Energy-Efficient Virtual Machine Live Migration in Cloud Data Centers

[1]Inderjit Singh Dhanoa, [2]Dr. Sawtantar Singh Khurmi
[1]Dept. of Computer Sciences, Desh Bhagat University, Mandi Gobindgarh, Punjab, India
[2]Ex. Professor, Bhai Maha Singh College of Engg., Muktsar, (affiliated to PTU Jalandhar), Punjab, India

Abstract
Cloud computing services will play an important role to meet various requirements of the clients in daily lives. In cloud computing, virtualization is an important issue to minimize cost incurred to manage data centers across the world. The energy consumption has become the reason for higher cost in operating data centers. Savings can be achieved by continuous consolidation with live migration of VMs depending upon the utilization of the resources, virtual network topologies and thermal state of computing nodes. This paper presents a review of research work done by researchers based on energy-aware Virtual Machine live migration from one host to another in cloud data centers and highlighting its key concepts with research challenges.

Keywords
Virtual Machines (VMs), Live Migration, Energy Overhead, Data Center

I. Introduction
Cloud computing is gaining importance day-by-day. The large number enterprises and individuals are shifting opting for cloud computing services. Thousands of servers have been employed worldwide to cater to the needs of customers for computing services by big organisations like Amazon, Microsoft, IBM and Google. The round-the-clock reliable computational services, fault tolerance and information security are the main issues to be addressed while providing services to geographically spread customers sites [1]. Cloud computing, also known as "pay as-you-go" utility model is economy driven. It becomes necessary for the service provider ensures load balance and reliable computing services to its clients round the clock worldwide and keeping services ON means consuming power all the time[4].
Another major issue of concern comes into consideration is how to minimize energy consumption or going for Green computing [2, 3]. Energy consumed by servers and in cooling data centers of cloud system is a costly affair. The United States Environmental Protection Agency (EPA) in its report says that energy consumption of only federated servers and data centers in this nation was 100 billion KWh in 2011 and infrastructure and energy (I&E) cost will be 75 percent of the total operation cost in 2014 [5]. Energy consumption of data centers has risen 56percent from 2005 to 2010 worldwide, and in 2010 it is accounted to be between 1.1percent and 1.5percent of the total electricity use [6] as shown in figure 1.Thus, reducing energy consumption is important and designing energy-efficient data centers has recently received considerable attention of research community.
There are a number of technologies, services, and infrastructure-level configurations that make cloud computing energy-efficient. Virtualization in cloud computing is a mechanism to abstract the hardware and the system resources from a given operating system. This is typically performed within a cloud environment across a large set of servers using a Hypervisor or Virtual Machine Monitor (VMM), that lies in between the hardware and the Operating System (OS)[7].

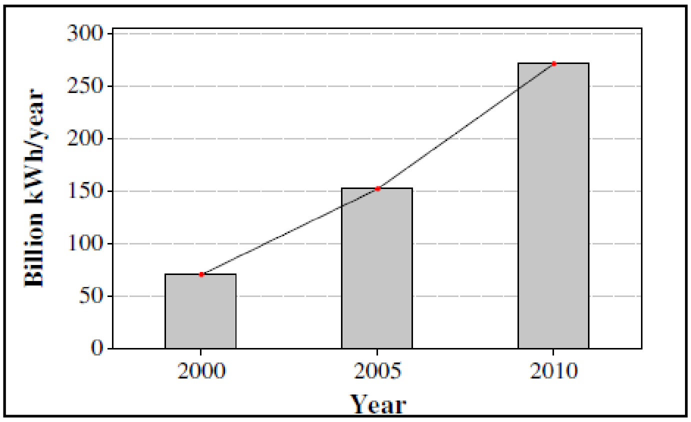

Fig. 1: Energy Consumption in Data Centers Worldwide

Through Hypervisor one or more virtualized OSs platforms can be made available (see fig. 2), one of the key advantage of cloud computing. Here cloud computing middleware is deployed on top of the virtualization technologies to exploit the capability to its maximum potential while still maintaining Quality of Service (QoS) to clients. QoS can be defined in terms of Service Level Agreement (SLA) between the service provider and the client.
The concept of consolidation of Virtual Machines (VMs) is applied to decrease energy consumption as it significantly reduces the percentage of idle power in the overall infrastructure. Such a consolidation can be done either statically or dynamically at run time. In the static approach, the mapping of the VMs to physical infrastructure can not be changed at runtime. A dynamic consolidation of VMs allows the reassignment of physical resources at runtime, when the load on the virtual machines increases or decreases. In case there is a low load on the VMs less physical resources need to be employed to provide certain performance level. In the other case, if the load on virtual machine increases, more physical resources can be assigned. The VMs can be migrated to another physical host if the current physical host gets overloaded. A dynamic consolidation of virtual machines mandates the cloud provider to monitor the resource utilization of virtual machines in order to determine how many physical resources have to be assigned for a particular event.

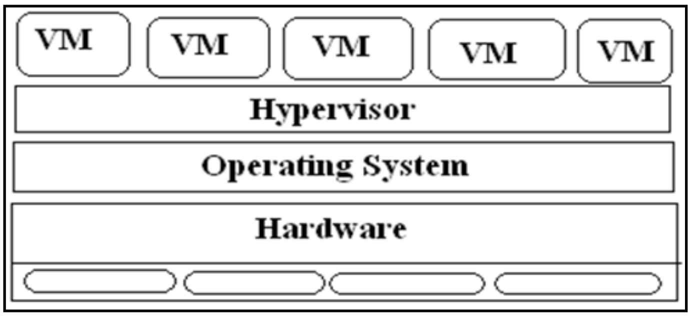

Fig. 2: Principle of Virtualization

Dynamic VM consolidation consists of two basic processes: Migrating VMs from underutilized hosts to minimize the number of active hosts; and Offloading VMs from hosts when those become

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

overloaded to avoid performance degradation as experienced by the VMs. The idle hosts automatically switch to a low-power mode to eliminate the static power and reduce the overall energy consumption by the system. Whenever required, the hosts are reactivated to accommodate new VMs or VMs being migrated.
Another capability provided by virtualization is live migration, which is the ability to transfer a VM between physical servers (referred to as hosts, or nodes) with a close to zero downtime. Using live migration, VMs can be dynamically consolidated to leverage fine-grained fluctuations in the workload and keep the number of active physical servers at the minimum at all times [8].
There are various advantages of live migration:

- Workload Balancing
- Maximum resource utilization
- Fault Tolerance
- Online System maintenance

A. Sources of Energy Waste

The inefficient use of computing resources consumes energy. The data collected from more than 5000 production servers shows that although servers are usually not idle, but their utilization rarely approaches cent percent. Most of the time servers operate at 10 to 50 percent of their capacity, leading to extra expenses on over-provisioning, and Total Cost of Acquisition (TCA) [9]. Moreover, managing and maintaining over-provisioned resources results in the increased Total Cost of Ownership (TCO) too. The problem of low server utilization is exacerbated by narrow dynamic power ranges of servers: completely idle servers consume up to 70percent of their peak power use [10].
VMs installed in data centers seldom communicate with each other. In virtual machine migrations, the communicating VMs may be hosted on distant physical nodes leading costly data transfers. The network communication may involve network switches that consume significant amount of energy [11].
Of course the energy overhead for VM migration cannot be considered negligible but the energy overhead from virtualized servers' increases as utilization of physical resources also [12].
Most of the energy spent in cooling ICT equipments in data centers energy is of the 45 percent of total energy costs [13].

II. Literature Review

Here the research work done is being highlighted in virtual machine live migration to minimize the power consumption of data centers.
L.Gergo et al introduced the energy-optimal allocation of virtualized services in a heterogeneous server infrastructure. Researchers proposed a model to predict the performance degradation of service when it is consolidated with other services. Two energy-efficient heuristics that approximate the energy-optimal and performance aware resource allocation problem are presented. This approach assumes a round robin process scheduler, CPU intensive services and response time as a performance metric [14].
Liang-Teh Lee et al proposed a mechanism to adjust the system voltage based on the CPU utilization, and migrating tasks in a heavy loaded machine to idle machines, so as to improve the resource utilization and reduce the energy consumption [15].
Yichao Jin et al investigated the impact of server virtualization on energy usage in physical servers and trade-off management between potential energy overhead introduced by hypervisor over the physical machine and reduction of maximum throughput for virtualized server [16].
Anton and Buyya presented a decentralized architecture of the energy aware resource management system for cloud data centers while meeting QoS requirements. The researchers present heuristics and three stages of continuous optimization of VM placement. Heuristics have been evaluated by simulation using extended Cloudsim toolkit in heterogeneous workload independent environment of VMs [17].
Nathuji and Schwan proposed architecture of energy management system for virtualized data centers, where resource management is divided into local and global policies. Consolidation of VMs is handled by global policies applying live migration to reallocate the VMs [18].
Song et al proposed resource allocation to applications according to their priorities in multi application virtualized clusters. But, it does not apply migration of VMs to optimize the allocation of resources to minimize power consumption [19].
Pinheiro et al proposed a technique for minimization of power consumption in a heterogeneous cluster of computing nodes serving multiple web-applications. The main technique applied to minimize power consumption is concentrating the workload to the minimum of physical nodes and switching idle nodes off. This approach requires dealing with the power / performance trade-off, as performance of applications can be degraded due to the workload consolidation. Requirements to the throughput and execution time of applications are defined in SLAs to ensure reliable QoS. The proposed algorithm periodically monitors the load of resources (CPU, disk storage and network interface) and makes decisions on switching nodes on / off to minimize the overall power consumption, while providing the expected performance. The actual load balancing is not handled by the system and has to be managed by the applications. The proposed approach can be applied to multi-application mixed-workload environments with fixed SLAs [20].
Manasa and Anirban conducted a survey of research in energy efficient computing and proposed architectural principles for energy efficient management of clouds, energy efficient VMs allocation policies and scheduling algorithms considering QoS expectations and power usage characteristics of the devices. It is validated by conducting a performance evaluation study using CloudSim toolkit [21].
Jing SiYuan considers dynamic on-demand resource provisioning which allows turning off part of idle servers to save energy. Author considers how to maximize the resource utilization without considering the overhead of virtual machine placement change. Author designed new method network-flow-theory based approximate algorithm to minimize the energy consumption and VM migration at the same time [22].
Hines et al. proposed a technique called dynamic self ballooning, where a driver runs in the guest VM, continuously reclaiming free and unused memory and giving it back to the hypervisor. This technique during migration reduces the amount of memory that is to be sent of the network since the reclaimed and unused memory is not transferred. The post-copy approach begins with the stop and copy phase and continues with the pull phase. In order for the destination machine to have all the data it requires, it retrieves them from the source continually. Several techniques can be used in order to retrieve the memory from the source, such as demand paging, active pushing, and adaptive pre-paging. The migration time of post-copy is mostly bounded by the amount of memory allocated to the VM since the memory is the bottleneck of saving and transferring the state. As opposed to pre-copy, post-copy will only transfer each memory page once [23].

Haikun Liu et al studied a more unorthodox method based upon tracing and replaying events instead of transferring data. This method, called system trace and replay, starts by taking checkpoint of the source VM. A checkpoint is a recorded state of the VM usually saved to disk to be resumed later. The system trace and replay instead transfers the checkpoint to the destination machine. Simultaneously, the source machine starts to record, or to trace, non-deterministic events such as user input and time variables. These events are recorded in a log and subsequently sent to the destination. The log transfer happens in a number of iterations similar to that of pre-copy. Thereafter, the destination executes, or replays, from the checkpoint and any non-deterministic event are read from the log. The replay mechanism is able perform faster than the original trace of events. This is required for the destination to catch up with the source. Deterministic events do not need to be recorded because the destination machine starts from a checkpoint and thus they will have the same deterministic events afterwards. The cycle of tracing and replaying goes on until the destination machine has a sufficiently small log; the source stops and copies the last of the log to the destination where the last of it is replayed and the migration has been completed at this point [24].

Kejiang Ye et al. focused on the live migration strategy of multiple VMs with different resource reservation methods. Authors performed series of experiments to investigate the impact of different resource reservation methods on the performance of live migration in both source machine and target machine. They analyzed the efficiency of parallel migration strategy and workload-aware migration strategy. Various Metrics like downtime, total migration time, and workload performance overheads were measured [25].

Experimental results showed that :(1) Live migration of virtual machine brings some performance overheads. (2) The performance overheads of live migration were affected by memory size, CPU resource, and the workload types.

Based on the observed results, they present corresponding optimization methods to improve the migration efficiency.

Takahiro Hirofuchi et al. reveal that their consolidation system with post-copy live migration is more efficient to reduce energy consumption than that of using pre-copy live migration. Authors focused on the comparison of pre-copy live migration technique and post-copy live migration technique with respect to energy efficient VM consolidation system. This system built up of three components mainly Load monitor to collect resource usage data every one second and put it into database, Relocation Planner to calculate optimal locations for VMs and VM controller to request live migration to server nodes according to the results from Relocation Planner. Then they found that to get the maximum energy saving, the consolidation system should be designed to be able to optimize VM locations at shorter intervals than one minute. Existing studies concerning VM packing have not addressed this kind of frequent optimization at such short intervals. Authors investigate that power consumption reduced by 11.8percent with post-copy live migration, and by 5.2percent with pre-copy live migration. Even this consolidation system based on post-copy live migration eliminated approximately half of the energy overheads as compare to pre-copy live migration [26].

Anja Strunk and Waltenegus Dargie experimentally investigate that live migration entails an energy overhead and the size of this overhead varies with the size of the virtual machine and the available network bandwidth. Authors classify the costs of virtual machine live migration into performance loss and energy overhead. The pre-copy and stop, and copy processes require additional resources, particularly, network bandwidth and some CPU cycles. This additional resource utilization cost during live migration creates energy overhead. The cost of migration in terms of power consumption is not negligible and power consumption during migration exceeds the idle power consumption by up to 63 percent [27].

Kateryna Rybina et al. experimentally investigate the magnitude of VM migration overhead in terms of energy consumption and service execution latency. Authors performed various tests with pre-copy live migration algorithm on two servers attached with Network Attached Storage. Authors migrated virtual machines under different configurations and bandwidth constraints in isolation from the source server to the destination server. During each migration, they measured and recorded the power consumption and the resource utilization (CPU and memory) of both servers. Authors investigate that during a VM migration the power consumption of both the source and the destination servers is higher than the power consumption of the servers before a migration was carried out. However, the power consumption of the source server was higher than the power consumption of the destination server in all the experiments. The power consumption of the source server during a migration was affected by the type of workload. The power consumption of the destination server during a migration did not depend on the workload type, since the VM was executing on the source machine during the migration. The energy overhead of a VM migration cannot be considered negligible. This is true regardless of the type of workload the VM was hosting. The VM migration time was not influenced by the type of workload running on the VM and was approximately the same for all types of workloads as long as the VM size and the network bandwidth were the same. The VM migration time was affected by the size of the VM and the available network bandwidth, particularly, by the network bandwidth. The energy overhead of a live VM migration significantly decreases with a higher network bandwidth [12].

Most of the existing or proposed approaches focus on the performance of live migration and measure migration time and down time, under different conditions. Work that explicitly investigates the energy cost of live migration is rare.

III. Research Directions

Designing energy-efficient data centers has recently received considerable attention. This problem has been approached from various perspectives:

- Energy efficient hardware architecture
- Virtualization of Computing Resources
- Energy-Aware Job Scheduling
- Dynamic Voltage and Frequency Scaling(DVFS)

These recent techniques are surveyed by Jyothi et al [28], but overcoming all the barriers for energy efficiency is not possible as each of the techniques throw light on different parameters with certain disadvantages of their own. We extend the set of challenging research directions by considering live migration issues in energy efficiency.

In this paper, recent research papers have been reviewed for research directions to minimize energy consumption of data centers. The studies showed that live migration feature of Virtualization Technology has ample scope to minimize energy consumption in data centers. There are some untouched issues which can be taken as future research work given as follows:

1. It would be crucial to investigate energy consumption of host servers during live migration of VMs with heterogeneous applications and multiple virtual machines.
2. VM network topology can help to minimize energy consumption, so it is necessary to observe the communication between VMs and to ensure placement of the communicating VMs on the same or closely located nodes.
3. Another challenge is to analyze the energy consumption of the various subsystems of the source and destination servers in order to manage for when and which VM should be migrated during service consolidation.
4. There is a need to develop energy efficient optimized live VM migration policy to minimize energy overhead during migration in data centers.
5. Another key research issue is to achieve a trade-off between application performance and energy efficiency during live migration of virtual machines in data centers.

IV. Conclusion

This paper presents survey of recent research on energy efficiency of Virtual Machine Live Migration in Cloud data centers. Energy efficiency in data centers is one of the most challenging issues faced by infrastructure providers today. Various papers have been reviewed with their policies to minimize energy consumption in data centers. But still there are some issues specially related with Live Migration of Virtual Machines which are not investigated for better energy management. Research directions have been discussed to further optimize energy consumption using live migration of Virtual Machines in various perspectives.

In future work it would be more significant to investigate various research directions given in this paper to optimize energy requirements for cloud services during Virtual Machine live migration in data centers.

References

[1] R. Buyya, C. Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility," Future Generation Computer Systems 25 (6) pp. 599– 616, 2009.

[2] Tharam Dillon, Chen Wu, Elizabeth Chang, "Cloud Computing: Issues and Challenges", 24th IEEE International Conference on Advanced Information Networking and Applications, IEEE Computer Society, 2010.

[3] Marios D. Dikaiakos, George Pallis, Dimitrios Katsaros, Pankaj Mehra, Athena Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific," IEEE Internet Computing, Vol.13, no.5, pp.10-13, Sept. 2009.

[4] Massoud Pedram,"Energy-Efficient Datacenters," IEEE Transactions on Computer –Aided Design of Integrated Circuits and Systems, Vol. 31, No. 10, pp. 1465-1484, October 2012.

[5] Rongbo Zhu, Zhili Sun, Jiankun Hu, "Special section: Green computing," Future Generation Computer Systems, pp. 368-370, 2012.

[6] Jonathan Koomey, "Growth in data center electricity use 2005 to 2010,"Analytics Press, Tech. Report.,2011.

[7] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, A. Warfield, "Xen and the art of virtualization," SIGOPS Operating Systems Review, Vol. 37, no. 5, pp. 164–177, Oct. 2003.

[8] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpet, I. Pratt, A. Warfield, "Live migration of Virtual Machines ," In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation - Volume 2, Berkeley, CA, USA: USENIX Association, 2005, pp. 273–286. [Online]. Available:http://dl.acm.org/citation.cfm?id=1251203.1251223.

[9] L.A Barroso, U.Holzle, "The case for energy-proportional computing," Computer,Vol.40,no.12,pp. 33-37,2007.

[10] X.Fan.W. D.Weber and L.A. Barrao, "Power provisioning for a warehouse-sized computer, "In proceedings of the 34th Annual International Symposium on Computer Architecture(ISCA),2007,pp.13-23.

[11] Anton Beloglazov, "Energy-efficient Management of Virtual Machines in Data centers for Cloud Computing", Thesis submitted to The University of Malborne, 2013.

[12] Kateryna Rybina et. al., "Investigation into the Energy Cost of Live Migration of Virtual Machines," The Third IFIP Conference on Sustainable Internet and ICT for Sustainability (SUSTAINIT 2013), October 30-31, Palermo, Italy; 2013.

[13] Kliazovich D,Bouvry P,Khan S,"Dens: Data center energy efficient network-aware scheduling", IEEE/ACM International Conference on Cyber, Physical and social computing (CPSCom),pp 69-75,2010.

[14] Gergo L., N.Florian and Hermann de Meer , "Performance Tradeoffs of Energy-Aware VM Consolidation ," Journal of Cluster Computing(Springer), Volume 16, Issue 3 , pp 481-496, 2013.

[15] Liang-The Lee, Kang-Yuan Liu and Chia-Ying Tseng, "A dynamic resource management with energy saving Mechanism for storing Cloud Computing," International Journal of Grid and Distributed Computing,Vol-6 Feb,2013.

[16] Yichao Jin, Yonggang Wen , "Energy Efficient and Server Virtualization in Data Centers: An Empirical Investigation," IEEE conference on Computer Communications Workshops (INFOCOM WKSHPS), Orlando, FL,2012.

[17] Anton Beloglazoy, Rajkumar Buyya , "Energy Efficient Resource Management in Virtualized Cloud Data centers," 10th International Conference on Cluster, Cloud and Grid Computing,2010.

[18] R. Nathuji and K. Schwan, "VirtualPower: Coordinated power management in virtualized enterprise systems," ACM SIGOPS Operating Systems Review, vol. 41, no. 6, pp. 265–278, 2007.

[19] Y. Song, H. Wang, Y. Li, B. Feng, and Y. Sun, "Multi-tiered on-demand resource scheduling for VM-based data center," In Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2009), Shanghai, China, pp. 148–155,2009.

[20] E. Pinheiro, R. Bianchini, E. V. Carrera, T. Heath, "Load balancing and unbalancing for power and performance in cluster-based systems," In Proceedings of the Workshop on Compilers and Operating Systems for Low Power (COLP),pp. 182–195,2001.

[21] Manasa H.B, B.Anirban, "Energy Aware Resource Allocation in Cloud Data Centers," International Journal of Engineering and Advanced Technology, Volume-2, Issue-5, June 2013.

[22] Jing SiYuan, "A Novel Energy Efficient Algorithm for Cloud Resource Management," International Journal of Knowledge and Language Processing, Volume 4, Number 2, 2013.

[23] Michael R. Hines, Kartik Gopalan, "Post-copy based live VM migration using adaptive pre-paging and dynamic self-ballooning," In Proceedings of ACM SIGPLAN/SIGOPS international conference on Virtual execution environments

(VEE), pages 51-60, New York, USA, 2009.
[24] H. Liu, C.-Z. Xu, H. Jin, J. Gong, X. Liao, "Performance and energy modeling for live migration of Virtual Machines," In Proceedings of the 20th international symposium on High performance distributed computing, ser. HPDC '11. New York, NY, USA: ACM, pp. 171–182, 2011.
[25] Kejiang Ye, Xiaohong Jiang, Dawei Huang, Jianhai Chen, Bei Wang,"Live Migration of Multiple Virtual Machines with Resource Reservation in Cloud Computing Environments," Cloud Computing (CLOUD), 2011 IEEE International Conference, pp.267-274, 4-9 July 2011.
[26] Takahiro Hirofuchi et.al, " Making VM Consolidation More Energy-efficient by Post-Copy Live Migration," International Conference on Cloud Computing, Grids and Virtulization,2011.
[27] Anja Strunk, Waltenegus Dargie , "Does Live Migration of Virtual Machines Cost energy?," In IEEE 27th International Conference on Advanced Information Networking and Applications, (AINA), pp 514-521,March 2013.
[28] Jyothi Sekhar ,Getzi Jeba, S.Durga, "A Survey on Energy Efficient Server Consolidation through VM Live Migration", International Journal of Advances in Engineering & Technology,Vol.5,Issue 1,pp 515-525,2012.
[29] Wu Y., Zhao M. "Performance Modeling of Virtual Machine Live Migration," In proceedings of IEEE, pp.492-499, 2011.

Inderjit Singh Dhanoa, Dept. of Computer Sciences, Desh Bhagat University, Mandi Gobindgarh, Punjab, India.