Adaptive Heuristics with Self-Healing for Efficient Dynamic Consolidation of Virtual Machines in Cloud Datacenters

Kshitiza Vasudeva
Department of Computer Science & Engineering
Jaypee University of Information Technology, Solan, Himachal Pradesh, India

Satya Prakash Ghrera
Head, Computer Science Engineering& Information Technology Department
Jaypee University of Information Technology, Solan, Himachal Pradesh, India

Abstract-Energy consumption in cloud data centers is major issue due to cost expenses, performance degradation and environmental impact. To reduce the power consumption many ideas have been proposed among which live VM migration is the most popular one. Live migration further increases the overhead, delay and degrading the performance. The concept of self healing is proposed to reduce the power consumption as well as number of virtual machine (VM) migrations. The main aim of this work is to self-heal the overloaded host before going for VM migration. The proposed algorithm is developed and implemented using CloudSim toolkit. The results demonstrate that the proposed system can handle dynamic workloads and show better performance.

Keywords-Cloud Computing, Virtualization, Dynamic Consolidation, Self-Healing.

I. INTRODUCTION

Cloud computing is among the most fast growing and symbolic contemporary technologies that has brought up sort of revolution in modern ICT [1]. Cloud Computing " is a model for allowing ubiquitous, convenient, and on-demand network access to a number of configured computing resources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2]. It has proved to be a powerful architecture which can handle large scale systems and perform complex computing. Cloud computing has wiped off the need to store, maintain and evaluate huge datasets through virtualization .It has not only minimized the cost of infrastructure storage and maintenance but also made systems more secure, reliable and efficiently manageable [3]. It efficiently addresses issues of rapid growth of economies and limited number of resources. Cloud computing gives full opportunity to organizations to just work on main logics related to business rather than being concerned about the infrastructure, availability of resources, man power, cost, security, etc. [4]

The fundamental basis of cloud is that content of users is not stored on local systems but is kept and processed in the datacenters through internet. The main technology used by cloud computing is virtualization. It is used for abstraction of the computing resources. The cloud providers are responsible for the management and maintenance of these data centers. The cloud providers provide Application Programming Interface (API) to the users so that they can access the data stored through any computing device connected to the internet.

1.1 ENERGY CONSUMPTION ISSUE IN CLOUD DATA CENTERS

Cloud computing service is a collection of virtual data centers with high optimization which not only software but also hardware and other resources.Companies and other organizations in need of any resource can use resources through pay-per-use model by simply connecting to the cloud.This curtails down their capital expenditure on extra resources at premises. To cope up with growing demand of computational power for high performance applications, companies need large scale data centers which are the core part of system. However, these data centers devour colossal amount of electrical power which has exceeded the cost of actual infrastructure. To make maximum profit, saving operational cost is preferred over performance. People have begun to pay more attention to energy consumption rather than only considering performance [5].

According to the survey presented [6], data centers in US consumed around 61 billion KWh in 2006 which is almost equivalent to 1.5% of the total energy consumption in US. Moreover, in 2011 it jumped to more than 100 billion KWh [7]. In spite of these economical reasons, energy consumption has adverse effects on the environment too. Servers emit CO_2 which is the main cause of greenhouse effect. And according to the survey, this high energy consumption and CO_2 emissions will keep growing in upcoming years.

Many different applications are run at the same data center which contains many heterogeneous servers and network devices. To keep these applications isolated and exploit features of cloud like elasticity, flexibility and reliability, cloud uses virtualization technology. Virtual machines(VMs) are the basic blocks of resources which are provided to customers either directly or indirectly through the provisioned applications.[8]

To save energy in data center best way is to efficiently utilize the resources i.e the VMs. VMs are consolidated to minimum number of hosts and idle hosts are switched off or put in other mode of operation like sleep mode. However, sometimes VM consolidation becomes too combative which may overload hosts and violate SLOs fixed in Service Level Agreement (SLA). Hence, there is a trade off between the QOS and energy consumption which is optimized by allocating VMs efficiently [9, 10].

Elasticity as a feature of cloud is the ability of the system to scale up and down the resources according to the current demand. So, when the load is high or low, the system should consolidate VMs accordingly with respect to QoS and saving power. Methods used for implementing elasticity are re-dimensioning, migration and replication.

VM allocation problem is divided in four steps [11]:

- **a)** Overloaded host detection
- **b)** VM selection for live migration
- **c)** Detecting underutilized host
- **d)** Migrating all VMs and turning off the host.

This live VM migration causes delay and overburdens the network as well as the physical hosts involved. Hence, violates SLA and degrades the performance of the system.

In this work we have proposed the concept of self-healing to reduce the number of VM migrations and power consumption without violating the SLA.

The remaining paper is setup as follows. Next section 3 discusses the literature review by other research scholars. Section 4 has the proposed system model with all the details of architecture and algorithm. Section 5 shows the simulated results and their analysis by comparing with the existing system. Section 6 concludes the paper with respect to proposal and results.

II. LITERATURE REVIEW

Energy consumed by distributed systems has brought up many issues and has become an outstanding question and requires consideration. Among all the existing methods to save energy, energy consumption can be reduced by proper scheduling of the applications and consolidating them to reduce the running servers.However, many scheduling approaches yet did not acknowledge the cost of energy consumption on network devices, which is also contributesto power consumption in data centers. Hierarchical Scheduling Algorithm (HSA) was proposed to curtail the energy consumed by both servers and network devices. In HSA, a Dynamic Maximum Node Sorting (DMNS) method dealt with optimizing the placement of applications on servers. To further lessen the number of working servers, Hierarchical crossing-switch adjustment is used. Results showed that the number of working servers as well as data transfer speed reduced to good extent.The HSA is simple and robust to minimize the energy consumption by effectively scheduling the applications but HSA and DMNS are not suitable for dynamic workload [12].

An immediate fix to curtail the power consumption in data centers is to utilize the modes with lower power. To measure the variation in energy consumption due to virtual machine scheduler's simulation was conducted and besides also demonstratedthe inability of default schedulers, using optimized scheduler. The customized scheduler has reduced the complete machine uptime by up to 60.1% after using many real simulation scenarios. OptSched optimizes the virtual machine to physical host mapping by utilizing the reservation length. The parameters covered in this study were heterogeneity of data centers and VMs, the long effect of run time distributions and sensitivity to batch requests. The cumulative machine uptime is balanced for heterogeneity of virtual machines but energy consumption is not efficient if the work load is highly dynamic [13].

Beloglazov et al. gave two step proposal to efficiently allocate the VMs. In the first step, new requests for VM provisioning are allowed and VMs placed on hosts, and in next step current allocation setup of VMs is optimized. The first part is somewhat like a bin packing problem with variable bin sizes and prices. As a solution, modification of the Best Fit Decreasing (BFD) algorithm is applied. In MBFD, VMs are sorted in decreasing order of CPU utilization and then VM is allocated to a host which shows minimum increase in power consumption after allocation.This gives a chance to choose the most efficient one with respect to power. The complexity of the algorithm is n·m, where n is the number of VMs. Optimization part of VM consolidation is further carried out in two steps: in first part VMs are selected to migrate and then chosen VMs are allocated to the host based on MBFD algorithm.The energy consumption is less with respect to the reliable QoS. The proposed approach does not follow the strict SLAs between the service provider and user under the dynamic workload [14].

The core technology used by cloud is virtualization so as to adequately consolidate the VMs into physical host for better utilization of resources and to save power. A survey done by many show that the average utilization of servers is still less than expected.A new concept was proposed for this dynamic consolidation of VMs by a dynamic programming algorithm which chooses the VMs from an overutilized host taking into consideration the overhead caused by migrating a VM. Since, all VMs are attached to a storage area network (SAN), the cost of live migration of a VM is decided by its memory imprint. Therefore, time taken by a VM to migrate is calculated by dividing the memory size of VM by network bandwidth. As a result, cost of migration is measured by memory size of the VM. Thus, while selecting the one with less memory size is the best. The cost based approach of VM migration minimizes the power consumption cost of the service provider but when the workload is variable with respect the application, the approach is failed to meet the SLAs [15].

The large-scale data centers contain thousands of servers whichconsume large amount of electrical power leading to high operating costs. Therefore, to curtailthis cost of power the cloud providers need to optimize resource usage effectively by consolidating VMs efficiently in order to improve energy efficiency. The problem of VM consolidation is divided into four sub-problems: physical host overload detection; host under-load detection; VM selection and VM placement. Each of the sub partswork together to optimize the trade off between energy and QoS. For dynamic consolidation of VMs, a new multi-agent system (MAS) was proposed to make the cloud system smarter by blending the five traits of multi agent systems which are ubiquity, intelligence, delegation, interconnection, and human orientation. MASs provide the cloud systems intelligent and insightful based software which can help in effective and better system. The proposed method has significantly reduced energy consumption and also kept constant with the objectives of the Service Level Agreements (SLA).The number of VM migration and energy consumption is effectively minimized but when the workload is variable with respect the application, the approach is failed to meet the SLAs [16].

Extra load on server machine causes performance degradation of applications because resources available are not sufficient. Currently all the proposed methods towards the issue of host overload detection are mostly heuristic-based, or depend onhistorical data analysed statistically. The drawback of this way is that it gives sub-optimal results.Beloglazov et al have given a solution to host overload detection problem by maximizing the mean inter-migration time keeping in mind the specified SLA based on a Markov chain model. Multisize Sliding Window workload estimation technique is also proposed to heuristically adapt the algorithm to unknown non-stationary workloads. It is quite probable that when the resources are utilized at max, the applications are more prone to lack of resources and performance degradation. To address this problem, most of the schemes for dynamic VM consolidation apply either heuristic-based technique, such as static utilization thresholds. The mean inter-migration time is reduced of the VM migration but the number of migration of VM is high which violates the SLAs [17].

III. THE PROPOSED MODEL

In this paper, the main framework considered is an Infrastructure as a Service (IaaS) environment. It represents N number of different physical hosts. Each host is portrayed by the CPU performance defined in MIPS, RAM and transfer speed. Virtual machines (VMs) are hosted on the physical machines on-demand of the users.Customers submit their requirements specifying MIPS, bandwidth, number of processors, etc. Different users hosting various types of applications use resources simultaneously. For this transparency, cloud system uses virtualization technology.

The system model shows software layer of the framework which consists of global resource manager, local managers and VMs hosted on physical machines. Each node has one local manager as a part of VMM which keeps a continuous check on resource requirements of VMs and also makes decision regarding selection and migration of VMs at time of overload. The global manager is a part of master node which is in contact with all the local managers and collects information so as to optimize the resource utilization and Service level objectives. It gives the orders regarding VM placement. Actual migration and resizing of VMs is performed by VMM.

3.1Proposed work

An adaptive heuristics is used for dynamic consolidation of VMs based on an analysis of previous data from the resource usage by VMs. The proposed algorithm significantly reduces energy consumption, while ensuring a high level of adherence to the Service Level Agreement(SLAs). The proposed algorithm performs dynamic consolidation of VMs at run-time on the basis of current utilization of resources which may involve live VM migration, changing the mode of unused host to lower power mode so that power can be saved.The system efficiently handles firm SLAand multi-core CPU architectures. The algorithm adapts the behaviour with respect to observations and characteristics of VMs.

3.2 Self-healing

With respect to the proposed work mentioned above, the live migration of virtual machine from the overloaded host is not performed at the first instance. Instead of that, for each of the over utilized host self healing is performed. All virtual machines utilization is analyzed in each overloaded host, then add the MIPS to the more utilized VM and remove MIPS from less utilized VM or if the host has some free PE or MIPS that can be added to more utilized VM. So, overloaded host adjusts VM parameters using self healing and balance the utilization without violating the SLAs. If self healing is not possible, then proposed approach performs the normal VM live migration algorithm where the migrating VMs are selected from the overloaded host and these VMs are migrated to the other host using some policy. For all underutilized hosts, all the VMs are migrated to the safe host.

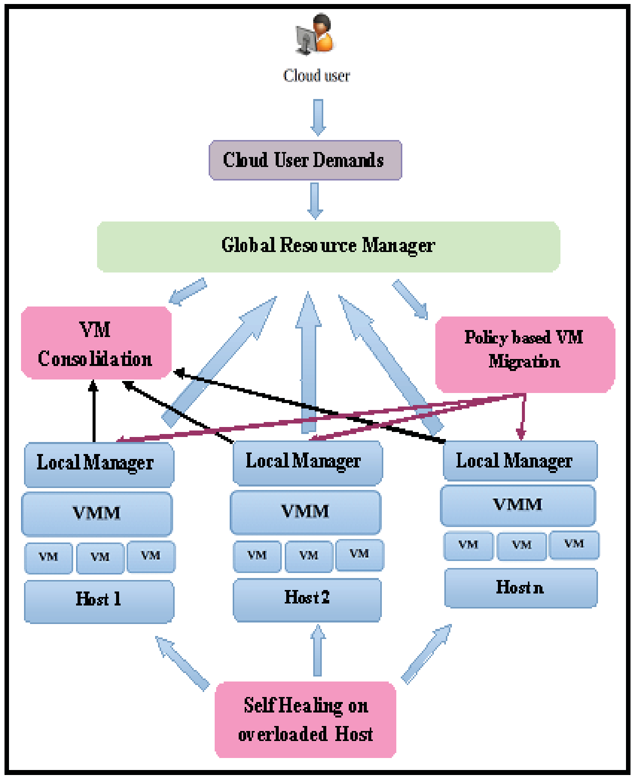

Fig 1. The System Model.

The proposed algorithm with self-healing so as to curtail down the number of migrations at first instance and energy consumption is shown below:

STEP 1: Create the user demands such as application configuration and VM configuration
STEP 2: Initialize the CloudSim and create the datacenter, broker, hosts, and VM based on the user demands
STEP 3: Create cloudlets (jobs or applications) for user requirements
STEP 4: Schedule the task on the VM based on VM allocation policy
STEP 5: Start simulation
STEP 6: Calculate the CPU utilization on the every host
STEP 7: Iteration
7a: get the first host in the list
7b: if host CPU utilization is lower than 0.2 then move the host to underutilized host list
7c: if host CPU utilization is greater than 0.8 then move the host to over utilized host list
7d: else mover the host to safe host list
Close the for loop
STEP 8: Iteration over utilized host list get the VM with maximum utilization
8a: get the available MIPS from the host of maximum utilized VM
8b: if MIPS is available then add available MIPS to over utilized VM
8c: else migrate the VM to safe host based on some policy
STEP 9: Iteration for each underutilized host
9a: Consolidate the every VM on overloaded host and move those
9b: VMs to migration list
Close the for loop
STEP 10: Sort the safe host in increasing order based on CPU utilization and migrate all the VMs based policy (migrate the VM with maximum utilization to host with minimum utilization to achieve the balancing)
STEP 11: Run applications
STEP 12: Stop simulation.

IV. SIMULATED RESULTS AND ANALYSIS

The CloudSim toolkit [20] has been selected as a platform for simulation, as it is a modern simulation framework setup for Cloud Computing environments. CloudSim 3.0.3 version is used. Data centers as physical nodes have been simulated half of which are HP ProLiant ML110 G4 servers, and the other half HP ProLiant ML110 G5 servers.

For evaluation of the proposed system with respect to existing system in cloudSim we have chosen two metrics.The two are energy consumption by the server machine in data center and Number of VM migrations due to application workloads. Table 1 shows the assumptions by which power consumption is calculated in cloudSim.

Table 1. Power consumption by the selected servers at different load levels in Watts

	CPU UTILIZATION IN PERCENTAGE (%)										
SERVER	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
HP ProLiantG4	86	89.4	92.6	96	99.5	102	106	108	112	114	117
HP ProLiant G5	93.7	97	101	105	110	116	121	125	129	133	135

For each particular policy given below, Power Consumption and number of VM migrations have been calculated for the existing as well as proposed algorithm by using combinations of host overloaded detection algorithms and VM Selection Algorithms.
Combinations used are:

- **IQR-MMT**(Inter quartile Range as host overloading detection algorithm and Minimum Migration Time Policy as VM selection policy)
- **LR-MMT**(Local Regression and Minimum migration Time policy)
- **MAD-MMT**(Mean Absolute Deviation and Minimum migration Time policy)
- **THR-MMT**(static Threshold and Minimum migration Time policy)
- **LR-RC**(Local Regression and Random Choice Policy)

According to proposed scheme simulated results have been shown and compared with the existing values of the metrics. Table 2 shows the results of Energy Consumption in kWh of the proposed algorithm and existing system.

Table 2. Energy consumption in kWh by the policies

Policy	**Existing system**	**Proposed system**
IQR-MMT	47.85	38.93
LR-MMT	35.37	36.61
MAD-MMT	45.61	32.65
THR-MMT	41.81	32.24
LR-RS	34.41	28.77

According to the proposed algorithm, above results can be visualized in graph to evaluate the performance of the algorithm in terms of energy consumption. LR-RS policy consumes minimum energy.

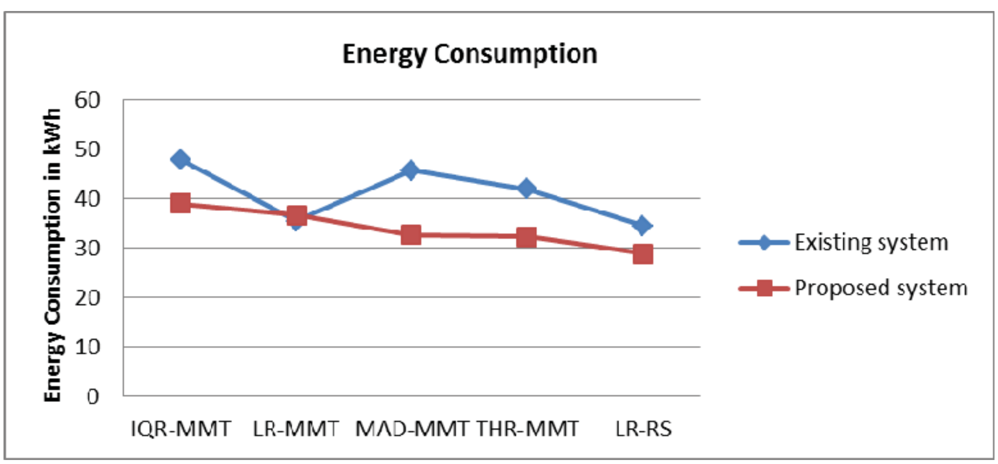

Fig 2. Comparison of policies based on Energy Consumption.

Above Figure 2 shows comparison of policies based on energy consumption using proposed and existing system.

Table 3. Number of VM migrations

Policy	Existing system	Proposed system
IQR-MMT	5502	3395
LR-MMT	2872	1867
MAD-MMT	5265	3103
THR-MMT	4839	3146
LR-RS	2434	1194

Table 3 shows the results of number of VM migrations of the proposed algorithm and existing system.

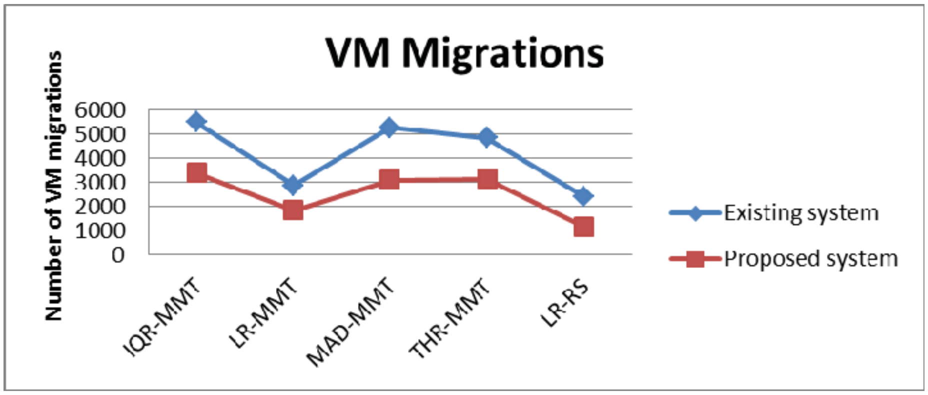

Fig 3. Comparison of policies based on VM Migrations.

Above Figure 3 shows comparison of policies based on number of VM migrations using proposed and existing system.

V. CONCLUSION

In order to cope up with growing dynamic load and to keep balance with operating cost, cloud providers need to optimize the resource utilization. High Energy consumption at cloud data centers has received much attention in past few years due to SLA violations, high operating cost, CO_2 which has bad impact on environment. This problem of energy consumption with live VM migration is discussed above. Many research scholars have studied this issue and tried to solve by proposing different strategies which are mentioned in literature survey. With respect to the work done before, a concept of self-healing is proposed to reduce the VM migrations and power consumption too. The proposed algorithm is mentioned as well as the simulated results and their analysis. Hence the number of VM migration are reduced, the waiting time of the running application on the VM is avoided. Hence proposed strategy strictly follows the SLAs assured between the cloud users and cloud service providers. If the self healing is not possible the, VM migration is performed on the basis policies existing in cloudSim toolkit to reduce the host overload. The VMs consolidation is also effectively done by proposed strategy and VMs placement is estimated based on the current resources requirements of the every VMs running on the host.

REFERENCES

[1] Hashem, Ibrahim AbakerTargio, IbrarYaqoob, Abdullah Gani, “The rise of big data on cloud computing: Review and open research issues,” Information Systems, Elsevier, Vol. 47, pp. 98–115, 2015.
[2] Lu, Chih-Wei, Chih-Ming Hsieh, Chih-Hung Chang, and Chao-Tung Yang. "An improvement to data service in cloud computing with content sensitive transaction analysis and adaptation." 37th IEEE Annual Computer Software and Applications Conference Workshops (COMPSACW), pp. 463-468. IEEE, 2013.
[3] P.Mell, T.Grance, “The NIST definition of cloud computing(draft)”, NIST Special Publication, pp. 800-145.
[4] A.Giuseppe,B.Alessio,D.Walter,P.Antonio,“Survey cloud monitoring: a survey, Computer Network”, International Journal of Computer and Telecommunications Networking, Elsevier, Vol. 57 No. 9, pp. 2093-2115, 2013.
[5] Uddin M, Rahman A A, “Energy efficiency and low carbon enabler green IT framework for datacenters considering green metrics”,Renewable and Sustainable Energy Reviews, Elsevier, Vol. 16, No. 6, pp. 4078–4094, 2012.
[6] Brown R, “Report to congress on server and data center energy efficiency Public law109 431”, Lawrence Berkeley National Laboratory 2008.
[7] Energy Star program requirements for computer systems–draft 4. Washington, DC: Environmental Protection Agency 2009.
[8] Qi Zhang, Lu Cheng, and RaoufBoutaba, “Cloud computing: State-of-the-art and research challenges”, Journal of Internet Services and Applications, Springer, Vol. 1, No. 1, pp 7-18, 2010.
[9] Srikantaiah, Shekhar, Aman Kansal, and Feng Zhao. "Energy aware consolidation for cloud computing", Proceedings of conference on Power aware computing and systems, Vol. 10, pp. 1-5. 2008.
[10] RajkumarBuyya, Anton Beloglazov, and JemalAbawajy. 2010. “Energy-efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges”. In Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications, 6–17.

[11] AntonBeloglazov, and RajkumarBuyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers."Concurrencyand Computation: Practice and Experience, Vol. 24, No.13, pp. 1397-1420, 2012.

[12] Gaojin Wen and Jue Hong, “Energy-aware Hierarchical Scheduling of Applications in Large Scale Data Centers”, 2011 International Conference on Cloud and Service Computing, IEEE, pp.158-165, 2011.

[13] Knauth, Thomas, and ChristofFetzer, “Energy-aware scheduling for infrastructure clouds.” 4th IEEE International Conference on Cloud Computing Technology and Science (CloudCom), IEEE, pages 58-65, 2012.

[14] Anton Beloglazov and Rajkumar Buyya, "Energy efficient allocation of virtual machines in cloud data centers." 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid), IEEE, pp. 577-578, 2010.

[15] Asyabi, Esmail, and Mohsen Sharifi, "A New Approach for Dynamic Virtual Machine Consolidation in Cloud Data Centers", International Journal of Modern Education and Computer Science (IJMECS), Vol. 7, No. 4, pp. 61-66, 2015.

[16] EshaBarlaskar and Y. Jayanta Singh, “Dynamic Consolidation of Virtual Machines with Multi-Agent System”, International Journal of Computer Applications, Vol. 94, No. 9, pp. 30-38,2014.

[17] Anton Beloglazov and RajkumarBuyya, “Managing Overloaded Hosts for Dynamic Consolidation of Virtual Machines in Cloud Data Centers Under Quality of Service Constraints”, IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 7, pp. 1366-1379, 2012.

[18] Barroso LA, Holzle U, “The case for energy-proportional computing” Computer, Vol. 12, pp. 33–37, 2007.

[19] Fan, Xiaobo, Wolf-Dietrich Weber, and Luiz Andre Barroso, "Power provisioning for a warehouse-sized computer." In ACM SIGARCH Computer Architecture News, ACM, Vol. 35, No. 2, pp. 13-23, 2007.

[20] Calheiros RN, Ranjan R, Beloglazov A, Rose CAFD, Buyya R. “CloudSim: a toolkit for modeling and simulation of Cloud computing environments and evaluation of resource provisioning algorithms” Software: Practice and Experience pp.23–50, 2011