ISSN:0975-9646

Green Data Center Using Spearman's Ranking Algorithm

Sonu Choudhary, Dr. Abhay Kothari

Department of Computer Science & Engineering,
Acropolis Institute of Technology & Research
Indore, Madhya Pradesh, India

***Abstract*— Green computing is the technological investigation where the system performance is optimized for achieving high performance computing in fewer amount of resource and power consumption. Therefore the presented work is an initiative for optimizing power consumption and performance enhancement by VM allocation and selection policies considering QoS expectations of the devices, which maximize the computational ability and minimize the power consumption. Thus first the detailed survey on VM allocation and selection approaches is performed for finding the optimum technique of the power preserving techniques for cloud computing. It is concluded that power consumption can be minimized by the efficient VM scheduling. Thus the traditional techniques namely MAD (Median Absolute Deviation), Random Selection and Maximum Correlation Coefficient techniques are implemented using CloudSim simulator. The simulator is build using JAVA technology. Additionally a new technique for optimizing the power consumption is utilized namely Spearman's Rank Correlation Coefficient which provides the ranked value of VM CPU scheduling. After implementing the proposed and traditional techniques, the comparative performance is computed and demonstrated. According to the obtained results the proposed technique less violate the SLA terms and provides the gain over the power consumption.**

***Keywords*— Cloud Computing, QoS (Quality of Service), SLA (Service Level Agreements), VM (Virtual Machine), MAD (Median Absolute Deviation).**

I. INTRODUCTION

Cloud computing is a high performance computing infrastructure for new generation computing and service distribution. But during computing a huge amount of power is consumed. Therefore a green computing domain is introduced for efficient computing and green effect. Green computing or green IT, refers to environmentally sustainable computing or IT. It is the study and practise of designing, manufacturing, using and disposing of ICT—efficiently and effectively with minimal or no impact on the environment. In order to achieve green computing and reducing the power consumption of the computational cloud environment required to schedule VMs efficiently. VMs are the software implementation of the particular computer system and provide sharing of physical hardware resources with users.

Cloud computing facilitates resources to the user by a huge collection of resources which referred to as resource pooling. High level of resource pooling led to the large scale of data centers establishment. These data centers consume large amount of power and results CO_2 emission and high operating cost. The problem of power consumption in large scale datacentre can be mitigated with the help of virtualization technology which allows creation of multiple instances of VMs over the physical hosts deployed in data centers [2]. This paper focus upon the energy-efficient VM scheduling that can be applied in a large-scale data center. In this paper spearman's ranking algorithm for VM scheduling is proposed that can be reduced power consumption significantly. Further, the comparative study among MAD-MC, MAD-RS and proposed Spearman's Ranking VM allocation and selection policies performed.

The remaining part of this paper is organized as follows to find the optimum solution. Section2 discuss proposed work followed by problem statement and simulation architecture in order to achieve low power consumption. The implementation of VM scheduling discuss in section3. A performance analysis of the proposed method present in section4. Section5 conclude the paper with summary and future work.

II. PROPOSED WORK

Cloud data center frequently uses the virtual machine migration and allocation techniques to provide efficient computing experience. Basically cloud data centers are composed with a number of virtual machines and these virtual machines are free to select different CPUs. In order to provide the efficient computing VMs are scheduled according to available resource allocation policies. Thus if the VM scheduling becomes efficient and the time of allocation and selection of VMs are minimized then power consumption can be enhanced. Based on this concept, solution is suggested and simulated by simulation using real world workload traces from more than a thousand Planet Lab VMs. The complete study includes the following work.

- Firstly, different approaches for reducing power consumption in a data center are studied and most optimum technique is investigated for extension.
- A new method is proposed and designed, that is providing less energy consumption in a data center.
- The designated algorithm is implemented using the CloudSim simulator. Additionally the recently used methods are also implementing for comparative study.
- Finally the comparative analysis of methods used for justifying the newly proposed solution.

In this research the main objective is to develop efficient policies for virtualized data center so that cloud computing

can be more sustainable and echo-friendly technology for future generation.

A. Problem Statement

Cloud data center emerged as need for rapidly increasing computational power and results greater power consumption and high carbon footprints to environment. Thus a key issue is to minimize power consumption along with SLA considerations. There are different techniques have been proposed in the past few years to overcome power consumption in cloud data center. Live migration technique migrate a running VM from one physical host to another and require more VM management.VM consolidation techniques gives efficient power consumptions and VM migration in dynamic workload experience but more SLA violations. Therefore the following issues are considered in the proposed study.

- The available allocation and selection techniques are not much efficient.
- The available VM consolidation technique is efficient but more SLA violation take place.
- Maximum coefficient correlation selection technique presents only the linear relationship between datasets and random selection technique migrate VMs randomly, which is inefficient.

B. Solution Domain

In order to achieve low power consumption in data center, VM scheduling is performed using Spearman's Rank Correlation Coefficient for VM selection. In VM scheduling firstly VM allocation is performed using MAD technique which determines overloaded and under-loaded host. Once hosts are determined, VM selection policy applied to select VMs for migration from that hosts for utilization.

1) Median Absolute Deviation Allocation Policy: MAD is an adaptive utilization threshold method setting upper and lower utilization threshold for hosts and keeps the total utilization of the CPU by all the VMs between these thresholds. If the CPU utilization of a host below the lower threshold, all VMs have to be migrated from this host and the host has to be switched to the sleep mode in order to eliminate the idle power consumption. If the utilization exceeds the upper threshold, some VMs have to be migrated from the host to reduce the utilization in order to prevent a potential SLA violation.

For a uni-variate data set $X_1, X_2, \ldots, X_n$, the MAD is defined as the median of the absolute deviations from the data's median:

$$MAD = median_i\left(\left|X_i - median_j\left(X_j\right)\right|\right)$$

2) Spearman's Correlation Coefficient Selection Policy: Spearman's Correlation Coefficient used for finding the optimum VM selection according to the workload and availability of resources. It shows monotonic relationship between two variables. In a monotonic relationship both variables increases concurrently but not at the same rate. Spearman's rank correlation coefficient is denoted by r_s. It is appropriate when one or both variables are ordinal or skewed and is robust when extreme values are present. The formula for calculating the sample Spearman's correlation coefficient for two variable X and Y is given by:

$$r_s = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n(n^2-1)}$$

Where d=|X-Y|, X is the set of work load attributes and Y is the available resource parameters.

C. Simulation Architecture

This architecture presented by which the implementation work is performed. The present system model includes different modules having their own methodology to find the energy consumption of a data center.

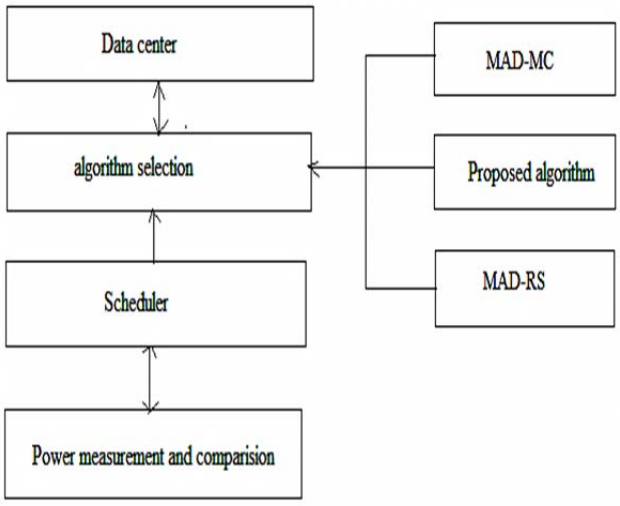

Fig. 1 Simulation Architecture

The given model is described as:

1) Data Center: We have simulated a data center that consists 800 heterogeneous half of which are hosts HP ProLiant ML110 G4 servers, and the other half consists of HP ProLiant ML110 G5 servers.

2) Algorithm Selection: This module select appropriate scheduling algorithm during simulation.

3) Scheduler: In this phase scheduling is performed according to selected algorithm.

4) Power Measurement and Comparison: This phase includes power measurement and comparison study among scheduling algorithms. The performance of these algorithms is evaluated in terms of time and space complexity and the energy consumption as well as SLA and VM migrations of the scheduling algorithms.

III. IMPLEMENTATION

A. Simulation Setup

In this research provides the desired configuration for simulation of power optimization and efficient VM scheduling approach in cloud environment.

TABLE I
SIMULATION SETUP

Simulation Properties	Values
Scheduling Interval	300ms
Virtual Machine RAM	870,1740,1740,613
Host RAM	4096,4096
VM Band Width	100000Mbit/s
Host Band Width	1000000Gbit/s
Host Storage	1000000GB
Host Power	HpProLiantMl110G4Xeon3040

B. Simulation Scenarios

In order to simulate VM scheduling for efficient power consumption the following simulation scenarios are prepared.

1) Simulation using MAD-MC VM Scheduling: In this simulation implementation the MAD (Median Absolute deviation) allocation and Maximum Correlation Coefficient selection policies of VM scheduling approach are implemented using the CloudSim discrete event simulator and performance of scheduling is calculated.

2) Scheduling using MAD-RS VM Scheduling: In this simulation implementation the MAD (Median Absolute deviation) allocation and Random Selection policies of VM scheduling approach are implemented using the CloudSim that is also predefined CloudSim VM scheduling methodology.

3) Scheduling using Proposed VM Scheduling: In this simulation analysis the enhanced correlation coefficient known as Spearman's Rank Correlation Coefficient is implemented with the CloudSim simulation technique and the performance of power consumption is compared.

IV. RESULT ANALYSIS

The comparative performance evolution of Spearman's Rank Correlation Coefficient technique is performed in this research paper over different performance parameters.

A. Energy Consumption

The comparative energy consumption of traditional techniques and presented technique is provided in this research study. The total energy consumption is the sum of energy consumed by the physical resources of a data center as a result of application workloads.

In given Fig.2, the red line shows the energy consumption using MAD-MC, green line shows the energy consumption using MAD-RS and the proposed method is given using blue line. For representing the performance evolution X axis shows the simulation time and Y axis shows the power used by data center.

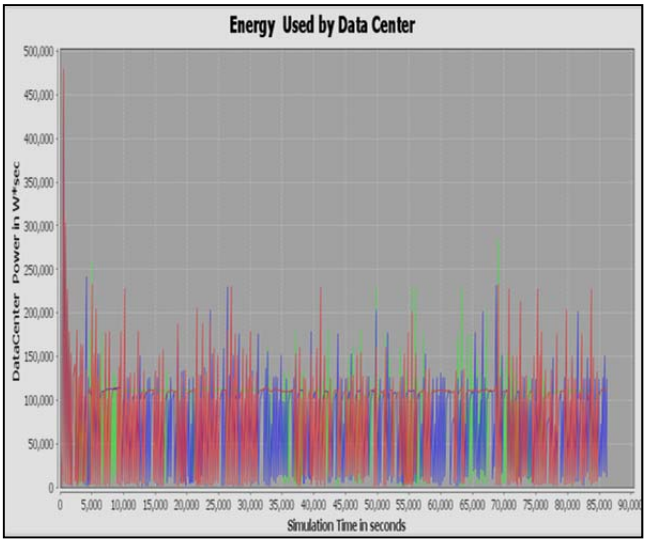

Fig. 2 Comparative Energy Consumption of Data Center

B. SLA Violation

When a VM cannot get the promised Quality of Service (QoS), SLA violation takes place. For example when a VM cannot get requested MIPS SLA violation issue occur. The performance in terms of SLA violation is measured in Fig.3, both the existing techniques violate frequently the SLA terms as compared to the proposed technique.

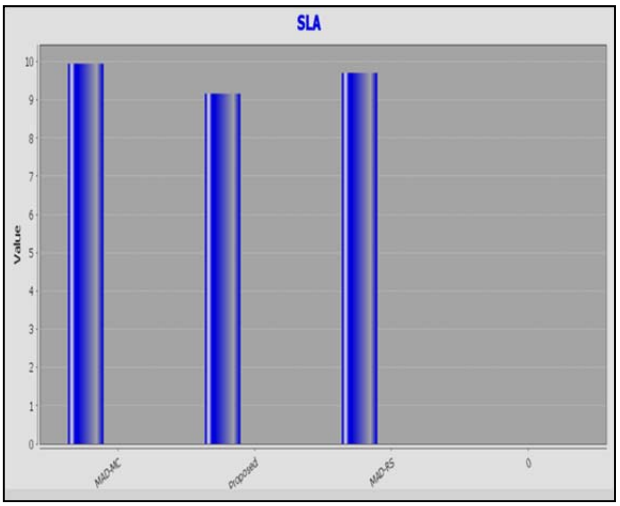

Fig. 3 SLA Violation

C. VM Migrations

For VM scheduling it is must require detecting overloaded and under-loaded host and once overloaded or under-loaded hosts found the VMs are selected for migration. Fig.4 shows the live VM migration and their switching capability.

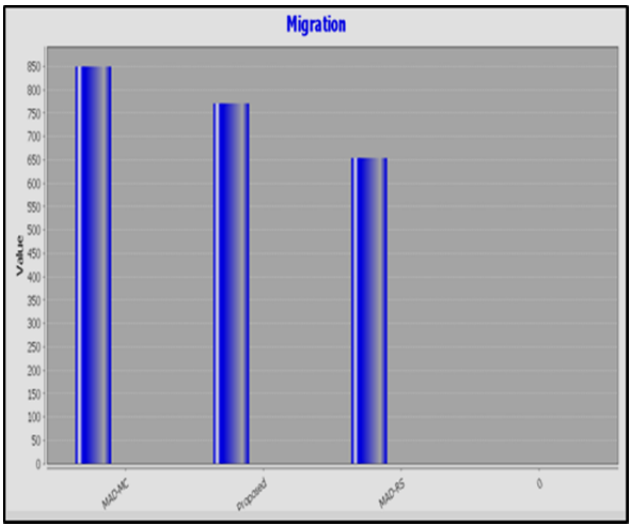

Fig. 4 Number of VM Migrations

D. *Memory Used*

The amount of main memory required for process the algorithms are known as the memory consumption or space complexity of the system. The space complexity of MAD-MC, MAD-RS and proposed algorithm is given in Fig.5. In this diagram the memory consumption of the algorithms are given in Y axis and the X axis contains the different experiments performed on the system.

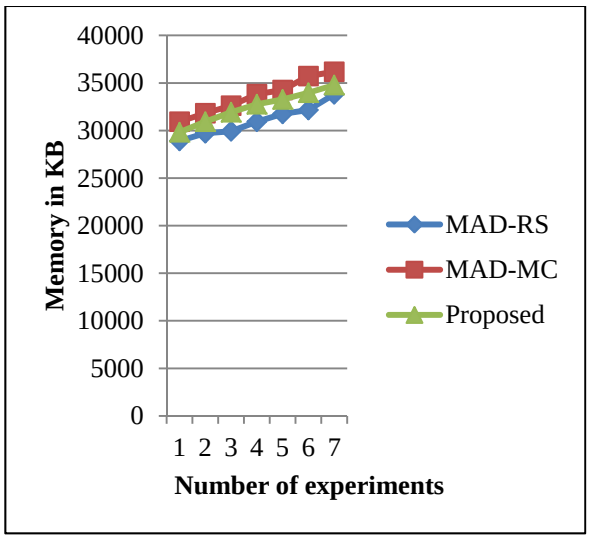

Fig. 5 Memory Consumption

E. *Time Complexity*

The amount of time required to schedule VM is known as time complexity.

The comparative time complexity of the implemented algorithms are given using Fig.6, in this diagram the X axis contains different experiments performed using the designed system and Y axis shows the observed time complexity of individual algorithms. According to the obtained performance the proposed method needs less time for efficient VM scheduling as compared to both the traditional methods.

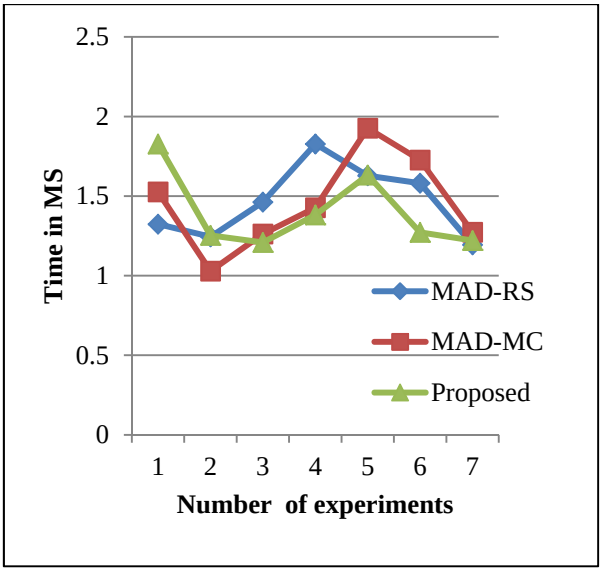

Fig. 6 Time Complexity

V. Conclusion And Future Work

The computational cloud is one of the most essential technologies to fulfill the demand of the new generation computational need. In recent years the demand of computational engines are increasing rapidly and new organizations and institutions are believe on the performance of the cloud infrastructure. The cloud computing infrastructure helps in efficient computation, data hosting and other various tasks. On the hand the cloud systems are affected with the huge amount of power consumption. Thus a new computational branch of the cloud is established as the green computing which keep in track the power consumption and green effect of the computational cloud.

In this presented work the cloud computing is investigated for their green computing technology and their power preservation techniques. There are a number of techniques recently developed for green computing and power preservation. Among them the efficient VM scheduling and performance improvement is an essential technique for power preservation. Therefore in this presented work the VM scheduling is enhanced using the Spearman's Rank Correlation Coefficient technique and the comparative performance is analyzed with respect to the MAD-MC (Median Absolute Deviation and Correlation Coefficient) and MAD-RS (Median Absolute Deviation and Random Selection) techniques. The implementation of all the techniques is performed over the CloudSim and JAVA technology. The presented Spearman's Rank Correlation Coefficient technique is good as compare to traditional techniques in terms of computational complexity and the power preserving capability.

The presented Spearman's Rank Correlation Coefficient technique for VM scheduling is found optimum, efficient and adoptable. In near future the proposed work is to be enhanced for preserving more power to achieve green computing.

REFERENCES

[1] R. Shaikh and M. Sasikumar, "Security Issues in Cloud Computing: A Survey", *International Journal of Computer Applicatins (0975-8887)*, vol.44, no.19, April 2012.

[2] Prof. R. A. Sheikh and Dr. U. A. Lanjewar, "Green Computing-Embrace a Secure Future", *International Journal of Computer Applicatins (0975-8887)*, vol.10, no.4, November 2010.

[3] A. Huth and J. Cebula, "The Basics of Cloud Computing", *Carnegie Mellon University, Produced for US-CERT*, 2011.

[4] Introduction to Cloud Computing, White Paper, Dialogic, 2013.

[5] N. Mirzaei, "Cloud Computing", Community Grids Lab, *Indiana University Pervasive Technology Institute*, 2008.

[6] M. Ricciuti, "Stallman: Cloud Computing is Stupidity", Available: http://news.cnet.com/8301-1001_3-10054253-92.html

[7] M. Ursacescu, "Green ICT Awareness in Organization-An Empirical Study in Romanian Companies", *EconomiaSeria Management*, vol.17, Issue1, 2014.

[8] E. K. Lee, H. Viswanathan and D. Pompili, "VMAP: Proactive Thermal-Aware Virtual Machine Allocation in HPC Cloud Datacenters", *this work was supported by the NSF Award no.CSR-1117263*.

[9] M. Mishra and S. Sudevalayam, "Introduction to Cloud computing and virtualization". Available: http://www.cse.iitb.ac.in/convergence/workshops/Intro_to_Virtualization.pdf

[10] VMWare, VMWareInc. Available: http://www.vmware.com

[11] Xen User Manual. Available: http:// bits.xensource.com/Xen/docs/user.pdf

[12] Tremulous official website. Available: http://tremulous.net

[13] R. Nathuji and K. Schwan, "VirtualPower: coordinated power management in virtualized enterprise systems", ACM Symposium on Operating Systems Principles, *Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles*, 2007.

[14] C. P. Sapuntzakis, R. Chandra, B. Pfaff, J. Chow, M. S. Lam and M.Rosenblum. "Optimizing the migration of virtual computers". *In Proc. of the 5th Symposium on Operating Systems Design and Implementation (OSDI-02)*, December 2002.

[15] S. Osman, D. Subhraveti, G. Su and J. Nieh. The design and implementation of zap: "A system for migrating computing environments". *In Proc. 5th USENIX Symposium on Operating Systems Design and Implementation (OSDI-02)*, pp. 361–376, December 2002.

[16] J. G. Hansen and A. K. Henriksen. Nomadic operating systems. Master's thesis, Dept. of Computer Science, University of Copenhagen, Denmark, 2002.

[17] H. Hartig, M. Hohmuth, J. Liedtke and S. Schonberg. "The performance of microkernel-based systems". *In Proceedings of the sixteenth ACM Symposium on Operating System Principles*, pp. 66–77. ACM Press, 1997.

[18] C. Clark, K. Fraser, S. Hand, J. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield. "Live migration of Virtual Machines". *In USENIX NSDI*, 2005.

[19] P. Chandran and Dr. V. Patil, "Green Computing Technologies towards the Development of ICT: A Critical Study", *Proceedings of National Conference on Emerging Trends: Innovations and Challenges in IT*, pp.19 -20, April 2013.

[20] L. Wang and S. U. Khan, "Review of performance metrics for green data centers: a taxonomy study", *Springer Science and Business Media LLC*, 2011.

[21] M. Kazandjieva, B. Heller, O.P. Gnawali, P. Levis, and C.Kozyrakis, "Green Enterprise Computing Data: Assumptions and Realities", *IEEE*, 2012.

[22] N. Ibrahim, "Towards Green Service-oriented Computing", ISBN: 978-0-9891305-3-0, *SDIWC*, 2013.

[23] M. Gupta and G. Gupta, "Green Computing – "A Step towards Better Milieu", Journal of Engineering", *Computers & Applied Sciences (JEC&AS)*, ISSN No: 2319-5606, vol. 2, no.9, September 2013.

[24] G. K. Sehdev and A. Kumar, "Power Efficient VM Consolidation using Live Migration- A step towards Green Computing", *International Journal of Science and Research*, vol.3 ,Issue 3, March 2014.

[25] K. Salah, R. Al-Shaikh and M. Sindi, "Towards Green Computing using Diskless High Performance Clusters", *7th International Conference on Network and Service Management (CNSM)*, 2011.

[26] B. Valeri, M. Baez and F. Casati, "Come Along: understanding and motivating participation to social leisure activities", *IEEE Third International Conference on Cloud and Green Computing*, 2013.

[27] N. J. Kansal and I.Chana, "Cloud Load Balancing Techniques: A Step Towards Green Computing", *IJCSI International Journal of Computer Science Issues*, vol. 9, Issue 1, no. 1, January 2012.

[28] J. Baliga, R. W. A. Ayre, K. Hinton and R. S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport", vol. 99, no. 1, *Proceedings of the IEEE*, January 2011.

[29] S. Singh and I.Chana, "Energy based Efficient Resource Scheduling:"A Step Towards Green Computing", *International Journal of Energy, Information and Communications*, vol.5, Issue 2, pp.35-52, 2014.

[30] K. Maurya and R. Sinha, "Energy Conscious Dynamic Provisioning of Virtual Machines using Adaptive Migration Thresholds in Cloud Data Center", *IJCSMC*, vol. 2, Issue. 3, pp.74 – 82, March 2013.

[31] K. Davenport, "Absolute Deviation Around the Median", *R news and tutorials contributed by (452) R bloggers*.

[32] S. Pandey, L. Wu, S. M. Guru and R. Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments", *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*.