the user. Users can only interact with the application layer and submit task to the application layer. Result is also received by the user through this layer. Platform layer is a set of hardware and software which is use by the developer to design new application whereas platform layer is set of virtual resources which is assign to the user. First level scheduling between user to VM, find the requirement of the VM and the second level scheduling between VM to host assign VM to the VM according to the VM description in the first level. This approach selects the VM on the first come first serve basis and assign to the PM which is lightest. Due to assigning the VM to the lightest PM this approach may increase the number of active servers.
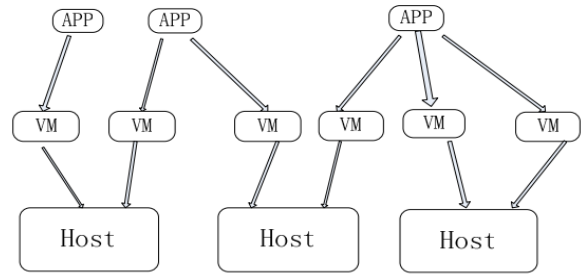

Figure 3: Two Level VM Scheduling Model

Armstrong et al. [12], proposed VM scheduling approach which uses execution time of the job to place the VM to the PM. This approach assigned VM to the PM which has minimum execution time regardless of the current load on that physical machine. They randomly select the VM from the pool and assign to the PM which it is expected to be executed fastest. This approach may increase the number of migration because this approach place the VM to the PM regardless of their load.

K. Dasgupta et al. [13], proposed Genetic algorithm for the load balancing in cloud environment. Main objective of the genetic algorithm is to minimize the make span of given task. typically, genetic algorithm is consist of four steps these are: Selection, Crossover, Mutation, and Termination. The genetic algorithm uses the natural PM selection strategy for the VM placement i.e., scan the entire PM list and assign VM to the first PM that is enough to hold VM. that is enough to hold VM. Cost function is use to assign priority to each VM and then place VM randomly to any PM. Problem with this approach is that it may have starvation. Higher priority VM is assign to the PM first, so if higher priority process comes regularly then lower priority VM is in starvation.

Nitin kumar Mishra and Nishchol Mishra [15], proposed a VM scheduling approach for the cloud which distribute the load equally among all the host. This approach proposed cost effective solution for the VM placement. Since time and cost are the two important parameter for the cloud provider and users, so algorithm which has lowest time and more throughput is the best. It is compare with round robin (RR) and throttle algorithm.

M. S. Pilavare and A. Desai [14], proposed VM placement approach based on the genetic algorithm [13]. Genetic approach is suffered by the starvation, so to avoid the starvation problem they prioritize the VM and then use the genetic approach to place the VM. To prioritize the VM they are using Logarithmic Least Squares Method which first find the comparison matrix of all VM then multiply each row to find the value of fitness function. Then assign the priority of each VM according to the fitness value. VM with higher fitness value will have higher priority. Following equation is use to find the fitness value for each VM.

$\zeta = w1 * \alpha(NIC \div MIPS) + w2 * L$

Where,

W1 and w2 are the weighting coefficient

NIC is the number of instruction in cloudlet

MIPS Million instruction per second requested by the VM

α is the cost of instruction execution

L is the delay cost

After assigning the priority to each VM, they use genetic algorithm for placing the VM, which randomly select the PM for placing the VM. Main problem with approach is that, they are not focusing on the energy consumption. VM placement play an important role in managing the energy consumption, hence energy consumption will be increase.

3. CONCLUSION

Resource scheduling in cloud different from resource scheduling in cluster and grid technology. Cloud uses the virtualization to increase the resource utilization. Virtualization increases the resource utilization of the PM but it also imposed the necessity of the resource scheduling. Since single resource is use by the multiple users, proper VM placement is important. Proper resource scheduling is a most challenging task in the cloud due to the dynamic nature of the user's requirement. Proper resource distributions not only optimize the system performance and also increase the provider profit. Several VM scheduling approach has been proposed in the past with their different objective. Main goal of the resource scheduling approach are to minimize the energy consumption, increase the profit and optimize the system performance.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] M. D. Dikaiakos, G. Pallis, D. Katsa, P. Mehra, and A. Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research", IEEE Journal of Internet Computing, Vol. 13, No. 5, September/October 2009, pages 10-13. Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[2] P. Mell and T. Grance, The NIST definition of cloud computing, http://csrc.nist.gov/ publications/nistpubs/800-145/SP800-145.pdf, 2012.

[3] R. W. Luckyet al., "Cloud computing", IEEE Journal of Spectrum, Vol. 46, No. 5,May 2009, pp. 27-45.

[4] R. K. Gupta et al., "A Complete Theoretical Review on Virtual Machine Migration in Cloud Environment", International Journal of Cloud Computing and Services Science (IJ-CLOSER), Vol.3, No.3, June 2014, pp. 172-178.

[5] R. K. Gupta et al.,," Survey on Virtual Machine Placement Techniques in Cloud Computing Environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 4, August 2014, pp. 1 -7.

[6] G. Pallis, "Cloud Computing: The New Frontier of Internet Computing", IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.

[7] D.Perez Botero, "A brief tutorial on live migration of virtual machine from a security perspective", 2011.

[8] C. Clark, K. Fraser, s. Hand and J.C. Warfield, "Live migration of virtual machine", proceeding of the 2nd conference on symposium on network system design and implementation, 2007.

[9] T. Wood et al., "Black-Box and Gray-Box strategies for virtual machine migration", proc. 4th conference on symposium on network system design and implementation, 2007.

[10] Mayank Mishra and Anirudha Sahoo, "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE 4th international conference on cloud computing in 2011.

[11] Y. Fang, F. Wang, and J. Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture Notes in Computer Science, Vol.6318, 2010, pages 271-277.

[12] T. R. Armstrong and D. Hensgen, "The relative performance of various mapping algorithms is independent of sizable variances in runtime predictions", in Proc. of 7th IEEE Heterogeneous Computing Workshop (HCW 98), pp. 79-87,1998.

[13] K. Dasgupta et al.,"A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing" in Proc. of Elsevier, Procedia Technology 2013.

[14] M. S. Pilavare and A. Desai," A Novel Approach Towards Improving Performance of Load Balancing Using Genetic Algorithm in Cloud Computing ", proceeding of EEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS'15, pp.1-4, 2015.

[15] Nitin kumar Mishra and Nishchol Mishra," CELBT: An Algorithm for Efficient Cost based Load Balancing in Cloud Environment", International Journal of Computer Applications (IJCA), Volume 134 – No.1, pp. 8-11, January 2016.