CPU Usage Prediction Technique for Live Migration of Virtual Machines

V.Ajantha
M.E - CSE [Final Year], Arasu Engineering College, Tamilnadu, India.

K.Suganthi
Assistant Professor, Department of CSE, Arasu Engineering College, Tamilnadu, India.

Abstract – Virtualization is a very important technology of cloud computing which partition the physical host into several Virtual Machines (VMs). The number of active host and their power consumption is easily managed by migrate the virtual machines based on their resource requirement and current status of the particular host. In cloud environment, Service Level Agreement (SLA) is necessary for maintaining reliable Quality of Service (QoS) between the user and data centers. The main objective is reduction of power cost and SLA violation. Live migration of the virtual machines manages the over load and under loaded situation of host which gives the ability of dynamic resource allocation on different host. The proposed technique will provide the ability of dynamic virtual machine consolidation using adaptive utilization threshold value based on CPU usage prediction which can easily manage the better SLA and reduces the number of VM migrations in between the host. The validation of the proposed technique on multiple workload traces of the Planet lab servers.

Index Terms – Virtualization, Resource, CPU, Prediction, Live migration, Consolidation, Virtual Machine.

1. INTRODUCTION

Cloud computing is a new computing paradigm which performs the delivery of computing services to the users without installing at local sites with minimal cost. These services are delivered to the user at on-demand way over the internet. The user must pay for the used resources when they release the resources. Cloud computing is the extension of existing technologies such as Service Oriented Architecture (SOA), Utility computing, Grid computing, Peer to Peer (P2P) computing, Virtualization. Many cloud service provider offers diverse services to the user, including Software as a service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Cloud computing enables user to store their information on remote side, so there is no need for storage infrastructure.

IaaS include all the basic infrastructure things, which provides the infrastructure for installing and maintaining the own environment in particular area. This layer is most useful in the data center perspective which can use the virtualization [2] technique for enhancing the service accessibility.

Virtualization is used to create a virtual version of a device or resource, such as a server, storage device, network and an operating system. By using virtualization technique, run multiple OS and applications on single server. Virtual Machine Monitor (VMM) or hypervisor is used in host for creating virtualization devices.

By virtualization technique, physical machines are divided into number of virtual machines for achieving the efficiency of the utilization by sharing the physical server resources into multiple virtual machines. The best feature provided by the hypervisors and virtualization technique is hot and cold live migration of virtual machines. The live migration is mainly used for load balancing, high performance throughput, fault tolerance and the physical server maintenance [3].

A virtual machine is used, that provides an operating environment for run a host OS. It is created and managed by VMM. The operations included in VM are multiplexing, suspension, provision, migration. It Offer solutions for underutilized resources and application inflexibility

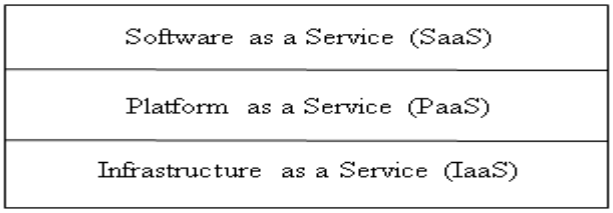

Fig.1. Cloud Services

Live migration is beneficial in the dynamic virtual machine consolidation according to their resource requirement and minimizes the number of active host for energy saving purpose. Migration of virtual machines in live environment can saves the power consumption as it keeping the low number of active host in the data center. Quality of service; managed by the service level agreement (SLA) which describes the response time and throughput of the system. The live migration should maintain the power management, SLA level and guarantee of the performance after the migration of the virtual machine.

This approach describes the phenomenon of the under load and overload detection whenever the host is under loaded and over loaded, then all the virtual machines from that host migrated on another host and that particular host is switched off for a specific time. Decreasing the number of active host automatically reduces the power consumption in data center. Moreover, the requested amount of CPU is exceeds the available capacity on the particular host, then that host considered as over loaded host. So, VMs migrate to another host for maintaining the SLA level.

Live migration and the SLA management addressing the following challenges in dynamic VM consolidation:

1. Determination of the overloaded host.
2. Determination of the under-loaded host.
3. Selection of the virtual machine for migration.
4. Placement of the selected virtual machine.

The first two issues considerably influence the decisions of other problems. So the expected CPU requirement of the host is used for deciding the overloaded and under-loaded host is important for migration in cloud environment. Predicting the CPU utilization on the basis of dynamic threshold will help in dynamic VM consolidation and generates the efficient load balanced structure on the every host.

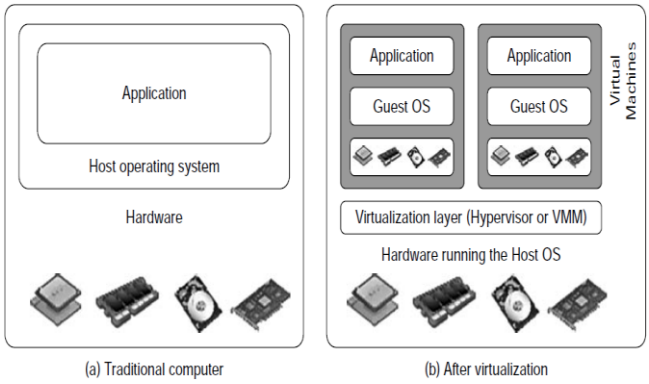

Fig.2. Virtualization

A system design which handles the auto live migration of virtual machines based on CPU utilization which manages the high level of SLA, reduces the number of migrations and decreasing the power consumption in data center. This addresses the following objectives: CPU utilization data of the servers as a input and Percentage of SLA violation, Number of VM Migrations and Power consumption as a output.

2. RELATED WORK

Nathuji and Schwan [8] have carried out the data center power management issues. They have dispersed the policies and methods in global and local level. At the local level, it manages the guest OS power management and gives the information to the global manager. Then the global manager catches the information from current resource allocation and applies its procedure to choose the virtual machine for correct placement.

Kusic et al. [6] have addressed the power management problem in the virtualized environment and the sequential optimization achieved it on the limited look ahead control (LLC). The main objective and purpose is to take the full advantage of the resource, revenue and reducing the power usage and SLA. Kalman filter has proposed the work on future demand forecasting and evaluate the necessary reallocation but require the modifications in the application which is not possible in the IaaS.

Verma et al. [10] address the continuous optimization problem of heterogeneous system as a power aware placement of the applications. This optimization was mainly useful in reducing the power usage and improvement in the performance. The performance of the application can be degraded due to the workload inconsistency. Recently, they have proposed this approach in to static, semi static, and dynamic consolidation. The multi-tier web application consolidation problem using live migration have expressed by the Jung et al. [4]

In static CPU threshold the decision is made on the threshold set in the policy if the CPU utilization value goes on the 85% of the total capacity then host is declared as a over loaded host but it do not adapt the workload changes so these are not suitable in the dynamic workload. Therefore, decision making on the dynamic workload based on the analysis of historical data is important.

Kumar et al [5] suggested an approach of dynamic VM consolidation established on the stability approximation. This is used as a possibility that the reallocation of the virtual machine will stay active for the provided time frame in the future.

The proposed adaptive migration threshold for dynamic VM consolidation at run time utilization of the resources and migration of the virtual machines switches off the idle nodes and reducing the power consumption. Live migration process varies according to the type of hypervisor and its supporting environment. We have conduct the review on architecture of Leading Hypervisors and Their Live Migration Techniques [2]; which explains the process of virtual machine migration and the architecture of hypervisor.

3. SYSTEM ARCHITECTURE

This scheme targets the data center especially the IaaS environment, which contains the N heterogeneous host and every node is characterized by the RAM, bandwidth, disk storage and the CPU performance (MIPS). The major illness is this environment not having the awareness of the workload and the time for virtual machines are provisioned in the data center. Multiple users are submitting the request for accessing of the

VMs. The virtual machines are owned by the independent users, so there is no chance for get the information of the type of applications installed by the other user.

This environment having the common storage connected to all the physical nodes. But without shared storage, live migration of virtual machines cannot be possible. The main advantage of using the shared storage while forming the cloud cluster, all the virtual machines has a virtual disk and all are placed on the shared storage and whenever the virtual machine needs to be migrated the disk, all Nodes are having connectivity to this storage and this disk will be remain on the shared storage and only memory pages will be transferred from one node to another. This migration will improve the performance and never gives the load on network and manages the application performance.

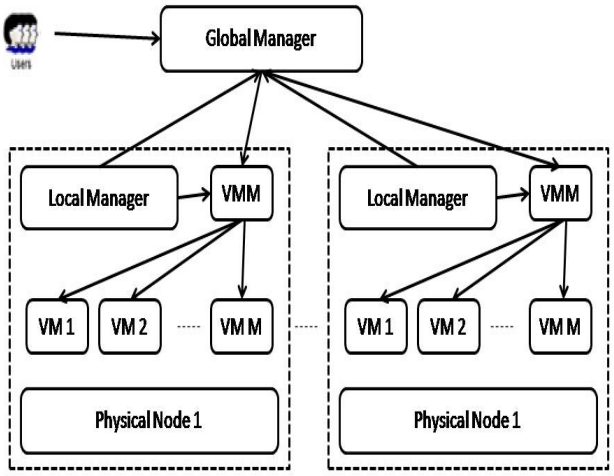

Fig.3. System Achitecture

For managing the virtualization, need to install the hypervisor on each host which is having the VMM (virtual machine monitor) and responsible for monitoring the local resources of the host server. The data is fetched by this monitoring system from this module. The system having multiple VMs created on the physical node and the information is taken by local manger from VMM. The local mangers are mainly responsible for monitoring the utilization of CPU and taking the correct decision for VM migration. Like as a VM scheduler the global manager works, which is connected to all VMM and local manager of each node for maintaining the over utilizations and commands are issued by local manager for the VM placement.

3.1. SLA Violation

The QOS is a main thing which should be considered and managed in the cloud computing environment by the cloud service provider. This requirement is addressed in the format of SLA which is characterized in the terms of maximum response time and minimal throughput. These characteristics can be varied from one application to another application. We have defined this generic terms in our experiment by fraction of the requested MIPS by all VMs Urj(t) and allocated MIPS in actual Uaj(t) relatively to the total requested MIPS during the life time of the VMs. Number of VMs is M.

$$SLA = \frac{\sum_{j=1}^{M} \int_t U_{rj}(t) - U_{aj}(t)\, dt}{\sum_{j=1}^{M} \int_t U_{rj}\,(t)} \qquad (1)$$

This equation represents the percentage of CPU has not been allocated when demanded by the applications.

3.2. Power Consumption

The CPU utilization always changes as workload is varied from time to time, the workload variability (t), to calculate the total energy consumption by the server with following model.

$$E = \int_t P\big(\,\mu(t)\big)\, dt \qquad (2)$$

3.3. Dynamic Threshold

The static threshold is not suitable for the dynamic and random workload which is unpredictable. We are looking for the system which is suitable for all types of workload and should adjust his behavior according to the workload. Therefore we propose the novel technique which adjusts the utilization threshold based on the analysis of past (historic) data of the virtual machine.

4. ALGORITHM

Following dynamic reallocation algorithm using the dynamic threshold utilize the minimum migration policy for VM selection. The major complexity of an algorithm is the addition of number of non over utilized host plus the product of over loaded host and their respective allocated VMs.

4.1. Reallocation Algorithm

Algorithm: Dynamic Thresholds (DT)

1. **Input:** hostList, vmList
2. **Output:** migrationList
3. vmList.sortDecreasingUtilization()
4. foreach h in hostList do
5. hUtil h.util()
6. bestFitUtil MAX
7. while hUtil > h.upThresh() do
8. foreach vm in vmList do
9. if vm.util() > hUtil - h.upThresh() then

10. t vm.util() hUtil + h.upThresh()
11. if t < bestFitUtil then
12. bestFitUtil t
13. bestFitVm vm
14. else
15. if bestFitUtil = MAX then
16. bestFitVm vm
17. break
18. hUtil hUtil - bestFitVm.util()
19. migrationList.add(bestFitVm)
20. vmList.remove(vm)
21. if hUtil < lowThresh() then
22. migrationList.add(h.getVmList())
23. vmList.remove(h.getVmList())
24. return migrationList

5. EXPERIMENTAL RESULT

In a cloud computing environment the proposed system should be evaluated in the large scale data center architecture environment. It is really difficult to conduct the experiment on real cloud set up as it requires huge amount of resources and cost. Whenever performing the live migration of virtual machines it needs the nodes with large amount of resources and shared storage must be attached to all nodes. If the proprietary hyper visors is chosen, then there is a need to purchase the license of their required software's and few open source hyper visors varies in the supporting guest OS. Whenever there is a need to reproduce the experiment with same condition to compare the multiple algorithms it is quite difficult to have the same state of the environment. The CloudSim framework is used for testing the algorithms in which the data center environment is developed. This modern framework is widely used in the cloud computing environment for testing the multiple algorithms due to the support for large scalability. In this framework we can scale the resources as well as we can apply the dynamic workloads. CloudSim can be easily integrated with the Eclipse which requires JRE and JDK 1.7 above version.

6. CONCLUSION

Virtualization is the concept of creating a virtual version of a device or resource. By using above algorithms the performance of the cloud servers is increased and the number of migrations of VMs from one physical machine to another physical machine is decreased. The proposed system supports the both static and dynamic threshold based migration. This system essentially focuses on the dynamic VM consolidation and the total number of physical host requirement is reduced. For maximization of Return on Investment (ROI), Cloud providers need to apply multiple strategies, like dynamic VM consolidation with switching idle servers to power-saving modes. A technique for implementing distributed dynamic VM consolidation in IaaS Clouds under workload-independent QoS constraints is proposed. This technique increases the utilization of data center resources and energy consumption is reduced by migration, as well as satisfying the defined QoS requirements.

REFERENCES

[1] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R.Neugebauer, I. Pratt, and A. Warfield, Xen and the Art of Virtualization, Proceedings of the nineteenth ACM symposium on Operating Systems Principles (SOSP03), pp.164-177, 2003.

[2] A. Beloglazov and R. Buyya, Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers, Concurrency and Computation: Practice and Experience (CCPE), Vol.24, pp.1397-1420, 2012.

[3] Fahimeh Farahnakian, Pasi Liljeberg, and Juha Plosila LiRCUP: Linear Regression based CPU Usage Prediction Algorithm for Live Migration of Virtual Machines in Data Centers 39th Euromicro Conference Series on Software Engineering and Advanced Applications 2013.

[4] Jung G, Joshi KR, Hiltunen MA, Schlichting RD, Pu C. A cost-sensitive adaptation engine for server consolidation of multitier applications. Proceedings of the 10th ACM/IFIP/USENIX International Conference on Middleware (Middleware 2009), Urbana Champaign, IL, USA, 2009; 120.

[5] Kumar S, Talwar V, Kumar V, Ranganathan P, Schwan K. vManage: loosely coupled platform and virtualization management in data centers. Proceedings of the 6th international conference on Autonomic computing (ICAC 2009),Barcelona,Spain, 2009;127136.

[6] D. Kusic et al. Power and performance management of virtualized computing environments via lookahead control. Cluster Computing, 12(1):115, 2009.

[7] V. Malik, C. R. Barde, "Survey on Architecture of Leading Hypervisors and Their Live Migration Techniques", International Journal of Computer Science and Mobile Computing,IJCSMC, Vol. 3, Issue. 11, November 2014.

[8] R. Nathuji and K. Schwan. Virtualpower: Coordinated power management in virtualized enterprise systems. ACM SIGOPS Operating Systems Review, 41(6):265278, 2007.

[9] BP. Rimal, E. Choi, I. Lumb, A Taxonomy and, Survey of Cloud Computing Systems, Proceedings of the Fifth International Joint Conference on INC, IMS and IDC, pp.4451, 2009.

[10] A. Verma et al. pMapper: power and migration cost aware application placement in virtualized systems. In Proc. of the 9th ACM/IFIP/USENIX Intl. Conf. on Middleware, pages 243264, 2008.

Authors

V.Ajantha received B.Tech degree in the department of Information Technology from Sudharsan Engineering College in 2014. She is now pursuing M.E in the department of Computer Science and Engineering in Arasu Engineering College. Her area of interest is Cloud Computing, Mobile Computing and Big Data.

K.Suganthi has received B.Tech in the department of Information Technology from Anjalai Ammal Mahalingam Engineering College in the year 2005 and M.Tech in the department of Computer Science and Engineering from PRIST University in the year of 2014. And now she is working as an assistant professor in Arasu Engineering College and also have 8 year and 2 months of teaching experience in the same Institution. Her area of Interest is Mobile Computing, Network security, Wireless Network and Image Processing.